Functional and structural characterisation of the epigenetic regulator, SMCHD1

Alexandra Dolores Gurzau

ORCID: 0000-0001-8664-0698

A thesis submitted in total fulfillment for the
degree of Doctor of Philosophy

Walter and Eliza Hall Institute
Department of Medical Biology
University of Melbourne

August 2020

Abstract

Structural Maintenance of Chromosomes Hinge Domain-containing protein 1 (SMCHD1) has been established as an epigenetic regulator, with critical roles in X-chromosome inactivation, autosomal gene silencing and genomic imprinting. Recently, variations in *SMCHD1* have been associated with two human conditions: facioscapulohumeral muscular dystrophy (FSHD) and Bosma arhinia microphthalmia syndrome (BAMS). There has therefore been a growing interest in unveiling SMCHD1's atomic structure and the molecular mechanisms underlying its function in both a healthy and diseased state.

To provide a better understanding of Smchd1's molecular structure and function, I successfully expressed and purified the full-length 2007-amino acid mouse Smchd1 protein. Electron microscopy analyses of the Smchd1 dimer revealed an elongated rod-like structure that displays a high conformational flexibility, similar to that of other structural maintenance of chromosomes (SMC) proteins. This flexibility is largely conferred by the intermediate region of the protein that connects Smchd1's two functional domains: the N-terminal GHKL ATPase and the C-terminal SMC hinge domain. In follow-up studies of the two individual domains, we revealed the first atomic-resolution structure of Smchd1's hinge domain, providing a novel insight into its DNA-binding and dimerisation modes. Contrary to previously suggested models describing the DNA interaction mode of canonical SMC proteins, I showed that nucleic acids are not threaded through the central pore region of the Smchd1 hinge domain. Subsequent immunofluorescence studies additionally revealed that the hinge domain targets full-length Smchd1 to chromatin, and that a functional hotspot within the hinge is required for chromatin localisation in cells.

SMCHD1's ATPase domain has been of particular interest due to the identification of disease-related variants that are frequently located within this region of the protein. However, the mechanisms by which some of these pathogenic variants affect SMCHD1 function are poorly understood. Using analytical ultracentrifugation, I demonstrated that the wild-type SMCHD1 ATPase undergoes dimerisation, which was reliant on the inclusion of both the UBL domain and the presence of substrate, ATP. Follow-up cellular

studies revealed that Smchd1's catalytic activity, as well as the presence of the newly-identified UBL domain, are both necessary for the localisation of full-length Smchd1 to chromatin.

Together, these studies provide an insight into the molecular basis of Smchd1 function and highlight how chromatin binding may be compromised in human disease. Future studies will further investigate the cellular localisation and dimerisation properties of disease-associated SMCHD1 variants, contributing towards our ongoing drug development program aimed at developing therapeutic treatments for FSHD patients.

Declaration

This is to certify that:

- this thesis comprises only my original work, except where indicated in the Preface
- due acknowledgement has been made in the text to all other material used,
- the thesis is fewer than 100,000 words, exclusive of tables, maps, bibliographies and appendices.

Alexandra Dolores Gurzau

August 2020

Preface

In accordance with the regulations governing the degree of Doctor of Philosophy at the University of Melbourne, I hereby submit that:

- The author's contribution to Chapter 3 was 85%. Dr Tracy Willson introduced point mutations in the full-length Smchd1 constructs. Differential scanning fluorimetry experiments were performed by staff members at the CSIRO Collaborative Crystallisation Centre (C3). Dr Jarrod Sandow performed mass spectrometry analyses.
- The author's contribution to Chapter 4 was 80%. Dr Jarrod Sandow performed the hydrogen deuterium exchange and mass spectrometry experiments. Analytical centrifugation experiments were performed by Dr Yee-Foong Mok and the data was analysed with the help of Dr Christopher Horne. SMCHD1-knockdown HEK293 cells were generated by Ms Ruoyun Wang and Ms Megan Iminitoff, where the shRNA guides were designed by Dr Andrew Keniry. Full-length Smchd1 constructs with point mutations or domain deletions were prepared by Dr Tracy Willson.
- The author's contribution to Chapter 5 was 85%. SMCHD1-knockdown HEK293 cells were generated by Ms Ruoyun Wang and Ms Megan Iminitoff, where the shRNA guides were designed by Dr Andrew Keniry. Full-length Smchd1 constructs with point mutations or domain deletions were prepared by Dr Tracy Willson.

Therefore, the author's overall contribution to the work presented in this thesis was 80%.

Acknowledgements

I would like to foremost thank my two supervisors, James Murphy and Marnie Blewitt, for giving me the opportunity to undertake this project. I was extremely lucky to have not only one, but two excellent mentors, who guided and encouraged me throughout this journey. Both James' and Marnie's admirable work ethic and enthusiasm about science were truly inspiring, and helped motivate me throughout the course of my PhD. I am also very grateful to my supervisors for providing me with the opportunity to learn various scientific techniques, allowing me to expand my skills and broaden my future career opportunities.

I would like to thank Kelan Chen, who dedicated a tremendous amount of her time and scientific expertise to help guide me through the early stages of my Honours and PhD studies. Kelan's findings also greatly informed our field and current knowledge of our favourite protein, and therefore a large amount of the work I presented in this thesis expands on her initial findings. I would also like to thank my PhD advisory committee members: Isabelle Lucet, Michael Griffin and Mark Dawson. Our discussions and informal meetings were incredibly beneficial, stimulating many new ideas and ways to tackle difficult scientific questions.

Many thanks to all the current and former members of the Murphy lab. Cheree Fitzgibbon, who is incredibly hard-working and was always happy to assist with various tasks. Sam Young for his perpetual assistance with cloning experiments, and his amicable and good-humoured nature. Emma Petrie, who was a great mentor and friend, and Andre Samson for his advice on several experimental techniques. Thank you to Chris Horne for his help with analytical ultracentrifugation studies, and his general suggestions and input towards my project. To past and present students, Katherine Davies, Lung-Yu Liang, Sarah Garnish, Yanxiang Meng, Catia Pierotti, Annette Jacobsen and Daniel Frank, thank you for many great discussions and for being so supportive during this journey.

A big thank you to all Blewitt lab members, especially Iromi Wanigasuriya and Megan Iminitoff who assisted with tissue culture and imaging experiments. Thank you to Natalia Benetti, Andres Tapia del Fierro, Kelsey Breslin, Sarah Kinkel, Quentin Gouil, Jenny Steiner, Andrew Keniry, Tamara McLennan and Ayush Semwal for their positive

and friendly vibes and for being such a pleasure to work with. I would also like to acknowledge all members of both the Inflammation (former Cell Signalling and Cell Death) and Epigenetics and Development (former Molecular Medicine) Divisions for providing a very supportive and friendly work environment.

I would like to thank Richard Birkinshaw who was highly involved in this project and whose expertise in structural biology and protein crystallography was very valuable. I would like to acknowledge Tracy Willson for her assistance with cloning experiments, which were pivotal to my project, Wilson Wong for introducing me to electron microscopy and assisting with initial experiments, and Jarrod Sandow who helped me perform and analyse several mass spectrometry studies. I would also like to thank Andrew Leis and Yee-Foong Mok from the Bio21 Institute who assisted me with electron microscopy and analytical centrifugation experiments, respectively.

A big thank you to Katherine Davies and Jonathan Bernardini for founding our very exclusive "coffee club". The wholesome chats and dopamine-inducing coffee made the past few years more enjoyable and helped me choose to see the cup half-full.

Despite some challenging times, the past few years were a tremendous experience that challenged my resilience, but helped me grow and evolve as an individual as well as a scientist. For better or worse, my family and close friends shared that experience with me. To Mum, Dad, Vali, and our furry family member, thank you for your understanding, your patience, and your unconditional love and support during these testing times. To my closest friends, Gabby and Claudiu, despite being scattered across different parts of the world, our friendship has withstood the tests of time, distance, and all the new challenges associated with adulthood. To Iromi, who closely shared this experience with me, thank you for being a great friend, for agreeing to go on spontaneous trips with me and for the many comforting discussions. Your endless support and encouragement has been incredibly valuable, and as cliché as it sounds, I wouldn't have made it this far without you all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Epigenetic Control of Gene Expression

1.1.1 Introduction

All the information required to sustain an organism is stored within its DNA, measuring approximately two metres in every nucleus of a human cell. It must therefore be very tightly packaged, yet in such a way that allows it to be easily accessible by enzymes such as transcriptional or DNA repair machinery. Chromatin, describing the complex of DNA with specialized associated proteins, is the basic organizational form of DNA in the eukaryotic nucleus and represents the first level of packaging into higher orders of organization. The functional repeat unit of chromatin is a nucleosome, where 147 base pairs of DNA are wrapped around a histone octamer that consists of two of each core histones: H2A, H2B, H3 and H4 (Figure 1.1). This structure is further held together by the linker histone H1, which interacts with the exterior of the nucleosomal core particle, around the DNA entry and exit sites [1, 2].

A higher-order chromatin arrangement is achieved, which is required to further compact the genome into one of two core chromosomal packaging patterns: heterochromatin or euchromatin (Figure 1.1). Genes located in heterochromatic regions are typically inaccesible to transcriptional machinery and less likely to be expressed, whereas euchromatin describes a more loosely packed structure that is therefore associated with transcriptionally active regions [3]. Differential chromatin packaging thus affects which underlying genes can be expressed, acting as a key regulator of gene expression regulation.

Chromatin is a very dynamic structure that globally alters its packaging state every cell cycle. During interphase, chromatin is found in its least condensed state and is more loosely distributed throughout the nucleus. As cells begin to divide, it remains largely condensed throughout the various stages of mitosis as transcription is globally silenced [4]. Proteins such as cohesin and condensin also contribute to the highly condensed state of chromatin during cell division. For example, cohesin promotes sister chromatid cohesion thereby allowing sister chromatids to separate and move towards opposite spindle poles [5], whereas condensin is thought to directly interact with chromatin and promote its condensation by linking distant segments of a single chromatid [6]. Both cohesin and condensin additionally play important roles during

interphase, as both are required for the spatial organization of the genome by extruding DNA into large loops [7]. Manipulating interactions between different chromatin regions is therefore another means of transcriptional regulation.

Epigenetics can be described as the study of mitotically heritable changes in gene expression that occur without altering the underlying DNA sequence [2]. Studies of epigenetic mechanisms largely focus on modifications added to DNA or histone proteins and the way these are able to influence the overall chromatin structure, resulting in changes in gene expression. As the epigenome essentially dictates a unique gene expression program in each cell type, it defines a cell's functional identity throughout development [8]. A precise coordination and organization of which genes can be expressed at a given time is therefore crucial for ensuring correct cellular function is maintained. As such, aberrant changes in epigenetic regulation are associated with a vast spectrum of consequences that include disorders such as cancer, autoimmune diseases and neurological disorders, overall highlighting the importance of epigenetic regulation throughout an organism's lifetime [9].

1.1.2 DNA Methylation

DNA methylation is one of the best characterized chemical modifications of chromatin. This modification preferentially occurs at sites known as CpG dinucleotides, describing sections of DNA where cytosine is positioned adjacent to a guanine nucleotide. CpG methylation is established by DNA methyltransferase (DNMT) enzymes, which catalyze the covalent addition of a methyl group at position 5 of the pyrimidine ring of cytosine [10]. Different subgroups of DNMTs are known to carry out different functions. For example, DNMT3a and DNMT3b are required to establish *de novo* DNA methylation at CpG sites, whereas DNMT1 is responsible for maintaining methylation patterns by propagating them between cell generations during replication. In mice, targeted mutations in DNMTs results in embryonic lethality, indicating that their dysfunction is incompatible with life [11].

Regions of DNA that are CpG-dense are known as CpG islands (CGIs) [12]. CGIs are highly prevalent in the human genome and are commonly located at gene promoter regions, where they play an essential role in transcriptional regulation. The majority of CGIs are found in an unmethylated state in somatic cells denoting a permissive

transcriptional state. However, CGIs can be hypermethylated, a phenomenon that is associated with compact chromatin that prevents transcription initiation and thereby gene expression. There are two ways that DNA methylation is thought to result in transcriptional repression: first by directly inhibiting transcription factors from binding their cognate DNA sequence, and secondly, by recruiting specialized proteins that directly mediate gene silencing [13]. CGI hypermethylation is a common occurrence at imprinted genes or genes located on the inactive X chromosome in females where constitutive DNA methylation is critical for gene silencing [13].

1.1.3 Histone Modifications

Each of the core histones that constitute a nucleosome has a long N-terminal tail that extends out of the nucleosomal core. These are highly conserved in sequence, owing to their critical role in regulating chromatin structure. Histone tails are subject to a variety of covalent post-translational modifications (PTMs), such as acetylation, methylation, phosphorylation, ubiquitylation and sumoylation [14] (Figure 1.1). Various modifications can lead to several different consequences, where the overall outcome is the ability to control chromatin structure [3]. For example, histone acetylation involves the transfer of an acetyl group from acetyl coenzyme A to lysine residues on histone tails via histone acetyltransferase (HAT) enzymes, a modification that tends to be associated with transcriptional activation. This is partly due to the addition of the acetyl group leading to the removal of the positive charge on the lysine, therefore destabilising the compacted chromatin state, as the positive charge of unmodified histones facilitates interactions with negatively charged DNA and promotes a closed chromatin state [15]. Conversely, deacetylation of lysine residues on histone tails is associated with chromatin compaction and transcriptional inactivation.

Histone methylation can either promote or repress transcription, depending on which residue is modified and how many methyl groups are added. For example, di- or tri-methylation of H3K9 is associated with transcriptional repression [16], whereas methylation of H3K4 is a conserved hallmark of promoters and enhancers and is associated with transcriptional activation [17]. H3K27 tri-methylation is also a repressive histone mark where its action often involves suppressing the expression of developmental genes. This histone mark is laid down by the polycomb repressive

complex 2 (PRC2) and is recognized by the polycomb repressive complex 1 (PRC1), which in turn catalyzes mono-ubiquitylation of lysine 119 of H2A (H2AK119ub), a histone mark that is also required for establishing or maintaining a heterochromatic environment [18]. Of note, PRC1 and PRC2 can also work in the reverse, with PRC1 being attracted first before PRC2. Therefore, aside from directly altering chromatin structure and conformation, histone modifications can also act as docking sites for "reader" enzymes which are able to further recruit transcriptional activators or suppressors. Additionally, as many combinations can result from different modifications added to different histone tails, at different locations, the number of combinatorial patterns and thereby the breadth of downstream effects is quite extensive.

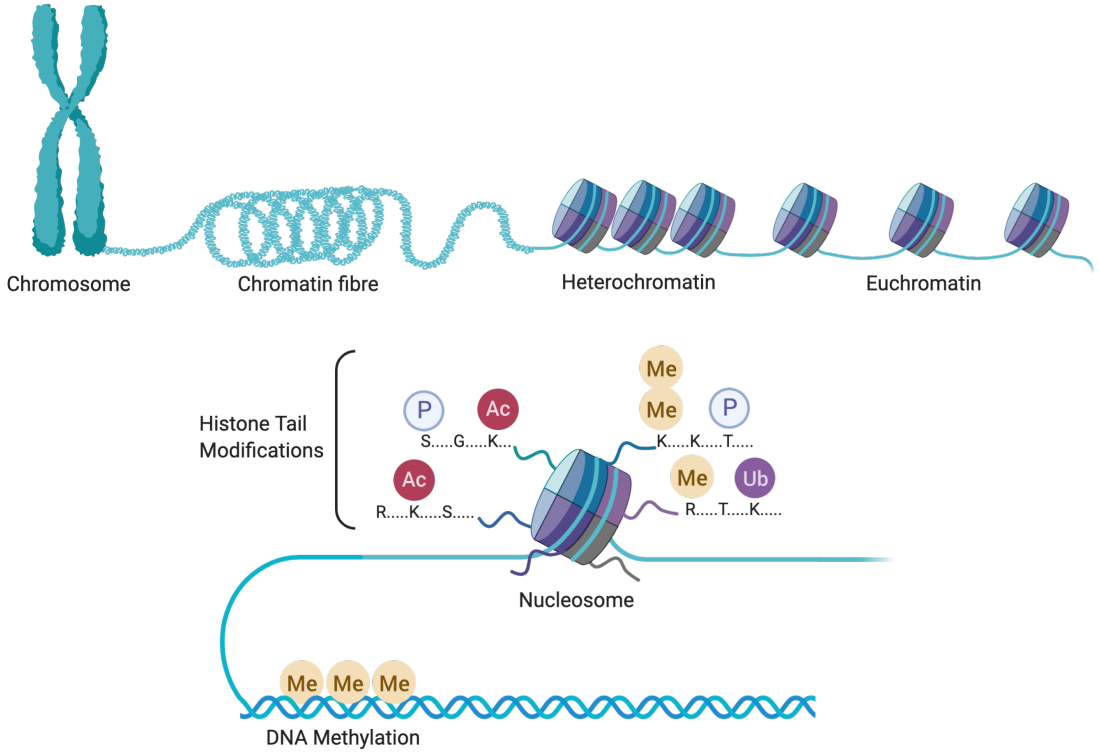


FIGURE 1.1: **Chromatin structure and epigenetic modifications.** DNA is wrapped around histone proteins, forming nucleosomes which are further packaged into chromatin. Chromatin fibers can form condensed structures, termed heterochromatin, or have a more accessible conformation, known as euchromatin. Both DNA as well as the histone tails protruding from nucleosomes are subject to post-translational modifications, including Methylation (Me), Acetylation (Ac), Phosphorylation (P) and Ubiquitylation (Ub).

1.1.4 Chromatin Remodelling

Histone PTMs commonly result in the recruitment of specialized proteins such as chromatin remodelling complexes that can directly contribute to chromatin compaction by manipulating nucleosomal position on DNA. This process is energetically unfavourable as the electrostatic interactions governing DNA-nucleosome interactions are substantial, and therefore it occurs in an ATP-dependent manner. Chromatin remodellers generally form large multi-subunit complexes where accessory subunits are thought to manipulate the ATPase activity of the core subunits [19, 20]. Accessory subunits commonly comprise interaction domains that are able to additionally facilitate interactions with transcription factors and other chromatin modifying enzymes, alongside targeting the complex to DNA or modified histones [20]. The exact mechanism of action of most chromatin remodelling complexes and specifically how they are recruited to precise chromatin sites are not well understood.

The first family of chromatin remodelling complexes identified is the large multi-subunit complex called SWI/SNF, which is highly evolutionarily conserved as homologous proteins have been identified in *Drosophila* flies, and even plants [21]. Additional chromatin remodelling complexes commonly include the INO80/SWR1, ISWI and CHD families. Members of these families are classified in these groups based largely on the sequence homology of their conserved ATPase subunits [22, 23]. Proteins that form the SWI/SNF complex are required for transcription by sequence-specific transcription factors, and the complex was therefore initially proposed to act as a general activator of transcription [21]. *In vivo* experiments suggested that SWI/SNF helps expose promoter regions by pushing the nearby nucleosomes away from each other and destabilizing or ejecting promoter nucleosomes [24]. ISWI and CHD chromatin remodellers are thought to equalize DNA linker length on both sides of a nucleosome and therefore promote uniform intragenic nucleosome spacing, whereas members of the INO80 family are implicated in deposition and removal of the histone variant H2A.Z [25–27]. Chromatin remodellers therefore play a variety of roles in transcriptional regulation, from activation to repression.

1.2 Monoallelic Gene Expression

Monoallelic gene expression occurs for various genes, describing the occurrence where only one of two alleles of a given gene is actively expressed while the other is silenced. Monoallelic gene expression can either be determined by the parent of origin, a phenomenon also known as genomic imprinting, or it can occur randomly [28]. Random autosomal monoallelic gene expression has been recognized for large gene families such as the protocadherins, the olfactory receptor gene family and immunoglobulins [29]. In these cases, monoallelic gene expression is thought to be important for generating cellular diversity and identity, in addition to providing a mechanism for fine-tuning the expression of particular genes.

1.2.1 X-chromosome Inactivation

The most widely-studied example of monoallelic expression is the process of X chromosome inactivation (XCI), which occurs during the developmental stages of most female mammals and is stably maintained throughout an organism's lifetime [30]. XCI ensures a balanced dosage of X-linked genes between XX females and XY males by silencing one of the two X chromosomes in female mammals. An incorrect dosage of X-linked genes can result in human disorders such as Turner (XO) and Klinefelter (XXY) syndromes [30]. There are two different forms of XCI: random XCI and imprinted XCI. Random XCI has no specific preference for which parental X chromosome becomes inactivated, and this takes place in all somatic cell lineages of placental mammals. Conversely, imprinted XCI results in the preferential inactivation of the paternal X chromosome and occurs in female marsupials and mouse placental tissues [30].

The long non-coding RNA (lncRNA), *Xist*, is a critical regulator of XCI [31, 32], with *Xist* deletion in female mice resulting in embryonic lethality during the early post-implantation stage. *Xist* is exclusively expressed by the inactive X-chromosome (Xi) where it initiates XCI by coating the Xi, resulting in a chromosome-wide inactivation of gene expression. This coordinated process involves the silencing of approximately 1000 genes and is completed in the epiblast lineage of the blastocyst which gives rise to all somatic cells [33]. XCI is thereby stably propagated, with the

same X chromosome maintained silenced in successive mitotic divisions. XCI is achieved via several mechanisms, such as by attracting polycomb repressive complex 1 (PRC1) complexes to the Xi which establish H2AK119ub repressive histone marks, exclusion of RNA polymerase II, removal of active histone marks and deposition of repressive marks, and robust DNA methylation. Taken together, these modifications lead to a strong condensation of the chromatin into a perinuclear structure which is also known as the Barr body, ensuring long-term transcriptional silencing [34, 35].

1.2.2 Genomic Imprinting

Both maternal and paternal contributions are required for mammals to develop into viable organisms, owing largely due to the parent-of-origin specific expression of imprinted genes. Genomic imprinting is a form of epigenetic regulation in which the expression of a gene depends on the parent of origin, such that there is unequal expression of the maternal and paternal alleles at an imprinted locus [36]. Therefore, a full complement of imprinted genes with contributions from both parents is essential for correct development to ensue. This phenomenon therefore subjects mammals to a greater genomic risk as a mutation in one allele can result in the complete absence of one or more gene products, which can lead to several imprinting disorders such as Beckwith-Wiedemann syndrome, Silver-Russell syndrome, and Prader-Willi and Angelman syndromes [37, 38].

Over the years, studies in mice and humans have shown that imprinted genes are essential for the prenatal development of normal embryonic and extra-embryonic components, but also for postnatal mechanisms that include behaviour, metabolism, and physiological adaptations [39]. A large selection of imprinted genes undergo tissue-specific imprinting, particularly genes found in the placenta [36]. The placenta is an extra-embryonic organ that is essential for supplying nutrients to the growing fetus. Consistent with this, imprinted genes have been implicated in insulin-like growth factor (IGF) signalling pathways, and as a consequence they have been recognized to play a role in human cancers [40].

The majority of genes that are subject to genomic imprinting are located in clusters of up to 1 Mb in length [36]. Present within each of these is an imprinting control region (ICR) which are regulatory sequences that influence the monoallelic expression of the

full cluster. ICR's exhibit parental allele-specific DNA methylation as a primary imprint signal, but are subsequently decorated with post-translational histone modifications, non-coding RNAs and a represive higher-order chromatin structure [36]. Upon germline deletion of an ICR, loss of imprinted gene expression is observed for the linked genes which further demonstrates the requirement for the ICR in establishing their imprinted expression [41].

1.3 Structural Maintenance of Chromosomes Hinge Domain-containing protein 1 (SMCHD1)

SMCHD1/Smchd1 is a 2005-amino acid chromosomal protein in human and a 2007 amino acid protein in mouse, that has been established as an epigenetic regulator. It was first identified via an *N*-ethyl-*N*-nitrosourea (ENU) transgene mutagenesis screen in mice aiming to find modifiers of gene silencing [42]. Blewitt *et al.* demonstrated that a homozygous mutation of *Smchd1* in female mice resulted in embryonic lethality, causing failed silencing of a subset of genes that are conventionally targets of X inactivation [42, 43]. Smchd1 was then revealed as essential for the maintenance of X-chromosome inactivation and has since been identified as a transcriptional repressor.

More recent studies have shown that Smchd1 also has a role in the silencing of various autosomal genes, as well as in genomic imprinting [44, 45]. Most importantly, heterozygous mutations in SMCHD1 are associated with autosomal dominant facioscapulohumeral muscular dystrophy (FSHD) and the rare craniofacial disorder Bosma arhinia and microphthalmia syndrome (BAMS) [46–48]. However, despite the identification of SMCHD1's many critical roles, how it is recruited to chromatin, its high-resolution structure, and complete mechanism of action remain to be elucidated. Characterisation of SMCHD1's structure and molecular function would therefore provide a tremendous insight, moreover contributing towards drug development aimed at therapeutic treatments for FSHD patients.

1.3.1 SMCHD1 is involved in transcriptional repression

It has been well-established that repeat sequences are subject to repeat-induced gene silencing via epigenetic mechanisms, such as DNA methylation [49]. The *N*-ethyl-*N*-nitrosurea (ENU) mutagenesis screen via which *Smchd1* was first identified was performed on a mouse line that consisted of a multicopy GFP transgene subject to repeat-induced silencing [42, 43] (Figure 1.2). Mice treated with the ENU mutagen can be screened for a phenotype of interest, leading to the mapping and identification of the responsible candidate gene. In this mouse model, an increased expression of the GFP transgene would indicate less effective silencing and therefore the potential targeting of an epigenetic repressor, whereas decreased GFP expression would represent an enhanced repressive ability or failed activation activity (Figure 1.2). The modifier of murine metastable epialleles dominant 1 *(MommeD1)* was a murine line generated via the ENU mutagenesis screen which introduced a nonsense mutation in *Smchd1* that resulted in nonsense-mediated decay of the transcript. These mice presented with a dose-dependent increase in the expression of the GFP transgene, providing the first evidence of Smchd1 acting as an epigenetic repressor [43].

1.3.1.1 SMCHD1 is critical for X-chromosome inactivation

Many studies have revealed that Smchd1 is critical in the process of XCI [42–45, 50–55]. Blewitt *et al.* first demonstrated that Smchd1 is enriched on the inactive X-chromosome (Xi), suggesting a role in XCI maintenance. Whereas *Smchd1*-null male mice were able to survive to adulthood, *Smchd1*-null female mice resulted in embryonic lethality by mid-gestation stage (E11.5) [42, 43]. Further analyses of *Smchd1*-null female embryos revealed that several X-linked genes either escaped XCI shortly after the initiation of silencing, or failed to maintain XCI altogether [43]. This phenomenon is associated with CpG island (CGI) hypomethylation on the Xi which is the most prominent phenotype observed in the absence of Smchd1, suggesting that Smchd1 acts upstream of DNA methylation [44, 45, 50]. Yet exactly how Smchd1 enables transcriptional repression remains unknown.

While Smchd1 has not been detected to directly interact with *Xist*, it was shown that Smchd1 localizes to the Xi in an *Xist*-dependent manner in both humans and mice [51,

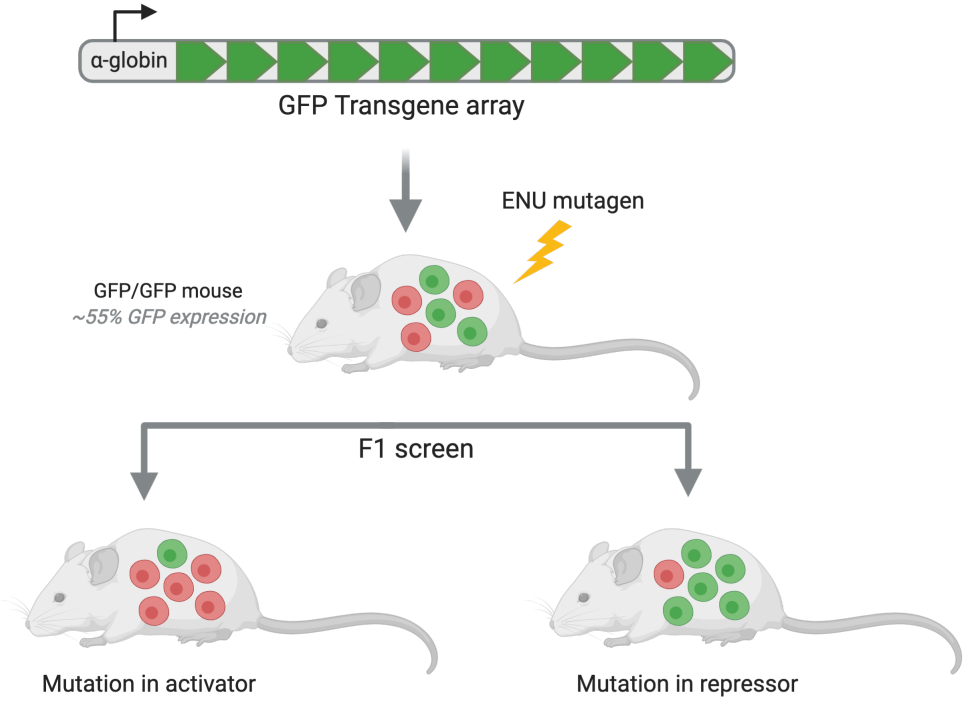


FIGURE 1.2: **Schematic of the ENU mutagenesis screen.** The GFP transgene array is linked to the human α-globin gene promoter and enhancer. Mice that are homozygous for the GFP transgene array have approximately 55% GFP-expressing erythrocytes. Upon ENU-induced mutagenesis, offspring of transgenic, mutant mice can be screened for changes in GFP-expressing erythrocytes. A decrease in GFP expression suggests the mutation targeted a transcriptional activator, whereas an increase in GFP expression indicates the mutation occurred in a transcriptional repressor, required for silencing the GFP transgene.

53, 55, 56]. Smchd1's localisation to the Xi is additionally dependent on the heterogenous nuclear ribonucleoprotein K (HnrnpK), a direct interactor of *Xist*. Jansz *et al.* [55] demonstrated that *HnrnpK*-knockdown in mouse embryonic fibroblasts (MEFs) results in the loss of Smchd1's localization to the Xi in approximately 85% of cells. However, no direct interaction between Smchd1 and HnrnpK was detected, suggesting that Smchd1 may depend on factors downstream of HnrnpK. One of HnrnpK's roles is the recruitment of the non-canonical PRC1 complex that catalyzes H2AK119ub deposition via its catalytic subunit Ring1A/Ring1B. Removal of Ring1A and Ring1B in MEFs resulted in the loss of Smchd1 in 80-95% of female nuclei. Importantly, this effect could be rescued by overexpression of Ring1B, but not a catalytically inactive Ring1B, suggesting that Smchd1's localization to the Xi, as well as global stability, is dependent on PRC1-mediated H2AK119ub-marked chromatin [55].

Nozawa *et al.* reported that SMCHD1 occupies sites on the Xi that are enriched for

H3K27me3 and *XIST*, and that its additional interaction with H3K9me3 contributes to Xi compaction [51]. More recent studies in mouse neural stem cells (NSCs) showed that there was no difference in the mean volume of Xi in wild-type compared to *Smchd1*-deleted NSCs. Chromatin accessibility studies additionally confirmed these findings as no X-linked changes in chromatin accessibility were detected, suggesting that Smchd1 is not involved in maintaining chromatin compaction of the mouse Xi [52]. Instead, Wang *et al.* reported that SMCHD1 interacts with the two chromatin compartments the Xi forms, S1 and S2, and merges them to create a compartment-less architecture, in agreement with the study by Jansz *et al.* [52]. The absence of SMCHD1 therefore leads to disruption of Xi compartments, alongside defects in *Xist* spreading and erosion of heterochromatic silencing [53]. It is therefore proposed that SMCHD1 is necessary for the maintenance of Xi architecture.

1.3.1.2 SMCHD1 regulates the expression of clustered genes

Genome-wide expression analyses of *Smchd1*-null mouse tissue samples revealed that autosomal monoallelic gene expression was disrupted at several imprinted gene clusters, as well as at clustered protocadherins (*Pcdh*) [44, 45]. Pcdh proteins are a group of cell-cell adhesion molecules that are predominantly expressed in the nervous system where they play an essential role in dendrite development and neural circuit formation. *Pcdh* genes are subject to random, combinatorial monoallelic gene expression, resulting in Pcdh diversity that relies on epigenetic regulation of promoter choice and alternative transcripts [57]. Chromatin immunoprecipitation followed by sequencing (ChIP-Seq) and quantitative gene expression analyses demonstrated that Smchd1 down-regulates the expression of *Pcdh* genes by direct interaction with promoters and enhancer regions of the cluster [58]. In the absence of Smchd1, a marked increase in the expression of *Pcdha* and *Pcdhb* clusters was detected [58]. Furthermore, this coincides with CGI hypomethylation at the cluster, suggesting Smchd1 is implicated in DNA methylation in this region. Interestingly, Smchd1 was reported to act in an opposing way to the transcription factor CCCTC-binding factor (Ctcf), occupying the same sites on promoter and regulatory elements controlling the expression of the *Pcdh* cluster where they display antagonistic effects [58].

Loss of Smchd1 results in biallelic expression of genes within the *Snrpn* and *Igf2r*-imprinted clusters, which normally exhibit monoallelic expression [45, 52, 58] (Figure 1.3). Several studies reported that *Peg12, Magel2* and *Ndn* genes within the *Snrpn* cluster exhibit complete loss of imprinted expression without Smchd1 [44, 45, 59]. ChIP-Seq studies additionally demonstrated that Smchd1 directly binds promoter and enhancer regions of these same genes from the *Snrpn* cluster [58], indicating a direct role for Smchd1 in regulating imprinted expression at this locus. While the mechanism by which Smchd1 regulates the expression of its target genes is not well understood, the most prevalent changes that occur in the absence of Smchd1 are DNA hypomethylation, architectural changes in the local chromatin environment and up-regulation of stably silenced genes.

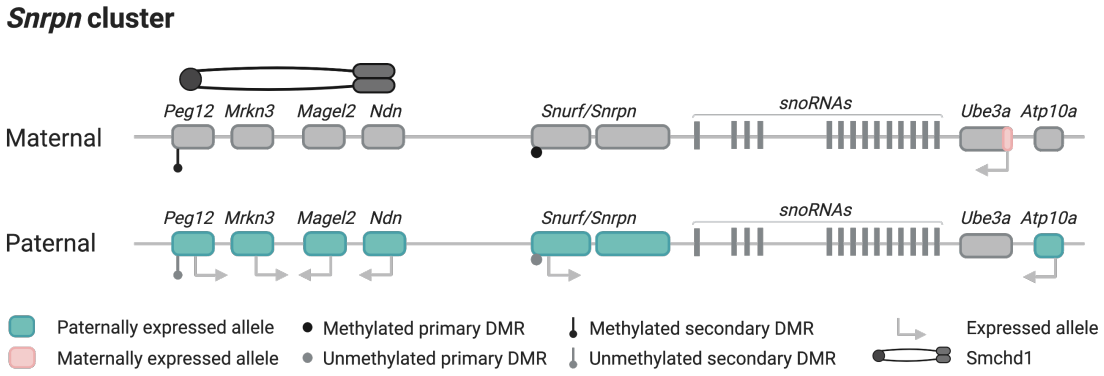


FIGURE 1.3: **A schematic representation of the *Snrpn* cluster of genes.** The maternal (top) and paternal (bottom) alleles of the *Snrpn* cluster are depicted. The maternal allele represents the genomically imprinted and therefore silenced allele, controlled by the methylated primary (germline) and secondary (post-zygotic) differentially methylated regions (DMRs). Here, Smchd1 has a role in silencing the four genes: *Peg12, Mrkn3, Magel2* and *Ndn*. The majority of genes from the paternal *Snrpn* allele are actively expressed, as highlighted in green, with unmethylated primary and secondary DMRs.

Studies investigating changes in chromosome conformation upon *Smchd1*-deletion in female cells revealed genome-wide changes to long-range chromatin interactions at Smchd1 targets, most significantly at the *HoxB* cluster for which Smchd1's role was not previously explored [52]. There are four Hox clusters in vertebrates: HoxA, HoxB, HoxC, and HoxD [60], all of which encode for transcription factors that are major regulators of development. The effects of *Hox*-gene dysregulation observed in *Smchd1*-null male mice were most pronounced in the embryonic skeleton, with varying malformations in the vertebral morphology that are consistent with *Hox*-gene activation. These studies therefore implicate Smchd1 in the developmental regulation

of *Hox*-gene silencing *in vivo*. As absence of Smchd1 correlates with changes in long-range chromatin interactions, it is suggested that Smchd1 mediates *Hox*-gene silencing via higher-order chromatin structure regulation, potentially behaving as an insulator that prevents promoter-enhancer interactions as a means of epigenetic repression.

1.3.2 SMCHD1 has a GHKL-type ATPase domain and an SMC-like Hinge Domain

SMCHD1/Smchd1 encodes a 230-kDa nuclear protein that is conserved across vertebrates. This large protein consists of an N-terminal GHKL-type ATPase domain and a C-terminal SMC hinge domain [61, 62]. The two terminal domains are separated by a long central region that shares no homology with other characterised proteins or functional domains. However, many FSHD2-associated mutations in *SMCHD1* which are known to lead a loss of protein function are located within this middle region [63], highlighting a functional importance that is yet to be determined. Recent studies on SMCHD1's isolated domains have revealed more detailed structural and functional information, which form the foundation of this thesis.

1.3.2.1 SMC family of proteins

The structural maintenance of chromosomes (SMC) family of proteins are considered key organizers of chromatin architecture in all living organisms, and are conserved across all kingdoms of life. In eukaryotic organisms, condensin and cohesin are the two main SMC complexes [64]. As their names suggest, their key roles include chromosome condensation and cohesion, which they are thought to carry out by tethering DNA in a way such that it results in the formation of DNA loops. Nonetheless, cohesin and condensin manipulate distinct cell processes and therefore hold different key roles [7].

Cohesin was first discovered as an essential factor in promoting sister chromatid cohesion and segregation during cell division. Cohesion of two newly-duplicated sister chromatids ensures their correct attachment to microtubules of the spindle by resisting their opposing pulling forces. Removal of cohesin then triggers the equal segregation of sister chromatids to opposite poles in anaphase [65]. It has since become evident that

cohesin also plays a key role in maintaining genomic stability by facilitating DNA repair and in interphase manipulating the 3D organization of chromatin, which is critical for gene expression regulation. Specifically, it mediates the formation of chromosome loops and topologically associated domains (TADs), which are described as chromatin regions that more frequently interact within themselves than amongst each other [7, 66]. TADs and their boundary regions are critical for correct gene expression, as their disruption has been shown to result in disease [66]. Defects in human cohesin and its regulators can additionally lead to genetic developmental disorders, including Cornelia de Lange syndrome and Roberts syndrome, and have furthermore been implicated in various cancers [67].

Condensins promote cell-wide chromatin condensation in preparation for chromosome segregation during mitosis, which is essential for the accurate genetic transmission to each daughter cell [68, 69]. Impaired condensin function has been shown to result in entangled chromosome arms which have failed to separate [70, 71]. Most eukaryotes possess two condensins, condensin I and II, which comprise of the same SMC core subunits but different sets of non-SMC subunits [68, 72]. Despite their structural and functional similarities, the way condensin I and II associate with chromosomes differ [73]. Condensin I is largely cytoplasmic during interphase and becomes chromosome-bound from prometaphase until telophase, whereas condensin II is nuclear during interphase and enriched on chromosomes from prophase until telophase [6].

In terms of protein architecture, cohesin consists of subunits SMC1 and SMC3, whereas condensin is composed of SMC2 and SMC4 (Figure 1.4). In either complex, each SMC subunit presents as an elongated rod-shaped protein of approximately 50 nm in length which consists of a centrally-located hinge domain that serves as the interface for both homo- and heterodimerization of SMC subunits, as well as a DNA interaction site [74, 75]. The hinge domain is flanked by long coiled-coil structures, in addition to a Walker A and B motif on either side consisting of phosphate-binding loops that are required for ATP-binding (Figure 1.4). Each SMC protein brings together its N- and C-termini to form a functionally active ABC-type ATPase domain, also known as an SMC head. This is followed by heterodimerization of SMC proteins at the hinge domain, resulting in the formation of a tripartite ring-like structure in the presence of other accessory proteins, such as kleisin, which links the ATPase heads of the cohesin complex [74, 75].

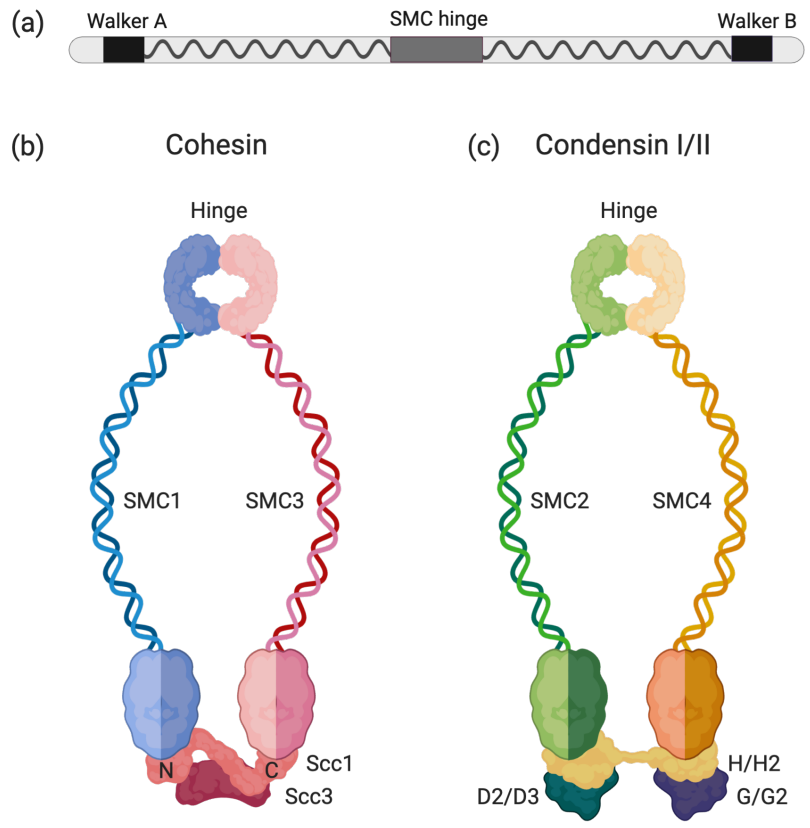


FIGURE 1.4: **Cohesin and condensin form ring-like structures.** Showing (a) a schematic of the canonical SMC protein architecture, with a central SMC hinge domain flanked by coiled coils and a Walker A and B motif on either side. Diagrams representing (b) cohesin and (c) condensin protein complexes, highlighting the long coiled-coils forming the SMC arms of the ring-like structures. (b-c) The hinge domain serves as the heterodimerization as well as DNA-binding site for SMC proteins, whereas the ATPase 'head' domains at the opposing end drive conformational changes of the SMC complexes. The head domains also interact with accessory proteins, such as (b) Scc1 in cohesin and (c) H/H2 in condensin I/II which are critical for closing the ring structure, in addition to promoting ATPase activity.

The presence of architectural similarities between different SMC complexes has suggested they may function via a similar, conserved mode of action. However, the relationship between SMC protein structure and their function is not yet fully understood. Most proposed models agree that the presence of long coiled-coil regions indicate the complexes are able to topologically embrace DNA, whereby for cohesin, for example, two sister chromatids become topologically entrapped within a single cohesin ring [76, 77]. The idea of topological entrapment can be envisioned as a 'ring on a string' model where DNA entrapment by SMC complexes is dependent on the structural integrity of their ring-like arrangement. Topological entrapment is a widely-accepted functional model for the initial loading onto DNA of SMC complexes, as they transition from a V- to a ring-like configuration following entrapment of either one or two DNA strands [74, 75].

Exactly how SMC complexes are able to progressively compact DNA has remained elusive until very recently. The 'loop extrusion' model became one of the most widely-accepted theories describing their ability to translocate along DNA, which relies on an intrinsic motor activity that allows DNA to pass through the central SMC ring [78]. Such an interaction between SMC complexes and DNA has been characterized as pseudo-topological or non-topological and is thought to describe the initial contact of SMC complexes with a promoter or enhancer region on DNA. This is followed by ATPase activity which drives loop elongation through the central ring, such that at least one or both DNA interaction sites translocate away from the other site (Figure 1.5). The loop extrusion model was recently confirmed for condensin using purified complexes from *Saccharomyces cerevisiae* via *in vitro* single-molecule imaging experiments. In addition to demonstrating that condensin is able to translocate along DNA and generate loops in an ATP-dependent manner [79–81], Ganji *et al.* also established that condensin-induced loop extrusion occurs in an asymmetric fashion as the complex anchors itself onto DNA and reels it in from one side [80]. This model may therefore elucidate how condensin is able to compact DNA in mitotic chromosomes in order to enable faithful chromosome segregation.

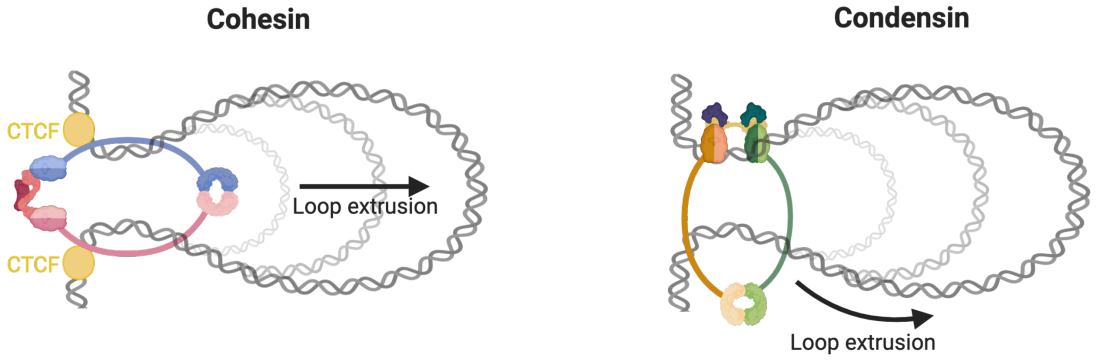


FIGURE 1.5: **The predicted loop extrusion mechanism of cohesin and condensin complexes.** (Left) DNA loop extrusion by cohesin is thought to occur bidirectionally as a DNA strand is extruded symmetrically from either side. CTCF and cohesin are known to co-localize, where CTCF acts as a barrier to the loop extrusion process by stabilizing cohesin at CTCF-binding sites and therefore creating the base of the established DNA loop. (Right) Condensin-induced loop extrusion was shown to occur asymmetrically as the complex is thought to anchor onto DNA via its 'head' domains, reeling DNA in from one side only (as indicated by arrow). In both cases, DNA loop extrusion occurs in an ATP-dependent manner.

While the exact molecular mechanism of function of condensins remains unclear, recently obtained structural evidence for yeast condensin has implied a model for its ATP-binding cycle. It illustrates that the binding of ATP to the Smc4 subunit induces

conformational changes that weaken and release its interaction with one of its accessory subunits, allowing Smc4 engagement with Smc2 via the bound ATP. Smc2-Smc4 dimerization induces further conformational changes that lead to the binding of the second ATP molecule and, simultaneously, to re-arrangements of the coiled-coils that open the condensin ring [82]. A DNA loop can hence be formed by the capture of DNA regions which would become the initial base of the loop within the tripartite ring-like structure of the condensin complex, expanding with the repeat of each ATPase cycle [83]. On the contrary, cohesin was previously reported to slide along DNA in an ATP-independent manner, relying solely on passive diffusion [84–86]. Most recently, however, two separate studies have indicated that human cohesin likewise holds the ability to generate loops by extrusion in an ATP-dependent manner, but it does so symmetrically in both directions, as opposed to condensin where DNA is reeled in from one side only [87, 88]. Although this loop extrusion mechanism may account for how chromosome organisation is achieved during interphase, the role it may play in cohesion is not yet clear.

Previous studies investigating cohesin function have established that its loading onto DNA depends on conserved cohesin-loader complexes, named Scc2-Scc4 in yeast and Nipbl-Nau2 in mammals [5, 89]. The Scc2 subunit interacts with Scc1, an accessory protein that links the two core SMC proteins at the head domain. Scc2 promotes cohesin's ATPase activity, which is thought to lead to the separation of the head domains and the opening of the ring structure which allows DNA entry. Scc4 does not directly participate in this reaction, however it was shown to stabilize Scc2 *in vivo*, as well as aid in targeting of the cohesin-loader complex to chromatin [90, 91].

Recent *in vitro* assays suggested that cohesin's interaction with DNA depends on the ATP-binding event rather than ATP hydrolysis. It was shown that the binding of non-hydrolyzable ATP-analogs, which faithfully mimic the ATP-bound state, is sufficient to initiate DNA loading without hydrolysis [92, 93]. The ATP-binding event is likely to therefore lead to conformational changes in the SMC subunits that facilitates DNA entry in the ring. A crystal structure of the yeast Scc3–Scc1 cohesin complex additionally revealed the complex undergoes direct DNA-binding via surface residues (Figure 1.6), suggesting the SMC head domains are also involved in DNA association [94].

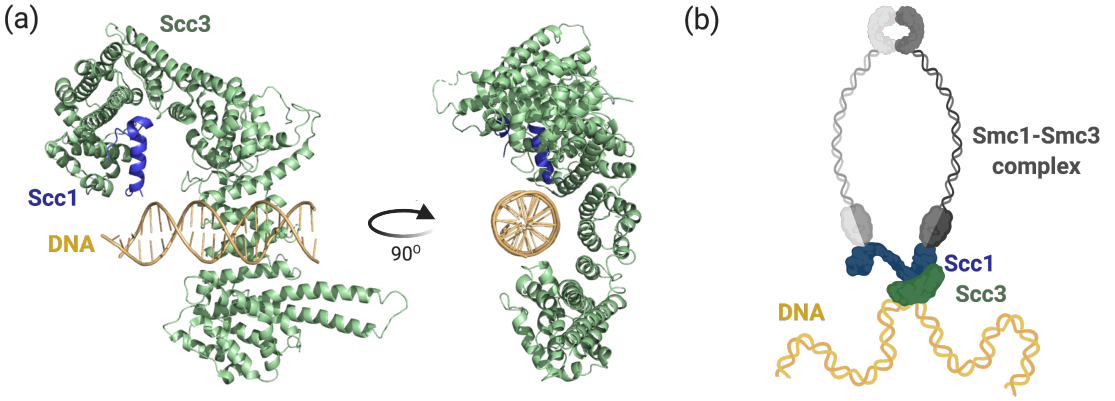


FIGURE 1.6: **Structure of the yeast Scc3-Scc1 cohesin complex bound to DNA.** (a) A cartoon representation of the yeast Scc3 (green) bound to a yeast Scc1 fragment (blue), in complex with a 19 base pair (bp) double-stranded DNA substrate (yellow). An image of a 90 degree rotation around the y-axis of the complex is shown on the right, highlighting the position of the Scc3-Scc1 complex around DNA (PDB: 6H8Q). (b) A model depicting Scc3-mediated DNA loading by cohesin complexes, where the Scc3-Scc1 DNA-binding interface recruits cohesin to target sites.

In humans, the association of cohesin with chromatin is known to additionally depend on CTCF [95–97]. CTCF plays a crucial role in transcriptional activation and repression as well as promoter and enhancer insulation, contributing to the establishment of a higher-order genome structure by defining TAD boundaries. CTCF has been shown to form a cohesin retention side as it physically restrict cohesin movement, acting as a boundary to further loop extrusion and therefore creating the base of the DNA loop established by cohesin [84, 97, 98]. CTCF-bound sites thereby co-localize with cohesin complexes, where both are required for the formation and maintenance of TAD regions.

1.3.2.2 GHKL ATPases

Gyrase, Hsp90, histidine kinase, MutL (GHKL) ATPases describe a large superfamily that encompasses members such as the molecular chaperone heat shock protein 90 (Hsp90), DNA mismatch repair proteins of the MutL family, DNA topoisomerases and members of the Microrchidia (MORC) family of proteins. All members of this diverse superfamily consist of a conserved Bergerat ATP-binding fold which is essential for their ATPase activity [99]. First described by Bergerat *et al.*, this fold comprises an α/β sandwich where three α-helices form a layer parallel to a four β-stranded antiparallel sheet (Figure 1.7). The lining of the ATP-binding pocket is formed by the three α-helices which are connected by a long flexible loop, termed the ATP-lid. This structure is the most diverse among members of the GHKL family, varying in both

composition and conformation, yet indispensable for function as it regulates ATPase activity by closing over the nucleotide-binding site [99].

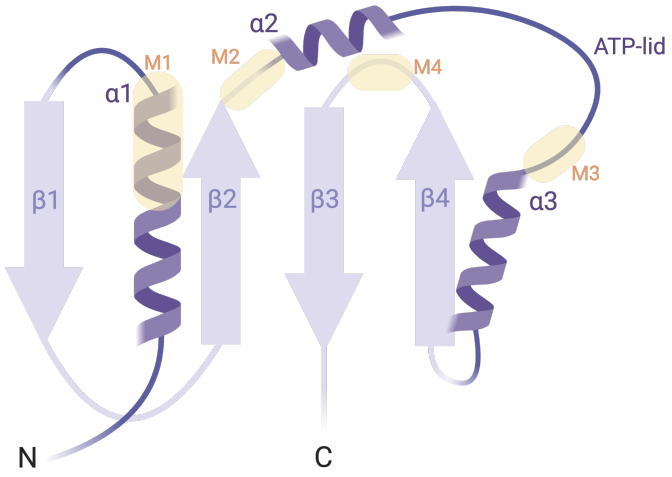


FIGURE 1.7: **Schematic representation of the Bergerat fold.** The Bergerat fold describes an α/β sandwich, where the bottom layer consists of four β strands (β1-β4, light purple) and the top layer comprises three α-helices (α1-α3, dark purple), where α2 and α3 flank the long, flexible ATP-lid. The location of the four conserved motifs are highlighted in yellow and labelled accordingly (M1-M4).

Highly-conserved residues of the GHKL module are largely involved in making direct contact with ATP. These are positioned within the four conserved motifs of the Bergerat fold, which are surprisingly situated within loop regions that connect the secondary structures of the GHKL ATPase, with the exception of Motif I which is located within the first α-helix [99]. In the classical Bergerat fold, various conserved residues situated within Motif I include the conserved glutamic acid (Glu) that activates the water molecule for ATP hydrolysis, and an asparagine (Asn) that is responsible for coordinating a Mg^{2+} ion to the active site, connecting all ATP phosphates to the GHKL entity via solvent-mediated hydrogen bonds [99]. Motifs II and III consist of conserved glycine (Gly) residues that confer ATP-lid flexibility, as they form the hinges on either side of the loop. Motif IV consists of only two residues, a Gly and a threonine (Thr), which contact residues within Motif II, helping maintain the structural integrity of the ATP-binding site [99]. An additional critical conserved site includes a lysine (Lys) or arginine (Arg) that is located within the transducer domain, downstream of the GHKL ATPase [99]. This residue is positioned within a 'switch loop' which extends into the ATP-binding site and becomes positioned adjacent to the γ-phosphate of the bound ATP, demonstrated

as necessary for catalytic activity as it is thought to stabilize the transition state of ATP hydrolysis [100].

The ATP-binding event induces additional intermolecular interactions that can result in the homodimerization of various GHKL ATPases [99]. A domain-swapping event of the N-terminal β-strand from each monomer occurs, which makes additional contacts to the ATP-lid and stabilizes the dimer state. Further conformational changes that take place upon ATP-binding include a rotation of the transducer domain with respect to the GHKL ATPase via a re-arrangement of the switch loop that positions the Arg/Lys into the ATP-binding site [101]. These changes result in an intertwined interaction between two GHKL monomers that are triggered by the ATP-binding event. In some cases, the cavity formed between two GHKL monomers is able to accommodate double-stranded DNA, reported for Gyrase B and MutL GHKLs [102, 103], or accessory proteins in the case of Hsp90 [104, 105]. ATP hydrolysis is then required to release the bound DNA or protein, as this event leads to the opening of the GHKL dimer. The ATPase cycle of GHKLs has therefore been described to drive a molecular clamp mechanism [99].

Hsp90

Hsp90 is a molecular chaperone that plays a vital role in cells and is highly conserved from bacteria to humans [106]. Under stress conditions, its function comprises folding, maturation and activation of unfolded substrate proteins. Hsp90 is a functional dimer whose ATPase cycle consists of large-scale conformational changes that include inter- and intra-protomer interactions. Each monomer consists of three regions: an N-terminal ATP-binding domain, a middle domain, and a C-terminal dimerization domain; all of which are connected by long, charged, flexible linkers that aid conformational rearrangements [106] (Figure 1.8). An N-terminal 'strap' swapping event is conserved across the Hsp90 GHKL, however it is not entirely essential for ATP hydrolysis as deletion of this strap in one Hsp90 subunit was demonstrated to abolish ATPase activity in the opposing monomer only [107, 108]. Nonetheless, this highlights that importance of inter-domain contacts between Hsp90 N-terminal domains.

Hsp90's activity is additionally regulated by the conserved ATP-lid structure that closes over the bound ATP. The Hsp90 ATP-lid is considerably longer compared to other GHKL ATPases, leaving the bound nucleotide completely exposed as the ATP-lid points

away from the ATP pocket [99, 109, 110]. Surprisingly, while deletion of the ATP-lid across both subunits completely abolishes ATPase activity, deletion of this loop in only one subunit dramatically enhances activity in the opposite subunit [111, 112]. Removal of the ATP-lid is thought to block early intra-protomer conformational changes that are required to drive ATP hydrolysis.

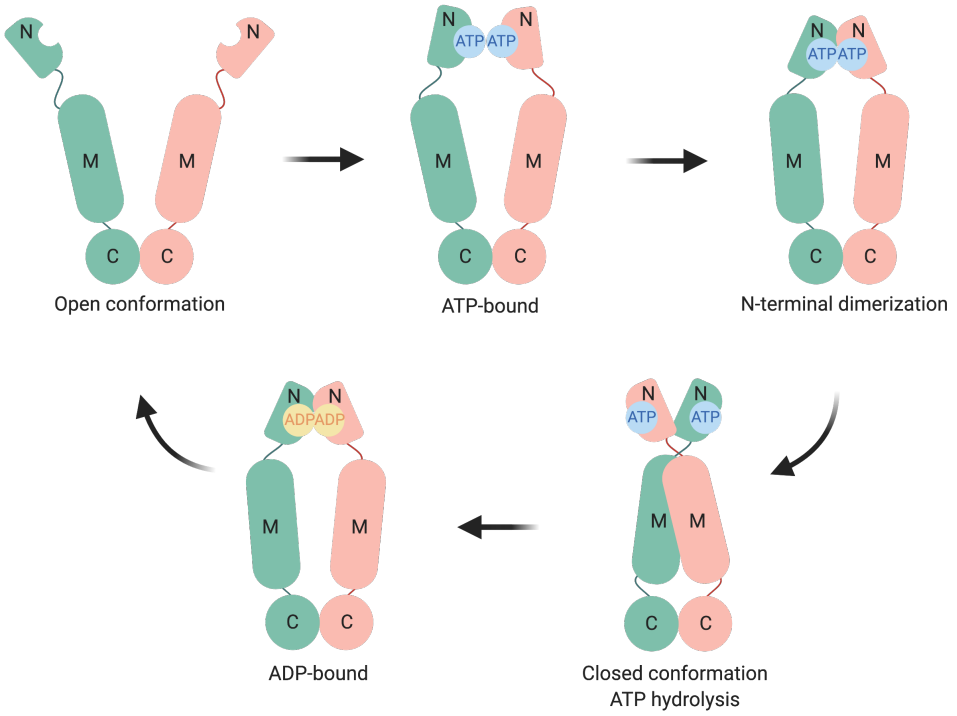


FIGURE 1.8: **The ATPase cycle of Hsp90.** The Hsp90 homodimer (green and pink monomers) depict the three different domains: N-terminal (N), middle (M) and C-terminal (C) domains. The Hsp90 dimer adopts a V-like conformation in the apo state. ATP-binding to the N domain induces a conformational change causing the ATP-lids to close over the bound nucleotide. This further induces the dimerization of the N domains, followed by an N-terminal strap swapping event which results in a closed conformation. Upon ATP hydrolysis, Hsp90 reverts back to a more open, ADP-bound dimer state, followed again by an entirely open, V-like conformation as ADP is released.

The activity of Hsp90 is additionally regulated by several co-chaperone proteins that control its progression through the ATP-dependent substrate activation cycle, likely by regulating the transition between Hsp90's conformational states that control ATP hydrolysis [110, 113, 114]. One of its most potent stimulators is Aha1, a co-chaperone that plays a crucial role in kinase activation and membrane protein folding in mammals [105]. It is suggested that Aha1 stimulates Hsp90 activity by accelerating its progression through different conformational states [112].

MORC family

MORCs are a family of transcriptional regulators that are highly conserved across eukaryotic organisms as they are key players in mechanisms of epigenetic regulation, with emerging roles in chromatin compaction processes [115–118]. In humans, there are four MORC proteins, MORC1-4, all of which share a similar domain architecture. They are large, multi-domain proteins that consist of an N-terminal GHKL domain, a central CW-type zinc finger (CW) domain and a C-terminal region that is thought to be involved in dimerization [115, 119, 120] (Figure 1.9). Their conserved GHKL ATPase has been shown to dimerize upon ATP-binding, contributing to the hypothesis that MORC proteins function as molecular clamps [120, 121]. The way that MORC proteins may interact with chromatin may therefore be analogous to SMC proteins such as cohesin and condensin, which use topological entrapping of DNA as their mechanism of action.

MORC1 is primarily expressed in male germ cells and regulates mammalian germ cell development and meiosis [122]. It was recently demonstrated that *Caenorhabditis elegans* MORC1 is able to form DNA loops in an ATP-independent manner, yet able to be stimulated by ATP or a non-hydrolyzable ATP analogue [123]. Studies additionally reported that MORC1 forms nuclear puncta as the protein was observed to diffuse along DNA and emerge into growing MORC1 foci in cells that can undergo phase-separation *in vitro* [123]. These observations therefore highlight a mechanistic insight into how MORC1 is able to to regulate silencing via chromatin compaction.

MORC2 is ubiquitously expressed in human cells and tissues [124–126]. It contains a chromo-like domain, commonly found in chromatin-associated proteins such as the chromodomain-helicase-DNA-binding (CHD) family members, chromobox protein homolog (CBX) members and chromatin remodelling complexes such as the SWI/SNF complex. Chromo-like domains are known to be involved in epigenetic regulation and gene expression control. MORC2 exhibited ATP-dependent chromatin remodelling activity, hence its function is associated with altering chromatin architecture [126]. In humans, MORC2 is part of the human silencing hub (HUSH) complex where it is implicated in epigenetic silencing at H3K9me3-marked chromatin sites, requiring both ATP-binding and dimerization of MORC2 for faithful silencing [115, 118].

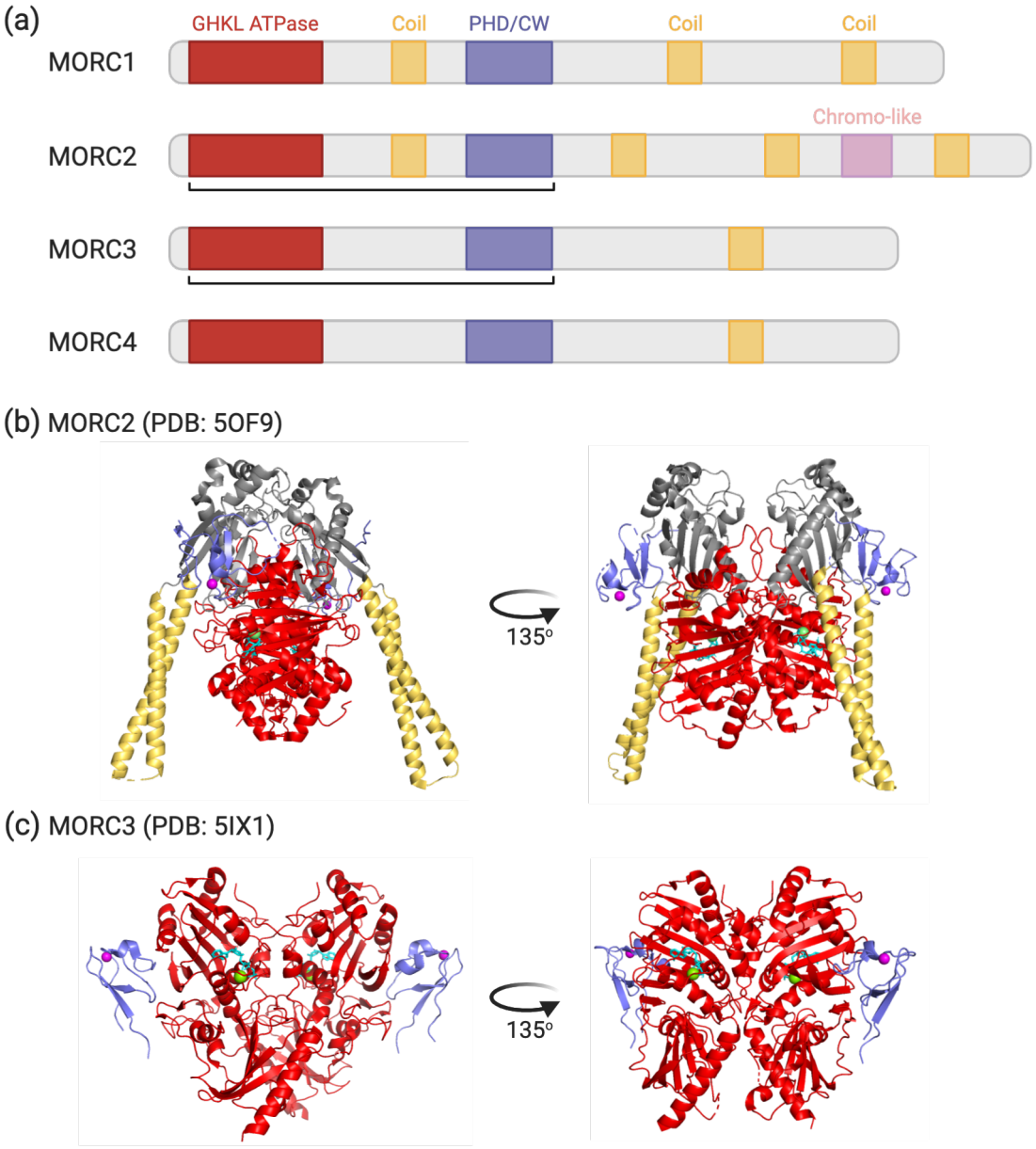


FIGURE 1.9: **Comparison of gene architecture and protein structure between different MORC family members.** (a) All members of the MORC family have a conserved GHKL ATPase domain at their N-terminus (red), a zinc-finger CW-type motif (CW) which is an N-terminally truncated version of the zinc-binding plant homeodomain (PHD) located in the central region (purple), and various coiled-coil domains (yellow). MORC2 additionally contains a chromo-like domain at its C-terminus (pink). Cartoon representation of crystal structures of (b) human MORC2 and (c) human MORC3, encompassing the underlined domains in black from panel (a). (b-c) Ribbon diagrams are represented in the colour scheme shown in panel (a), with the addition of a transducer-like domain (gray) shown for (b) MORC2. Zinc is shown in magenta, AMP-PNP is shown in cyan and Mg^{2+} is shown in green.

Recent structural and biochemical studies revealed that like many other GHKL ATPses, MORC3 also undergoes ATP-dependent dimerization [120]. The crystal structure of a MORC3 construct encompassing the ATPase and CW domains illustrated the binding of a H3K4me3 peptide within an aromatic cage of the CW domain of each protomer,

confirming its ability to localize to H3K4me3-marked chromatin [120]. Interestingly, the CW domains of MORC1 and MORC2 lack the ability to interact with H3K4me3, suggesting that different MORCs may engage with chromatin via different mechanisms. The ATPase domain of MORC3 was also reported to display a DNA-binding ability that is able to enhance its ATPase activity [127]. Similarly to MORC1, MORC3 was also shown to accumulate in nuclear bodies [121, 128] where its histone binding ability plays an essential role in its recruitment to chromatin [127].

Taken together, these studies demonstrate that GHKL ATPases are emerging players in epigenetic regulation, hypothesized to function as molecular clamps on DNA.

1.3.2.3 SMCHD1's GHKL ATPase region

SMCHD1 is considered a member of the GHKL superfamily owing to the presence of an N-terminal GHKL ATPase domain [61]. The first study that examined SMCHD1's ATPase region generated recombinant protein encompassing the predicted GHKL domain alongside the downstream region, which is homologous to the transducer domain in canonical GHKL proteins [61]. Chen *et al.* demonstrated that Smchd1's N-terminal region (residues 111-702) was catalytically active and present as a monomer in solution [61]. Interestingly, this study also revealed via small-angle X-ray scattering (SAXS) experiments that this construct adopts a conformation that is highly similar to that of full-length Hsp90, despite the proteins sharing only approximately 8% sequence identity [61] (Figure 1.10). The catalytic rate of Smchd1's N-terminal region was measured to be approximately 0.6 ADP/minute [61], making it a weak ATPase, yet comparable to the rate observed for Hsp90 [129, 130]. Interestingly, Like Hsp90, Smchd1's ATPase activity can be antagonized by the compound Radicicol [61, 131], which exhibits an IC50 value in the nanomolar range which is comparable to that observed for Hsp90 [61, 132]. Structural homology to Hsp90 may therefore suggest the extended Smchd1 ATPase could interact with additional proteins that may regulate its activity, which is a key feature of Hsp90.

The first crystal structure of a SMCHD1 ATPase construct (residues 25-580) was recently solved and it revealed the conserved Bergerat fold as its catalytic domain (residues 110-395) [133]. Pedersen *et al.* also identified a novel ubiquitin-like (UBL) domain at the N-terminus (residues 25-110) that undergoes a domain-swapping event

between two SMCHD1 monomers via an N-terminal β-strand (residues 110-120) [133] (Figure 1.10 a). The group was the first to describe a dimerization event of SMCHD1's ATPase domain, which is not an unexpected finding considering GHKL ATPases universally sustain the ability to homodimerize [99]. The group showed that SMCHD1 dimerization requires not only the UBL domain, but also the presence of ATP, in addition to the downstream transducer domain. Surprisingly, the solved structure does not reveal any interface contacts between SMCHD1 monomers at the transducer domain as dimerization appears to instead create a cavity between the transducer domains of two monomers. An explanation for the inability of SMCHD1 to dimerize upon absence of the transducer domains was not postulated. It would therefore be interesting to further investigate whether extending the SMCHD1 ATPase construct at the C-terminus may reveal contacts between the transducer domains, as these commonly serve as the sites of homodimerization across members of the GHKL family [99].

Despite providing highly valuable insights into the molecular structure of SMCHD1, it is important to note that the solved structure of the ATPase construct represents a catalytically inactive point mutant form. The E147A mutation introduced has previously been reported and shown to completely abolish the ATPase activity of Smchd1 [61], as the Glu residue is conserved across members of the GHKL superfamily and is indispensable for the ATP hydrolysis step [99]. As this crystal structure was solved in the presence of ATP, it is presumed the E147A mutant retains the ability to bind ATP but fails to hydrolyze it. Interestingly, dimerization triggered by ATP-binding is a common feature among GHKL ATPases [115, 119, 123, 134]. This is closely followed by the closing of the ATP-lid over the active site and the ATP hydrolysis event that leads to the dissociation of the dimer. If the E147A mutant is trapped in the ATP-bound state, this phenomenon likely justifies its preferential dimerization over wild-type SMCHD1. This idea is further supported by native PAGE analyses where only the E147A variant of SMCHD1 appears to migrate in a dimeric form under native conditions, while the wild-type or any other SMCHD1 variant tested remain largely monomeric [133]. Such a discrepancy raises the value of a wild-type SMCHD1 atomic structure and prompts more in-depth biophysical experiments to establish parameters such as dimerization affinity for both wild-type as well as the catalytically inactive E147A mutant.

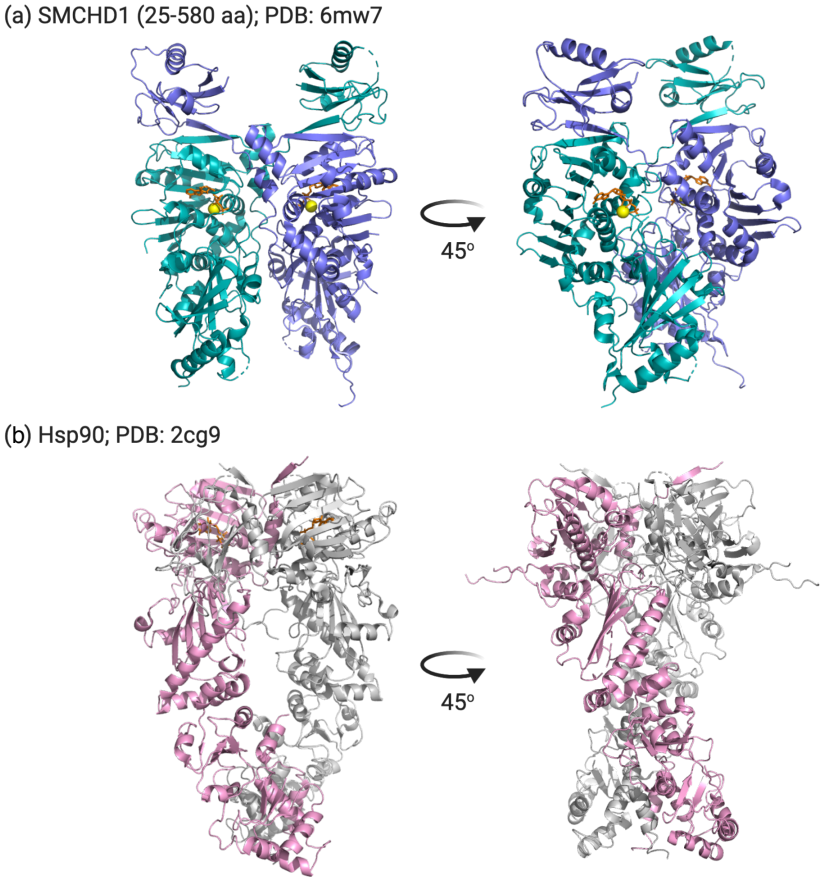


FIGURE 1.10: **Comparing the crystal structures of SMCHD1's ATPase region and full-length Hsp90.** Cartoon representations of (a) the N-terminal ATPase region of SMCHD1 (teal and purple monomers) and (b) full-length yeast Hsp90 (pink and grey monomers). (a-b) ATP is shown in orange in stick form and an Mg^{2+} ion is depicted as a yellow sphere for (a) SMCHD1. Images of a 45 degree rotation around the y-axis are shown on the right. These highlight (a) the N-terminal domain swap occurring in the SMCHD1 homodimer and (b) the compact, twisted structure formed by the Hsp90 homodimer where an N-terminal β-strand swapping event also occurs.

1.3.2.4 SMCHD1's SMC hinge domain

SMCHD1 is considered a member of the SMC family owing to presence of an SMC hinge domain [62]. As opposed to canonical SMC proteins, SMCHD1's hinge domain is located at the C-terminus of the protomer rather than centrally (Figure 1.11). Structural and biochemical studies including SAXS and analytical ultracentrifugation (AUC) analyses revealed that Smchd1's hinge domain likewise forms a homodimer [58, 62]. These SAXS models suggested that the globular central hinge domain is flanked by two stalks of different lengths emerging from either side [62]. This is consistent with the presence of two intermolecular parallel coiled coils surrounding the Smchd1 hinge domain, consisting of N- or C-terminal α-helices. Such a configuration differs from the long, intra-molecular

coils that flank the hinge domain of canonical SMC proteins (Figure 1.11). Overall, these data suggests that the full-length Smchd1 protein may assemble into a head-to-head parallel dimer without adopting a hairpin shape at the hinge domain, which differs from the dimeric arrangement that is conventionally found in all members of the SMC protein family characterized to date. Further evidence for this model is provided by previously reported negative stain electron microscopy images where full-length Smchd1 appears to form extended dumbbell-like rods with globular domains observed at either end of the Smchd1 dimer [131], likely corresponding to the ATPase and SMC-hinge domains.

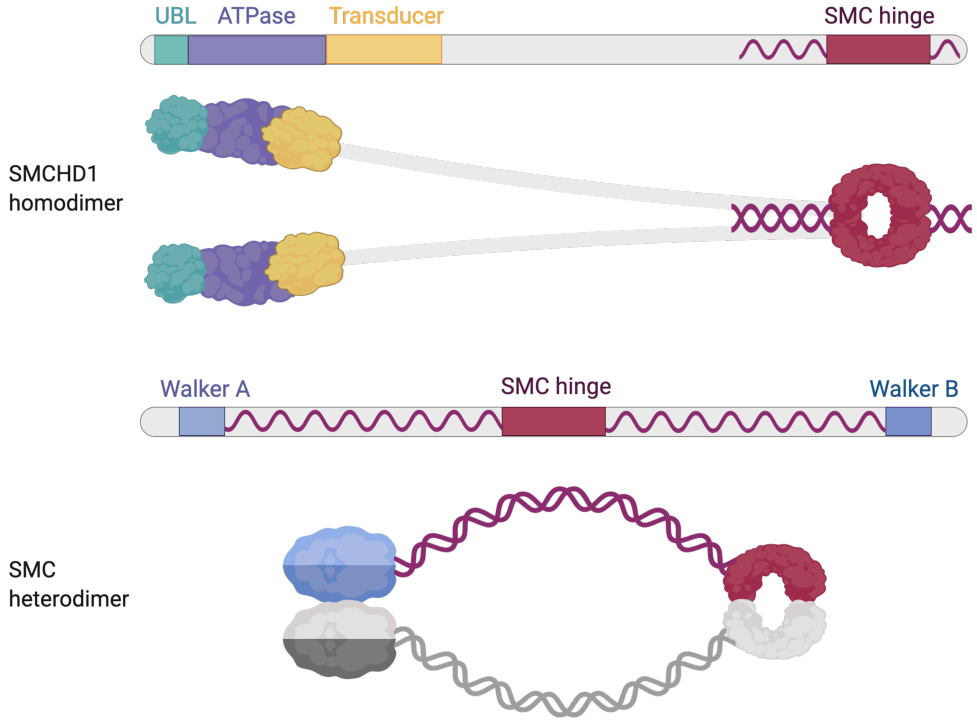


FIGURE 1.11: **Comparison of canonical SMC proteins and the proposed arrangement of full-length SMCHD1.** (Top) Schematic representation of the proposed full-length SMCHD1 arrangement, comprising of an N-terminal ATPase region that consists of a UBL domain (green), a GHKL-type ATPase (purple) and a transducer domain (yellow). The SMC hinge domain (deep red) is located at the C-terminus and is connected to the N-terminus by a long uncharacterized 'middle' region (grey). Two SMCHD1 protomers are thought to dimerize in a parallel fashion, where the hinge domain is flanked by short inter-molecular coiled-coils. (Bottom) Schematic representation of a full-length canonical SMC protein, highlighting the centrally located SMC hinge domain (deep red) that mediates heterodimerization of SMC proteins and has a role in DNA-binding. The flanking long intra-molecular coiled-coils fold in an anti-parallel manner, connecting the terminal Walker A and B motifs to form a functionally active ABC-type ATPase domain as it folds upon itself. Upon heterodimerization with another SMC protein, the SMC complex forms a ring-like structure.

Chen *et al.* showed that in addition to providing an interface for Smchd1 homodimerization, the hinge domain also exhibits DNA-binding activity [58]. *In vitro* DNA-binding assays as well as AUC experiments demonstrated that Smchd1's hinge domain displays the highest affinity for single stranded DNA, and in some instances methylated DNA. Chen *et al.* also reported that Smchd1 has the potential to bind RNA [58], suggesting the protein may interact with RNA molecules as part of its epigenetic function, yet without any sequence specificity *in vivo* as demonstrated recently by Jansz *et al.* [55]. The additional investigation of an FSHD2-compromised mutation located within Smchd1's hinge domain, R1867G, revealed the mutant's DNA-binding ability was substantially compromised [58], providing a likely explanation for the loss of SMCHD1 function exhibited by the patient. Taken together, these findings raise the possibility that Smchd1's hinge domain is a likely candidate for recruiting the full-length protein to chromatin binding sites.

1.3.3 Mutations in *SMCHD1* lead to human disease

1.3.3.1 Facioscapulohumeral Muscular Dystrophy (FSHD)

FSHD is the third most common form of muscular dystrophy, with a prevalence of approximately 1 in 8000 people. The condition is specifically characterized by facial, shoulder and upper arm muscle weakness [135] and is separated into two disease types: FSHD type 1 (FSHD1) and type 2 (FSHD2), with FSHD1 being the more common form. In both cases, FSHD occurs due to the de-repression of a macrorepeat array, *D4Z4*, located in the subtelomeric region of chromosome 4q. Healthy individuals have between 11 and 100 *D4Z4* repeats, with each repeat unit of approximately 3.3 kb in size [136]. In FSHD1 patients, this array of repeats is contracted to 1-10 units, which results in a more permissive local chromatin structure [137]. In the presence of a polymorphic polyadenylation signal (PAS) on the last *D4Z4* repeat, a stabilized *DUX4* mRNA can be produced which is followed by the ectopic expression of DUX4 in skeletal muscle (Figure 1.12). DUX4 is a transcription factor that is normally expressed in germ cells in the testis and early pre-implantation embryos, but is epigenetically silenced in somatic cells [138, 139]. Its specific function and exactly how it confers muscle toxicity is not well-understood. However, it is thought that its aberrant expression triggers the activation of signalling pathways that are solely required during embryonic development [138]. This

ultimately leads to skeletal muscle cell death and consequently the phenotypic changes and symptoms common to patients suffering of either FSHD1 or FSHD2, as ectopic DUX4 expression appears to underlie both FSHD types [140].

Conversely, FSHD2 patients do not have a contracted *D4Z4* repeat array, yet approximately 80% of them harbour heterozygous *SMCHD1* mutations, in addition to a *D4Z4* allele with a permissive polyadenylation signal (4qA) on chromosome 4 [141]. Therefore, FSHD2 follows a digenic inheritance pattern that requires both the 4qA allele as well as the presence of a *SMCHD1* mutation [48] (Figure 1.12). SMCHD1 has been shown to directly bind and maintain a repressed chromatin structure at *D4Z4* repeats [45, 137]. In FSHD2 patients, reduced functional SMCHD1 levels results in the de-repression of the local chromatin structure at the *D4Z4* repeat array, leading to the downstream *DUX4* mRNA expression. A 30-40% reduction in DNA methylation levels within *D4Z4* CGI's have also been reported in FSHD patients [142], alongside a loss of repressive histone 3 lysine 9 tri-methylation (H3K9me3) marks [143, 144]. These results are consistent with SMCHD1's repressive epigenetic role at the *D4Z4* locus. Further studies have identified *SMCHD1* mutations in FSHD1 patients, where these resulted in an augmented disease severity. These findings demonstrate that mutations in *SMCHD1* are not exclusively found in FSHD2 patients and they may also act as a disease modifier in FSHD1 patients [145].

FSHD2-associated mutations in *SMCHD1* are loss-of-function mutations that range from deletions and frameshift mutations, to chromosomal translocations and deletions. These span the entire length of the 2005-amino acid protein, but are least prevalent in the DNA-binding C-terminal hinge domain [63]. Nonsense mutations are thought to lead to haplo-insufficiency and therefore less overall presence of functional SMCHD1 dimers. On the other hand, because SMCHD1 forms a homodimer, missense mutations could exert a dominant negative effect in the dimeric protein, with the presence of a mutant SMCHD1 subunit in an estimated three-quarters of the dimer population resulting in malfunctioning dimers [137]. We have recently shown via *in vitro* studies utilizing the extended ATPase domain of Smchd1 that in the presence of FSHD2-associated missense mutations in Smchd1, a decreased ATPase function may contribute to its overall loss of function as an epigenetic repressor [146]. This highlights the possibility of therapeutically targeting the ATPase region in an attempt to boost the activity of the functional SMCHD1 subunits in FSHD2 patients.

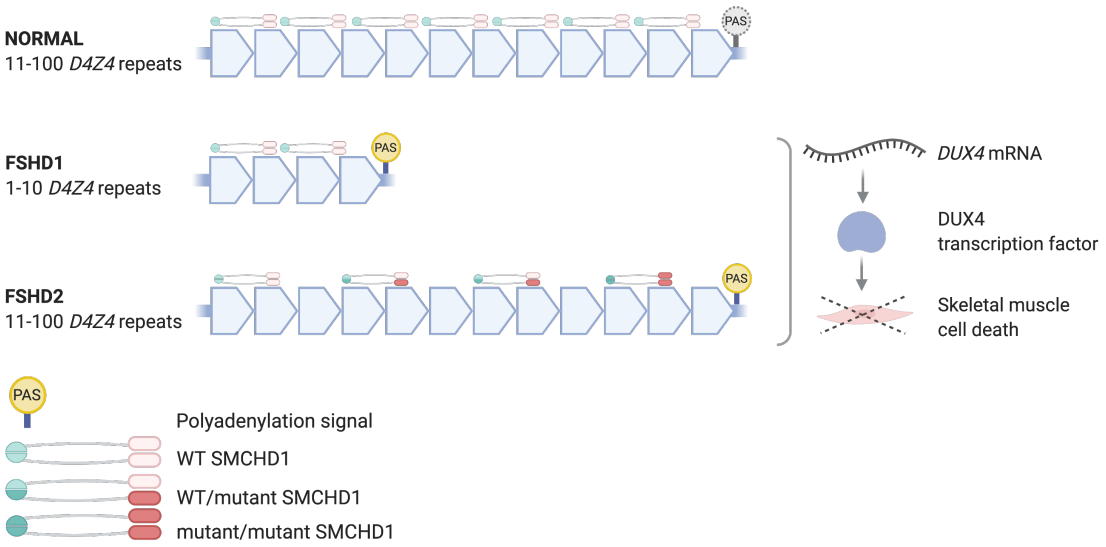


FIGURE 1.12: **Molecular basis of FSHD.** Healthy individuals have between 11-100 *D4Z4* repeats, alongside functional SMCHD1 protein that directly binds and represses these. FSHD1 patients have a contracted *D4Z4* array, whereas FSHD2 patients have a normal *D4Z4* array yet they harbour heterozygous mutations in *SMCHD1*. In both cases, this results in the de-repression of the D4Z4 locus, and in the presence of a polyadenylation signal (PAS) which is permissive of *DUX4* transcription, this ultimately leads to skeletal muscle cell death. Healthy individuals can also harbour a PAS on the last *D4Z4* repeat, hence depicted in gray. The SMC hinge and ATPase domains of the dimeric SMCHD1 protein are depicted in shades of green and pink, respectively, where the light shades indicate a wild-type (WT) copy of SMCHD1, and the dark shades indicate a SMCHD1 protein that harbours a mutation. Because FSHD2-associated mutations target only one *SMCHD1* allele, this can result in three possible forms of SMCHD1 heterodimers: WT/WT SMCHD1, WT/mutant SMCHD1 or mutant/mutant SMCHD1. This consequentially leads to less overall functional SMCHD1 protein dimers in FSHD2 patients, highlighted by a decreasing presence of SMCHD1 bound to the *D4Z4* repeats in FSHD2 patients.

1.3.3.2 Bosma Arhinia and Micropthalmia Syndrome (BAMS)

Bosma Arhinia Microphthalmia Syndrome (BAMS) is a very rare congenital condition, with fewer than 100 patients identified to date worldwide. The condition is primarily characterized by an absence of the nasal area, and is often accompanied by ocular defects [147]. Affected newborns therefore suffer from severe respiratory and feeding difficulties that typically require reconstructive surgery to allow for their normal development. BAMS is thought to arise from a developmental defect during the early embryonic stages, affecting the nasal placodes or surrounding neural crest-derived tissues [147, 148]. *De novo* heterozygous mutations in *SMCHD1* were recently confirmed in studied patient samples, suggesting a potential new role of SMCHD1 in craniofacial development [46, 47]. Interestingly, all identified mutations are missense and they all reside in the extended ATPase domain of SMCHD1 (Figure 1.13) [46, 47]. This indicates that BAMS-associated mutations may affect SMCHD1 function by potentially altering its catalytic

activity. Our studies have shown that several patient SMCHD1 mutants lead to a gain of ATPase function *in vitro*, reflected also in *Xenopus laevis* embryo models where the same mutations lead to ocular defects that partially recapitulate the phenotypic changes observed in human patients [46, 146]. Further supporting a gain-of-function model, treatment of *Xenopus* embryos with increasing amounts of wild-type SMCHD1 similarly resulted in a small-eye phenotype [46].

The rare occurrence of both loss- and gain-of-function mutations in the same protein leading to two distinct conditions, BAMS and FSHD, has triggered a further interest in uncovering SMCHD1's molecular mechanism of function. While both conditions share *SMCHD1* mutations as an underlying factor, the idea that FSHD and BAMS arise from distinct defects in SMCHD1 function has become fairly controversial. A study by Shaw *et al.* reported that either knockdown or knockout of *smchd1* in zebrafish resulted in a small-eye phenotype, suggesting that SMCHD1 loss-of-function rather than gain-of-function may instead underlie the pathogenesis of BAMS [47]. However, they were unable to reproduce this phenotypic effect in mouse models upon the introduction of BAMS-related mutations in *SMCHD1*. To further validate our gain-of-function model, we showed via studies in *Xenopus* embryos that upon mRNA injection of FSHD2-associated *SMCHD1* mutations, no phenotypic changes are observed [46].

Only two *SMCHD1* mutations common to both FSHD2 and BAMS patients have been reported to date, G137E and L107P. Upon examination of the patient carrying the L107P mutation, it was revealed their muscle phenotype was atypical of FSHD [149], whereas studies in the *Xenopus* model revealed that the G137E mutant resulted in a decreased eye diameter, alongside a gain of ATPase activity in our *in vitro* biochemical assay [146]. This provides further evidence that the G137E and L107P mutations in *SMCHD1* are likely an underlying cause of BAMS and not FSHD2, and that the patients may instead suffer from a phenotypically similar muscular dystrophy, such as limb girdle muscular dystrophy, as the two conditions are often misdiagnosed for each other [150]. Crucially, the potential gain-of-function phenotype observed in BAMS-associated SMCHD1 mutants reveals the possibility of specifically targeting active or allosteric sites in the extended ATPase domain to activate the one wild-type copy of the two SMCHD1 alleles in FSHD2 patients, and boosting SMCHD1 function in FSHD1 patients even without SMCHD1 mutations. Of particular interest are

small-molecule compounds that can increase SMCHD1's catalytic efficiency in individuals who are at high risk or in the early stages of developing FSHD2 in anticipation of preventing further muscle cell death.

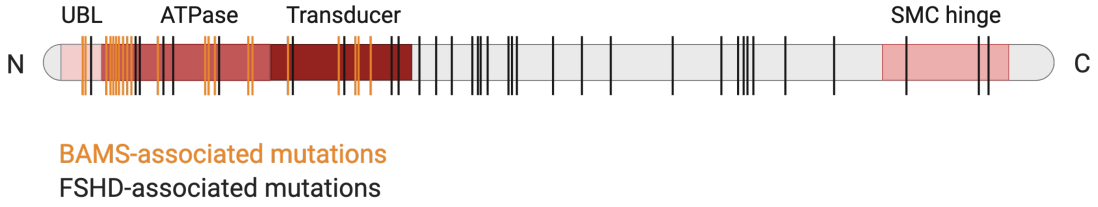


FIGURE 1.13: **Comparison of FSHD2- and BAMS-associated *SMCHD1* mutations.** Schematic of the full-length SMCHD1 gene architecture, highlighting the different functional domains in various shades of red. FSHD2-associated mutations in *SMCHD1* are depicted in black lines and are found across the full length of the protein, whereas BAMS-associated mutations in *SMCHD1* are highlighted in orange lines and are found exclusively in the extended ATPase region encompassing the UBL, ATPase and transducer domains.

1.3.3.3 Prader-Willi Syndrome (PWS)

PWS is another rare developmental condition that occurs in approximately 1 in 20,000 births. It was first identified about 50 years ago and is characterized by many features that include infantile hypotonia, hyperphagia leading to severe obesity, short stature, secondary hypogonadism with genital hypoplasia, and mild cognitive impairment [151, 152]. There are 11 known paternally expressed genes which are thought to be involved in the pathogenesis of PWS. This cluster of genes is normally imprinted, therefore it is only expressed from one of the parental alleles, in this case the paternal allele. PWS patients have lost the active paternal allele and therefore lack the ability to express several genes within the cluster entirely [153, 154]. SMCHD1 epigenetically represses four of the genes on the inactive maternal allele, including a critical gene called *MAGEL2* [44, 58]. Inhibiting SMCHD1 may therefore provide a potential treatment for PWS by enabling re-activation of the maternal allele. Importantly, *MAGEL2* and the nearby PWS genes are only considerably expressed in the hypothalamus in humans [155], enabling the idea of tissue-specific targeted treatment.

1.4 Aims

This project is focused on the biochemical and structural characterisation of Smchd1, with the aim of providing a better understanding of its function as an epigenetic regulator. In order to provide a complete understanding of Smchd1's molecular function, analysis of the full-length protein is essential. Therefore, I foremost aimed to successfully express and purify the full-length Smchd1 protein with the intent of performing structural analyses via electron microscopy. The crystal structure of SMCHD1's N-terminal ATPase region was solved by Pedersen *et al.* during the course of my studies, which led to my next aim of performing more in-depth biophysical analyses investigating the dimerization requirements of SMCHD1's ATPase domain and how these may be affected by disease-associated mutations in *SMCHD1*. With the availability of the recently solved structure of Smchd1's hinge domain, I additionally sought out to further characterise its functional properties, with a primary focus on investigating the DNA-binding mode of Smchd1's hinge domain.

Chapter 2

Materials and Methods

2.1 Molecular cloning

cDNA encoding *Smchd1* (mouse) or *SMCHD1* (human) was PCR-amplified and cloned into pPROEx HTb vector (Life Technologies) for expressing N-terminally 6-His-tagged recombinant protein or into pGEX-2T-TEV (Life Technologies) for expressing N-terminally GST-tagged recombinant protein, in BL21(DE3) Codon Plus-RIL *Escherichia coli (E. coli)* bacterial cells. For recombinant protein expression in *Sf*21 insect cells, *Smchd1* or *SMCHD1* were cloned into pFastBac HTb vector and introduced into DH10MultiBac *E. coli* cells (ATG Biosynthetics) to produce bacmid DNA for expressing N-terminally 6-His-tagged recombinant protein, as per manufacturer's instructions (Bac-to-Bac Baculovirus Expression System, Thermo Fisher). For transfection and immunofluorescence studies in human embryonic kidney (HEK293) cells, Smchd1 constructs were sub-cloned into a pcDNA3 vector (provided by Dr Tracy Willson, Walter and Eliza Hall Institute of Medical Research).

2.1.1 Oligonucleotides

A list of all oligonucleotide sequences used is outlined in section 2.7.2.

2.1.2 PCR-mediated site-directed mutagenesis

To introduce point mutations, the 5' and 3' arms of the construct of interest were PCR-amplified in separate reactions with the corresponding mutagenesis primers (see section 2.7.2). 50 μL reactions were prepared with 0.5 μL Phusion High-Fidelity DNA Polymerase (Life Technologies), 10 μL 5X Phusion High-Fidelity buffer (Life Technologies), 50 ng of template plasmid (for Smchd1, a pS1180 plasmid containing the murine Smchd1 cDNA construct was provided by Dr Graham Kay, QIMR Berghofer Medical Research Institute, Brisbane), 100 ng 5' primer, 100 ng 3' primer, 1.5 μL 10 mM dNTPs and up to 50 μL with dH_2O. PCR reactions were carried out at 98° C for 2 min, followed by 30 cycles of 98° C for 15 s, 55° C for 30 s, 72° C for 30-120 s (depending on length of insert calculated by the extension rate of 15 s/kb); and a final cycle of 72° C for 2 min with 4° C hold. Amplified PCR products were cleaned up with a PCR clean-up kit (Qiagen) according to the manufacturer's instructions. PCR products were eluted in 30 μL EB buffer [10mM Tris (pH 8)] and

digested with the appropriate restriction enzymes in reactions containing 1 μL of each restriction enzyme (20,000 units/mL, New England Biolabs), 5 μL 10x Cutsmart buffer (New England Biolabs) and up to 50 μL with dH_2O. Reactions were incubated at 37° C for 1.5 s and the digested fragments were gel purified, as outlined below.

2.1.3 Agarose gel electrophoresis and DNA purification

A 1% agarose gel (w/v) was prepared in 1X TAE buffer [40 mM Tris (pH 8.0), 1 mM EDTA, 0.1% acetic acid], supplemented with 2 μg/mL ethidium bromide. The gel was left to set at room temperature before immersion in 1X TAE buffer. DNA products were mixed with 6X purple loading dye (New England Biolabs) to a final 1X concentration before loading into wells, and electrophoresis was performed at 100 V for 30–40 min. A Generuler 1kb Plus DNA ladder (Thermo Fisher Scientific) was used as a size marker. DNA bands were visualized under UV light and bands of interest were excised and purified via a QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instructions. DNA was eluted in 30-50 μL of the provided elution buffer [10 mM Tris-HCl (pH 8.5)] and stored at -20° C.

2.1.4 Ligation

Ligation reactions consisted of 100 ng of cut vector, 3 μL of insert prepared above, 1 U of T4 DNA ligase (Promega) in 1X T4 ligase buffer (Promega), and H_2O to a final volume of 10 μL. Ligation reactions were incubated overnight at 16° C.

2.1.5 Transformation

2 μL of ligation reaction from above was added to 50 μL of freshly ice-thawed *E. coli* DH10B electrocompetent cells (Thermo Fisher), transferred to a 0.2 cm aperture cuvette and electroporated (2.5 kV) using a Micropulser Electroporator (Bio-Rad). Cells were recovered with 200 μL Superbroth [3.5% (w/v) tryptone, 2.0% (w/v) yeast extract, 0.5% (w/v) NaCl, 5 mM NaOH] and plated on LG agar plates [1.0% (w/v) tryptone, 0.5% (w/v) yeast extract, 0.5% (w/v) NaCl, 0.2% (w/v) glucose, 1.5% (w/v) Difco agar, 10 mM Tris (pH 7.4), 1 mM $MgCl_2$] containing 100 μg/mL ampicillin and incubated at 37° C overnight.

2.1.6 Plasmid extraction

1.5 mL Superbroth containing 100 μg/mL ampicillin was inoculated with a single colony and incubated overnight at 37° C, shaking at 220 rpm. Plasmids were extracted using the QIAprep Miniprep Kit (Qiagen), following the manufacturer's instructions.

2.1.7 Sequencing PCR

DNA sequencing was performed via the BigDye Terminator v3.1 Cycle Sequencing kit (Thermo Fisher Scientific). Each reaction mix consisted of 1 μL Big Dye v3.1, 1X sequencing buffer, 250 ng plasmid DNA, 4 μL of 1 μM primer and H_2O to a total reaction volume of 20 μL. PCR conditions were 96° C for 2 min, 25 cycles of 96° C for 10 s, 50° C for 10 s and 60° C for 4 min. DNA was precipitated using 1 volume 3 M sodium acetate, 5.5 volumes 100% ethanol (v/v) and 3.5 volumes H_2O, and incubated on ice for 10 min. DNA was pelleted by centrifugation at 18,000 x g for 20 min at room temperature. Pellets were washed with ice-cold 70% (v/v) ethanol and dried at 37° C for 45 min. DNA sequencing analysis was performed by Micromon DNA Sequencing Facility. Primers are listed in 2.7.2.

2.2 Cell Culture

2.2.1 *Sf*21 Insect Cells

*Sf*21 cells were cultivated in Insect-XPRESS Protein-free Insect Cell Medium with L-glutamine (Lonza) and passaged when reaching a density of 3.0-3.5x10^6 cells/mL in order to maintain log phase growth. For cultivation and viral amplification, cells were grown in suspension in 1L Schott bottles with up to 200 mL total volume, shaking at 130 rpm, at 27° C. For protein expression, cells were grown in 0.5 L cultures in 2.8 L unbaffled Fernbach flasks shaking at 90 rpm, at 27° C.

2.2.2 Mouse Embryonic Fibroblasts (MEFs)

Smchd1$^{GFP/GFP}$ MEFs were derived from a Smchd1$^{GFP/GFP}$ mouse strain generated on a C57BL/6 background. This strain carries a GFP cDNA knock-in immediately

prior to the stop codon in Smchd1 to produce a Smchd1-GFP fusion protein [55]. Smchd1$^{GFP/GFP}$ MEFs were immortalized using a p53 knockdown, as described by Dickins *et al.* [156]. MEFs were cultured in DMEM (Gibco) supplemented with 10% (v/v) FBS (Life Technologies) at 37° C in a humidified atmosphere with 5% (v/v) CO_2 and O_2. Cells were maintained by passaging every 3 days using 0.5% Trypsin-EDTA (Life Technologies) to detach the cells from the plates.

2.2.3 Human Embryonic Kidney cells (HEK293)

Both HEK293 and HEK293T cells were cultured in DMEM (Gibco) supplemented with 10% (v/v) FBS (Life Technologies) at 37° C in a humidified atmosphere with 10% (v/v) CO_2. Cells were maintained by passaging every 3 days using 0.5% Trypsin-EDTA (Life Technologies) to detach the cells from the plates.

2.2.4 shRNA-mediated *SMCHD1* knockdown in HEK293 cells

shRNA nucleotides were designed by Dr Andrew Keniry (Walter and Eliza Hall Institute of Medical Research) to target the 3' untranslated region (UTR) of SMCHD1 (listed in section 2.7.3). These were cloned into the LMPEBFP2 (LTR miR30 Puromycin IRES EBFP2) vector [157]. Retrovirus was prepared as previously described in Majewski *et al.* [158] using HEK293T cells. HEK293T cells at 80% confluency were transfected using calcium phosphate with the MD1-gag-pol structural vector, CAG-Eco and shRNA retroviral construct in the ratio 8:24:1. Plasmid DNA was made up in 250 mM CaCl2 and precipitated in 2X HBS solution, then added to the cells in media containing 25 μM chloroquine (Sigma-Aldrich). The media was changed 8 and 24 h post-transfection, and collected 48 and 72 h post-transfection via centrifugation to remove residual HEK293T cells.

For SMCHD1-knockdown, HEK293 cells were transduced with shRNA retroviral constructs, either non-silencing or designed for SMCHD1-knockdown. Retroviral supernatant was prepared 1:10 in media containing 4 μg/mL polybrene (Sigma-Aldrich) and added to HEK293 cells at 50% confluency. After 24 h, the media was changed and 5 μg/mL puromycin (Sigma-Aldrich) was added for selection of transduced cells.

2.2.5 Transfection of *SMCHD1*-knockdown HEK293 cells

Smchd1-knockdown cells were transfected for over-expression of wild-type or mutant full-length Smchd1 constructs. 2x10^4 *SMCHD1*-knockdown HEK293 cells were seeded in a 12-well plate on a 13 mm coverslip (Marienfield Superior). 24 h after plating, cells at 80% confluency were transfected using calcium phosphate-mediated transient transfection. 1 μg plasmid DNA was made up in 250 mM $CaCl_2$, precipitated in 1 volume 2X HBS solution [50 mM HEPES (pH 7.05), 10 mM KCl, 12 mM dextrose, 280 mM NaCl, 1.5 mM Na_2HPO_4] and added drop-wise to the plated cells. Cells were fixed 24 h later in preparation for immunofluorescence studies.

2.3 Recombinant protein expression and purification

2.3.1 Bacterial expression

Recombinant proteins were expressed in BL21(DE3) Codon Plus-RIL expression competent *E. coli* cells (Agilent). Cells were cultured in Superbroth supplemented with 100 μg/mL ampicillin until an optical density of 0.6-0.8 was reached, measured by A_{600}. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 18° C, shaking at 220 rpm. Cells were harvested by ultracentrifugation at 5,000 x g for 10 min, at 4° C.

2.3.2 Insect cell expression

Baculovirus containing the recombinant gene of interest was generated by transfecting *Sf*21 cells with bacmid DNA. 1 μg bacmid DNA diluted in 100 μL Insect-XPRESS protein-free medium with L-glutamine (Lonza) was mixed with 6 μL of CellFectin II (Life Technologies) diluted in 100 μL insect cell medium and incubated at room temperature for 45 min. 0.9x10^6 *Sf*21 cells were dispensed per well into a 6-well plate and incubated at 27° C for 20 min to allow adhesion. 800 μL of insect cell medium was added to each transfection mix and the mixture was pipetted dropwise onto the adhered cells. Cells were incubated at 27° C for 5 h, after which the transfection mix was removed and fresh media was added. After cells were maintained at 27° C for four days, passage 1 (P1)

baculovirus supernatant was harvested by centrifugation at 1,500 x g for 5 min at room temperature. 1 mL P1 virus was added to 100 mL *Sf*21 cells at 1.5x10^6 cells/mL and grown in 1L Schott bottles with agitation at 130 rpm at 27° C for a further four days. P2 viruses were harvested by centrifugation at 1,500 x g for 5 min at room temperature, and stored at 4° C protected from light. For protein expression, 50 mL P2 baculovirus was added to 500 mL 3x10^6 cells/mL in Fernbach flasks and cultured at 27° C for 2 days, shaking at 90 rpm. Cells were harvested by centrifugation at 1,500 x g for 5 min at room temperature and cell pellets were snap-frozen in liquid nitrogen and stored at -80° C.

2.3.3 Cell lysis

Purification was performed as previously described by Babon *et al.* [159]. Cells were re-suspended in lysis buffer [0.5 M NaCl, 20 mM Tris-Hcl (pH 8.0), 20% (v/v) glycerol, 5 mM imidazole (pH 8.0), 0.5 mM TCEP] supplemented with 1 mM PMSF and 1X cOmplete EDTA-free protease inhibitor (Roche). Sonication was performed on ice for 5 cycles with 1 s on, 0.2 s off, 22 s per cycle at 50% amplitude, using the Bandelin sonicator fitted with the VS 70/T probe. To remove insoluble material, lysates were centrifuged at 45,000 x g for 30 min at 4° C.

2.3.4 Protein purification

2.3.4.1 Immobilised metal-ion affinity chromatography (IMAC)

Following cell lysis by sonication, lysate supernatant from cells expressing N-terminal 6-His-tagged proteins were incubated with nickel-nitrilotriacetic acid (Ni-NTA) cOmplete His-tag purification resin (Roche) for 1 h at 4° C, on rollers. 1 ml of 50% resin slurry was used per 1 L of cell culture. The resin was pelleted by centrifugation at 1,500 x g for 5 min at 4° C and the supernatant was removed as the unbound sample. The resin was washed twice with 5 mM imidazole buffer (pH 8.0), followed by two washes with 35 mM imidazole (pH 8.0) and eluted in 250 mM imidazole buffer (pH 8.0). The 6-His-tag was cleaved by incubation of the pooled elutions with tobacco etch virus (TEV) protease (a gift from Dr. Cheree Fitzgibbon, Walter and Eliza Hall Institute of Medical Research) overnight at 4° C. Next day, cleaved protein was concentrated with a 30-kDa molecular

mass cut-off concentrator (Millipore) by centrifugation for 5 min at 2,500 x g, 4° C, then diluted either with lysis buffer with 5 mM imidazole (pH 8.0) prior to subtractive Ni-NTA chromatography which was performed to remove uncleaved protein and any remaining 6-His-tagged TEV protease, or diluted in the corresponding buffer for the following chromatography step.

2.3.4.2 Subtractive immobilised metal-ion affinity chromatography

For purification of recombinant Smchd1 hinge domain protein, subtractive IMAC was performed with Ni-NTA cOmplete His-tag purification resin (Roche) to eliminate undigested protein and 6-His-tagged TEV protease. The cleaved protein was diluted in lysis buffer [0.5 M NaCl, 20 mM Tris-HCl (pH 8.0), 20% (v/v) glycerol, 5 mM imidazole (pH 8.0), 0.5 mM TCEP] and incubated with Ni-NTA resin for 1 h at 4° C, on rollers. The resin was centrifuged at 1,500 x g at 4° C for 5 min and the unbound portion, containing the cleaved protein, was retained. The resin was washed with increasing imidazole concentrations and samples of eluting fractions were verified via SDS-PAGE. Fractions of interest containing the cleaved protein were pooled and concentrated, after which size exclusion chromatography was performed.

2.3.4.3 GST affinity chromatography

For GST-tagged proteins, cell lysis was performed as described previously but cells were instead re-suspended in a size exclusion chromatography (SEC200) buffer [200 mM NaCl, 20 mM HEPES (pH 7.5), 5% (v/v) glycerol] supplemented with 1 mM PMSF and 1X cOmplete EDTA-free protease inhibitor (Roche). Lysate supernatants were incubated with glutathione agarose resin (Ubiquitin-Proteasome Biotechnologies) for 1 h at 4° C, on rollers; using 3 mL of 50% slurry for 1L of cell culture. The resin was pelleted by centrifugation at 1,500 x g for 5 min at 4° C and the supernatant was removed as the unbound sample. The resin was washed three times with lysis buffer after which TEV protease was added to the resin and incubated overnight without agitation at 4° C for cleavage of the GST-tag. Next day, the resin was washed three times with 5 mL lysis buffer and these elution samples were pooled together and concentrated with a 30-kDa molecular mass cut-off concentrator (Millipore) by centrifugation for 5 min at 2,500 x g, 4° C.

2.3.4.4 Ion exchange chromatography

Ion exchange chromatography was performed for the N-terminal ATPase region of Smchd1 as well as for full-length Smchd1 protein. For the ATPase region, protein concentrated in Buffer A [50 mM NaCl, 25 mM HEPES (pH 7.5), 0.5 mM TCEP, 10% (v/v) glycerol] was loaded onto a MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with Buffer A and exchanged into Buffer B [500 mM NaCl, 25 mM HEPES (pH 7.5), 0.5 mM TCEP, 10% (v/v) glycerol] in a 0-100% gradient over 20 column volumes for protein elution. For full-length recombinant Smchd1, two separate ion exchange chromatography conditions were trialled as a third purification step, following size exclusion chromatography. The protein, which was concentrated in the corresponding Buffer A, was either loaded on a MonoQ 5/50 GL column (GE Healthcare) pre-equilibrated with Buffer A [50 mM NaCl, 25 mM Tris-HCl (pH 9.5), 0.5 mM TCEP, 10% (v/v) glycerol] and exchanged into Buffer B [500 mM NaCl, 25 mM Tris-HCl (pH 9.5), 0.5 mM TCEP, 10% (v/v) glycerol], or loaded on a MonoS 5/50 GL column (GE Healthcare) pre-equilibrated with Buffer A [50 mM NaCl, 25 mM MES (pH 5.5), 0.5 mM TCEP, 10% (v/v) glycerol] and exchanged into Buffer B [1 M NaCl, 25 mM MES (pH 5.5), 0.5 mM TCEP, 10% (v/v) glycerol] in a 0-100% gradient over 20-column volumes for protein elution. Fractions containing the N-terminal ATPase region of Smchd1 eluted at approximately 45% Buffer B. These were pooled and concentrated to be further purified by size exclusion chromatography. For full-length Smchd1 protein, fractions of interest from the MonoQ column eluted at approximately 45% Buffer B, whereas from the MonoS column they eluted at approximately 55% Buffer B. Samples were taken directly from eluting fractions and subjected to negative stain and electron microscopy analysis.

2.3.4.5 Size exclusion chromatography (SEC)

For purification of recombinant full-length Smchd1 protein, size exclusion chromatography was performed on a Superose-6 Increase column (GE Healthcare) pre-equilibrated with SEC500 buffer [500 mM NaCl, 20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 0.5 mM TCEP]. Chromatography of recombinant Smchd1 N-terminal ATPase region or the C-terminal hinge domain was performed on a Superdex-200 10/300 GL column (GE Healthcare) pre-equilibrated with either SEC100 buffer [100

mM NaCl, 20 mM HEPES (pH 7.5)] or SEC200 buffer [200 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10% (v/v) glycerol, 0.5 mM TCEP] for protein prepared for small-angle X-ray scattering (SAXS) experiments. Fractions containing the recombinant proteins of interest were pooled concentrated and snap-frozen in liquid nitrogen for storage at -80° C. Protein concentration was measured using a DS-11 FX Spectrophotometer (DeNovix), using protein A_{280} absorbance values for 0.1 % (=mg/mL) of 1.15 for SMCHD1 ATPase region, 0.78 for full-length Smchd1 and 0.687 for the Smchd1 hinge domain.

2.4 Protein biochemistry

2.4.1 SDS-PAGE

Protein samples were mixed with 4X SDS sample reducing buffer [200 mM Tris-Cl (pH 6.8), 400 mM Dithiothreitol (DTT), 8% SDS, 0.4% bromophenol blue, 40% glycerol] to a final 1X dilution and heated at 95° C for 5 min. Samples were resolved by SDS-PAGE via either 4-12% Bis-Tris gels (Novex) or 4-20% Mini PROTEAN TGX Stain-Free gels (Bio-Rad), using MES running buffer [50 mM MES, 50 mM Tris, 1 mM EDTA, 0.1% (w/v) SDS; at a final pH of 7.3] or Tris-Glycine buffer [25 mM Tris, 190 mM glycine, 0.1% (w/v) SDS], respectively. Gels were either stained with Coomassie SimplyBlue Safestain (Life Technologies) or visualised on a ChemiDoc Imaging System (Bio-Rad) using a Stain-Free setting.

2.4.2 Preparation of whole cell extracts

Media was aspirated from cultured cells and cells were scraped off plates with cold PBS buffer on ice, then lysed in KALB lysis buffer [150 mM NaCl, 50 mM Tris-Cl (pH 7.5), 1% (v/v) Triton X-100, 1 mM EDTA (pH 7.5)] supplemented with 1 mM PMSF and 1X cOmplete EDTA-free protease inhibitor (Roche) on ice for 30 min with resuspension every 10 min. Insoluble material was removed by centrifugation at 20,000 x g for 10 min at 4° C and the supernatant was transferred to fresh tubes. Total protein concentration of lysates was quantified via the Pierce BCA protein assay kit (Thermo Fisher Scientific), following the manufacturer's instructions.

2.4.3 Immunoprecipitation

Immunoprecipitation (IP) of Smchd1-GFP from Smchd1$^{GFP/GFP}$ MEFs was performed with GFP-Trap_MA beads (Chromotek). 25 μL of GFP-Trap_MA beads per IP were washed three times in 500 μL dilution buffer [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 mM EDTA] and added to 500 μL lysate with incubation for 1 h at 4° C with end-to-end rotation. The supernatant was discarded and the beads were washed three times in 500 μL dilution buffer. Bound proteins were eluted from the beads by re-suspension in 40 μL of 0.2 M glycine (pH 2.5) for 30 s under constant mixing, followed by neutralisation with 10 μL of 1M Tris-HCl (pH 10.4). This step was repeated twice more to obtain a total of three elution samples. Samples were analysed both by western blot and by mass spectrometry.

Immunoprecipitation experiments testing for potential Smchd1 interacting partners were performed using Smchd1-GFP MEFs, where 500 μg of whole cell lysate was incubated with 5 μg antibody of interest at 4° C for 2 h with end-to-end rotation. The lysate-antibody mix was incubated with 20 μL pre-washed Dynabeads Protein G (Life Technologies) for 1 h at 4° C under rotation. The unbound fraction was collected by using a magnetic rack to separate the Dynabeads. Beads were washed with 500 μL KALB lysis buffer three times and the immuno-complex was eluted from the beads with 40 μL 2X SDS buffer, heated at 95° C for 10 min, after which samples were analysed by SDS-PAGE and Western Blot, immunoblotting against Smchd1.

2.4.4 Western Blot analysis

Samples were resolved by standard SDS-PAGE analysis on 4-12% Bis-Tris gels (Thermo Fisher Scientific) in MES buffer and transferred to a PVDF membrane (Osmonics, GE Healthcare) by wet transfer at 100 V for 1 h in transfer buffer [25 mM Tris, 192 mM glycine, 20% (v/v) methanol]. Membranes were blocked with a 5% (v/v) skim milk powder in 0.1% (v/v) Tween-20/PBS for 1 h at room temperature. Primary antibody was added to the membranes in 5 mL blocking buffer and incubated overnight at 4° C in a capped tube, on rollers. Membranes were washed for 30 min at room temperature with 0.1% (v/v) Tween-20/PBS, followed by incubation with secondary antibody for 1 h at room temperature, which was diluted in 5 mL blocking buffer. The 30 minute washing

step was repeated as before and antibody binding was visualised using the Luminata ECL system (Millipore) following the manufacturer's instructions.

2.4.5 Blue-Native PAGE

Cell lysates or recombinant protein samples were supplemented with 10x loading dye [100 mM Bis-tris, (pH 7.0), 500 mM 6-aminohexanoic acid, 5% (w/v) Coomassie Brilliant Blue] and samples were loaded onto a 4-12% pre-cast Bis-Tris gradient gel (Thermo Fisher Scientific) and run at a constant current of 8 mA. The anode buffer remained unchanged during electrophoresis, whereas the Coomassie-cathode buffer was replaced with cathode buffer when the dye front had run approximately one third of the way through the gel. Both buffers were prepared as per manufacturer's instructions (Thermo Fisher Scientific). The transfer of proteins onto PVDF membranes was performed as stated in 2.4.4. Molecular masses were approximated using unstained native protein markers (Thermo Fisher Scientific) following incubation under agitation with destaining solution [50% (v/v) methanol and 25% (v/v) acetic acid]. Membranes were rinsed in distilled water prior to western blot transfer.

2.4.6 Silver stain

Samples were resolved by standard SDS-PAGE analysis. The resulting gel was fixed in a 30% (v/v) ethanol and 10% (v/v) acetic acid solution for 1 h, followed by a 1.5 h incubation in a 30% (v/v) ethanol, 0.4 M sodium acetate, 12.5 mM sodium thiosulfate and 2% (v/v) glutaraldehyde solution. The gel was washed with H_2O for 30 min with several changes, and incubated with a 0.1% (v/v) silver nitrate and 0.02% (v/v) formaldehyde solution for 30 min. The gel was briefly rinsed with H_2O and developed with a 2.5% (v/v) sodium carbonate and 0.01% (v/v) formaldehyde solution until a desired darkness was achieved. The reaction was quenched in a 12% (v/v) methanol and 7% (v/v) acetic acid solution.

2.4.7 Differential Scanning Fluorometry (DSF)

DSF was performed for recombinant full-length Smchd1 at the CSIRO Collaborative Crystallisation Centre, Melbourne, Australia. Reactions were set up in 20 μL volumes,

containing 0.3 μL full-length Smchd1 at 1.3 mg/mL, 0.3 μL SYPRO orange dye and 19.4 μL of corresponding buffer condition tested. 14 different buffer conditions were analyzed, each set up in triplicates, alongside lysozyme as a positive control for each testing condition.

2.4.8 Cycloheximide chase assay

0.36 x 10^6 *Sf*21 cells were seeded per well in a 12-well plate, to which P2 viruses were added in corresponding volumes and incubated at 27° C. After 48 h, a fixed 10 μL volume of cycloheximide (CHX) (Sigma-Aldrich) was added to corresponding wells at final concentrations of either 50 or 200 μg/mL, or 10 μL DMSO for control wells. At either 0, 8 or 24 h after CHX addition, the media was aspirated and 100 μL of Cytobuster lysis buffer (Merck, Millipore) was added to each well. The lysates were collected and pelleted at 16,000 x g for 5 min at 4° C, and a Western Blot was performed (section 2.4.4).

2.4.9 Circular Dichroism (CD) spectroscopy

Far UV CD spectra were recorded for recombinant full-length Smchd1 protein (2.5 μM, in PBS supplemented with 10 % glycerol) using an Aviv 410SF CD spectrometer (Lakewood, NJ). Spectra were acquired between 190 and 260 nm (far UV), using a 1 mm path-length quartz cuvette, 0.5 nm step size, and 4 s averaging. CD spectroscopy data were analysed for secondary structure content using the K2D2 online program [160].

2.4.10 Limited Proteolysis and N-terminal Sequencing

400 μg of purified SMCHD1 N-terminal region protein (residues 111-702) in SEC100 buffer [100 mM NaCl, 20 mM HEPES (pH 7.5)] was incubated with varying concentrations of trypsin (Promega), starting from an original 1 μg/μL stock solution (diluted in SEC100 buffer), in ratios of 1:2000, 1:1000, 1:500, 1:250 and 1:125 (v/v) to SMCHD1 protein. Similarly, proteolytic digestion with chymotrypsin (Promega) (prepared in SEC100 buffer), was set up from an original 1 μg/μL stock solution in 1:1000, 1:500, 1:250, 1:125, 1:62.5 and 1:31.25 (v/v) ratios. Reactions were performed

at room temperature for 30 min and samples were analysed by reducing SDS-PAGE. For N-terminal sequencing, samples were transferred to a PVDF membrane via Western Blot (section 2.7.3). The membrane was stained with Ponceau S stain [0.1% (w/v) Ponceau S, 5% (v/v) acetic acid] to visualise individual bands, and the band of interest was excised with a clean scalpel and sent for N-terminal sequencing to the Australian Proteome Analysis Facility (APAF), Macquarie University, Sydney, Australia.

2.4.11 Hydrogen-Deuterium Exchange and Mass Spectrometry (HDX-MS)

HDX-MS was performed as outlined in Petrie *et al.* [161]. Briefly, a final SMCHD1 (111-702 aa) protein concentration of 1 mg/mL was achieved by dilution in deuterated buffer [10 mM Tris-d_{11} (pH 7.5)]. Aliquots were taken at multiple time points (30 s, 60 s, 10 min and 60 min) with the hydrogen/deuterium exchange reaction suppressed by acidification of the sample to pH 2.5 using formic acid before snap-freezing in liquid nitrogen. Digestion of the protein was carried out by thawing the sample in a ten-fold dilution of H_2O before addition of an equimolar concentration of pepsin (Sigma-Aldrich) for 5 min on ice. Peptides were subjected to LC-MS analysis using an 1100 series HPLC (Agilent) coupled to an LTQ-Orbitrap XL (Thermo Fisher Scientific). Peptides were loaded onto an in-house packed, reverse-phase trap column (ReproSil-Pur C18 (Dr. Maisch GmbH, Germany), 2 X 2 mm^2, 5 μm) before separation on an in-house packed, reverse-phase analytical column [ReproSil-Pur C18 (Dr. Maisch GmbH)], 200 μm X 150 mm, 3 μm) housed at 1° C. Peptides were loaded onto the trap column at 5% acetonitrile and 0.2% formic acid, and elution was performed using a gradient rising from 5% to 40% acetonitrile over 12 min, then 85% acetonitrile for 5 min before reconditioning the column at 5% acetonitrile for 15 min. Spectra were acquired in positive ion mode with m/z range from 350 to 1850. Deuteration of peptides was determined by analysis of samples using HDX workbench followed by additional analysis of the HDX workbench output was performed using in-house scripts written in R (version 3.3.1). Deuterium exchange per residue was calculated by summing and averaging deuterium exchange across overlapping peptide regions for each residue.

2.4.12 Preparation of Fab fragments

Fab fragments were prepared with the Pierce Fab Preparation Kit (Thermo Fisher Scientific) following the manufacturer's instructions, using the in-house anti-Smchd1 (N-terminal ATPase region) monoclonal antibody (clone 1D6, see section 2.7.1) raised in rats at the WEHI antibody facility.

2.4.13 Fluorescence polarization assays

Fluorescence polarization (FP) assays were performed with the Transcreener ADP^2 FP assay kit (BellBrook Labs), as outlined in the manufacturer's instructions and summarised below.

2.4.13.1 ATPase assay

ATPase assays were performed as outlined in Chen *et al.* [61]. 10 μL reactions were set up in triplicates in 384-well low flange, black, flat-bottom plates (Corning) containing 7 μL reaction buffer [50 mM HEPES (pH 7.5), 4 mM $MgCl_2$, 2 mM EGTA], 1 μL recombinant protein at concentrations ranging from 0.1-0.6 μM or SEC buffer control, 1 μL nuclease-free water and 1.25-10 μM ATP substrate. Reactions were incubated at 25° C for 1 h in the dark. Reactions were stopped by the addition of 10 μL detection mix [1X Detection buffer, 4 nM ADP AlexaFluor 633 Tracer, 128 μg/mL ADP^2 antibody] and incubated for another h in the dark. Fluorescence polarization readings (mP) were measured with an Envision plate reader (PerkinElmer Life Sciences) fitted with excitation filter 620/40 nm, emission filters 688/45 nm (s and p channels) and D658/fp688 dual mirror. Readings from a free tracer (no antibody) control were set as 20 mP as the normalization baseline of the assay for all reactions. The amount of ADP produced by each reaction was estimated by a 12-point standard curve, as outlined in the manufacturer's protocol. Data were plotted and analysed in GraphPad Prism.

2.4.13.2 DNA-binding assay

12.5 nM of 6-Fam fluorescently-labeled single-stranded DNA (HPLC purified, IDT) (listed in section 2.7.4) as previously reported [162] was incubated with recombinant

Smchd1 hinge domain protein at final concentrations of 0-100 μM diluted in SEC100 buffer [100 mM NaCl, 20 mM HEPES (pH 7.5)] in 10 μL reactions. Reactions were set up in 384-well low flange black flat bottom plates (Corning) in duplicate and incubated at room temperature for 15 min in the dark. The emission polarization values were measured with Envision Multilabel plate reader (PerkinElmer) with a 480 nm excitation filter, a 535 nm static and polarized filter, and FITC FP dual mirror. Data were plotted and analysed in GraphPad Prism.

2.4.14 ADP-Glo ATPase assay

A second type of ATPase assay was performed using the ADP-Glo Kinase Assay kit (Promega). Each reaction was performed in a total volume of 5 μL, consisting of 0.5 μM or 1.0 μM full-length Smchd1 protein, 10 μM ATP and reaction buffer [50 mM HEPES (pH 7.5), 4 mM $MgCl_2$, 2 mM EGTA]. Reactions were incubated at room temperature for 1 h, followed by the addition of 5 μL ADP-Glo reagent to terminate the reaction by depleting ATP, and a further incubation for 40 min at room temperature. 10 μL of Kinase Detection Reagent was then added to each reaction and incubated for a further 60 min at room temperature. Luminescence was measured using the FLUOstar Omega microplate reader (BMG Labtech). Data were plotted and analysed in GraphPad Prism.

2.4.15 Electromobility Shift Assay (EMSA)

EMSA was performed as previously described by Griese *et al.* [163]. Briefly, 50 nM of 6-Fam-labelled oligonucleotides (Integrated DNA Technologies), as listed in section 2.7.4, were mixed with recombinant Smchd1 protein (either full-length or hinge domain only, wild-type or mutants) in 0-, 5-, 10-, 25-, 50-fold molar excess over DNA in SEC100 buffer [100 mM NaCl, 20 mM HEPES (pH 7.5)] in a total volume of 20 μL. Samples were incubated for 30 min at room temperature, following the addition of 5 μL of 50% (v/v) glycerol. The samples were loaded onto a 0.5% (w/v) agarose gel in 1X TBE buffer [890 mM Tris base, 890 mM boric acid, 20 mM EDTA (pH 8.0)] and separated for 1.5 h at 4 V/cm at 4° C. Gels were scanned on a Typhoon 9410 fluorescence scanner with a 526-nm short-pass filter (GE Healthcare).

2.4.16 Filter-Aided Sample Preparation (FASP) and Mass Spectrometry

Elutions obtained following immunoprecipitation experiments were reduced with 100 mM TCEP for 30 min at room temperature. Each sample was added onto a FASP column, to which 200 μL of urea/Tris solution (12.5 M urea, 100 mM Tris-HCl) was added and centrifuged at 14,000 x g for 15 min at room temperature. This step was repeated once more and the flow-through was discarded. 100 μL of 50 mM iodoacetamide was added to each FASP column and vortexed briefly before incubation in the dark for 20 min. The columns were centrifuged at 14,000 x g for 10 min at room temperature. 100 μL of urea/Tris-HCl was added to each FASP column and centrifuged at 14,000 x g for 15 min at room temperature; this step was repeated a total of three times. 100 μL of 50 mM ammonium bicarbonate was added to each FASP column and centrifuged at 14,000 x g for 10 min at room temperature; this step was performed a total of three times. 2 μg of Trypsin Gold (Promega) was added to each FASP column and incubated overnight at 37° C under agitation. Next day, the FASP columns was transferred to new FASP collection tubes. 40 μL of 50 mM ammonium bicarbonate was added to each FASP column and centrifuged at 14,000 x g for 10 min at room temperature; this step was performed twice. The peptides collected in the final spin were acidified with formic acid at a 1% (v/v) final concentration and stored at -80° C prior to lyophilising.

Peptides were resuspended in 2% (v/v) acetonitrile and 1% (v/v) formic acid and injected and separated by reverse-phase liquid chromatography on a M-class UHPLC system (Waters) using a 250 mm x 75 mm column (1.7 mm C18, packed emitter tip, Ion Opticks, Australia) with a linear 90-minute gradient at a flow rate of 400 nl/min from 98% (v/v) solvent A (0.1% (v/v) formic acid in H_2O) to 35% (v/v) solvent B (0.1% (v/v) formic acid. 99/9% acetonitrile). The nano-UHPLC was coupled in-line to a Q-Exactive Orbitrap mass spectrometer equipped with an EASY-spray ionization source (Thermo Fisher Scientific). The Q-Exactive was operated in a data-dependent mode, switching automatically between one full scan and subsequent MS/MS scans of the ten most abundant peaks. The instrument was controlled using Exactive series version 2.8 build 2806 and Xcalibur 4.0. Full-scans (m/z 350–1,850) were acquired with a resolution of 70,000 at 200 m/z. The 10 most intense ions were sequentially isolated with a target

value of 100,000 ions and an isolation width of 2 m/z and fragmented using higher-energy collisional dissociation with stepped normalised collision energy of 19.5, 26, 32. Maximum ion accumulation times were set to 80 ms for full MS scan and 200 ms for MS/MS.

Raw files consisting of high-resolution MS/MS spectra were processed with MaxQuant (version 1.5.8.3) [164] for feature detection and protein identification using the Andromeda search engine [165]. Extracted peak lists were searched against the *Mus musculus* Uniprot database in addition to a separate reverse decoy database to empirically assess the false discovery rate (FDR) using a strict trypsin specificity allowing up to 2 missed cleavages. The minimum required peptide length was set to 6 amino acids.

2.4.17 Analytical Ultracentrifugation (AUC)

Sedimentation velocity experiments were performed with a XL-I analytical ultracentrifuge (Beckman Coulter) using double sector quartz cells and epon center-pieces in an An-50 Ti 8-hole rotor. Data were obtained at 50,000 rpm using 380 µL protein at 1.0 mg/ml concentration in buffer SEC100HEPES (100 mM NaCl, 20 mM HEPES pH 7.5) as the sample, and 400 µL of buffer SEC100HEPES as the reference. A total of 100 scans were collected at 20° C using radial absorbance scans at 290 nm and a step size of 0.003 cm. All data were analysed using SEDFIT [166]. Sedimentation data were fitted to a continuous size distribution [c(s)]. Fit data is presented using GUSSI [167]. The buffer density, buffer viscosity and an estimate of the partial specific volume of the protein sample based on the amino acid sequence were also determined using SEDNTERP.

2.5 Structural Studies

2.5.1 Small angle X-ray scattering

2.5.1.1 Data collection

SAXS data were collected on the SAXS/WAXS beamline at the Australian Synchrotron, coupled with in-line size exclusion chromatography. 55 μL at 4.5 mg/mL of recombinant Smchd1 hinge domain samples were each loaded onto a Superdex-200 5/150 (GE Healthcare) pre-equilibrated in purification buffer [200 mM NaCl, 20 mM Tris-HCl (pH 8.0), 10% (v/v) glycerol, 0.5 mM TCEP]. 55 μL at 4 mg/mL of recombinant wild-type full-length Smchd1 was loaded onto a Superose-6 Increase 10/300 GL (GE Healthcare) column, pre-equilibrated in SEC500 buffer [500 mM NaCl, 20 mM HEPES (pH 7.5), 10% (v/v) glycerol, 0.5 mM TCEP]. For Smchd1 hinge domain, samples were eluted via a 1.5 mm glass capillary at 16° C positioned in the X-ray beam. Diffraction data were collected with a 1M, 170 mm x 170 mm Pilatus detector at 2 s intervals over the course of the elution. For full-length Smchd1, samples were eluted via a 1.5 mm glass capillary at 16° C positioned in the X-ray beam and diffraction data was collected with a 1M 170 mm x 170 mm Pilatus detector at 2 s intervals over the course of the elution. Data were processed by the beamline control software, ScatterBrain. 2D intensity plots from the size exclusion chromatography peak of the eluting protein sample were radially averaged and normalised to sample transmission. Scattering profiles from buffer alone were averaged for background subtraction of 1D profiles.

2.5.1.2 Data analysis

Data analyses were performed with the ATSAS suite [168]. PRIMUS [169] was used to perform Guinier analysis for examining scattering curves at small angles (qRg below 1.3). From this, estimation of two parameters can be obtained: the radius of gyration (Rg) value, which represents the square root of the average distance of each scattering atom from the particle centre, and zero angle intensity ($I(0)$), which is proportional to the molecular weight and the concentration of the protein. The linearity of the Guinier plot reflects the quality of the scattering data obtained, indicating the absence

of high molecular weight aggregates or inter-particle interference. Real-space interatomic distance distribution function, $P(r)$, and maximum dimension of the scattering particle, D_{max}, were computed by indirect Fourier transform via GNOM [170]. Low resolution shape envelopes were generated via the *ab initio* bead-modeling program, DAMMIF [171]. 10 independent models were built from each scattering profile and were further compared in the program DAMSEL, where the mean value of the normalised spatial discrepancy (NSD) of the 10 models were calculated. The most probable models were aligned by DAMSUP, and an averaged model was obtained via DAMAVER. The final model was adjusted to correspond with the experimentally obtained data via the program DAMFILT [169].

2.5.2 Crystallisation trials

Crystallisation trials were conducted at the CSIRO Collaborative Crystallisation Centre, Melbourne, Australia. Recombinant SMCHD1 protein (residues 111-702) at concentrations of either 5 or 10 mg/mL in SEC100 buffer [100 mM NaCl, 20 mM HEPES (pH 7.5)], either in the apo state or pre-incubated with 0.5 mM Radicicol (Sigma-Aldrich). Crystal trays were set up via the sitting drop method in 96-well plates under various chemically defined conditions, using the C3 screens: C3-1, C3-2, C3-3, C3-4, C3-5, C3-6, Shotgun and Proplex, all at 20° C. For each well, 150 nL protein was mixed with 150 nL crystallant and equilibrated against 50 μL crystallant in the reservoir, via nano-dispensing robots.

A lysine methylation reaction was prepared prior to one crystallisation trial, as described by Walter *et al.* [172]. Briefly, 1 mg of SMCHD1 protein (residues 111-702) in 1 mL volume was mixed with 20 μL dimethylamine-borane (Sigma-Aldrich) and 40 μL of 1M formaldehyde (Sigma-Aldrich). The mixture was incubated for 2 h at 4° C, on rollers. 20 μL dimethylamine-borane and 40 μL of 1M formaldehyde was again added to the reaction, and incubated for a further 2 h. The process was repeated once more and the reaction was incubated overnight 4° C, on rollers. The following day, the sample was centrifuged at 5,000 x g and soluble protein was collected and applied onto a Superdex-200 10/30 GL (GE Healthcare) column for size exclusion chromatography. Chosen fractions containing the protein of interest were pooled and concentrated to 15 mg/mL for crystallisation.

2.6 Microscopy

2.6.1 Electron Microscopy

To prepare samples for negative stain, 1.2/1.3 μm grids with 2 nm ultrathin holey carbon (300 mesh) (Quantifoil) were glow-discharged (PELCO easiGlow) for 30 s. 200 ng of purified full-length Smchd1 protein was applied on the grid and incubated for 30 s, after which excess protein was blotted with filter paper. The grids were washed and blotted twice with H_2O and twice with 1% (v/v) uranyl acetate solution, then air-dried for 20 s or until completely dry. Prepared grids were visualized on a FEI Talos L120C transmission electron microscope (Bio21 Institute, University of Melbourne, Australia). Images were obtained at x 150,000 magnification.

2.6.2 Immunofluorescence

Immunofluorescence was performed on *SMCHD1*-knockdown or *SMCHD1*-knockout HEK293 cells, either untransfected as a negative control or transfected with various full-length *Smchd1* constructs containing point mutations of interest. The protocol followed is described in Chaumeil *et al.* [173], with minor modifications. Briefly, prepared cells which were cultured on coverslips were washed in PBS and fixed in 3% (w/v) paraformaldehyde made in PBS for 10 min at room temperature. Cells were washed three times in PBS for 5 min each, followed by permeabilisation in 0.5% Triton X-100 in PBS on ice for 5 min. Cells were washed three times in PBS for 5 min each and then blocked in 1% (w/v) Bovine Serum Albumin (BSA) (Life Technologies) for 15 min at room temperature. Cells were incubated with primary antibody diluted in 1% (w/v) BSA overnight at 4° C in a dark and humid chamber. Next day, cells were washed three times in PBS for 5 min each at room temperature and incubated with a secondary antibody conjugated to a fluorophore diluted in 1% (w/v) BSA for 40 min at room temperature in a dark and humid chamber. Cells were washed three times in PBS for 5 min each and counterstained with DAPI for 10 min at room temperature in a dark chamber. Coverslips were mounted in Vectashield Mounting Medium H-1000 (Vector Laboratories). Cells were visualised on a LSM 880 (Zeiss) microscope at 63x magnification and z-stacks were acquired. Images were analysed using the open source ImageJ distribution package, FIJI.

2.7 Reagents

2.7.1 Antibodies

Antibody	**Manufacturer**	**Catalogue Number**	**Use**	**Host**
α-Smchd1	WEHI, in-house	Clone 1D6	WB (1:2000)	Rat
α-Smchd1	WEHI, in-house	Clone 2B8	IF (1:100)	Rat
α-Tubulin	Sigma-Aldrich	T5168	WB (1:1000)	Mouse
α-H3K27me3	Millipore	07-449	IF (1:100)	Rabbit
α-H2AK119ub	Cell Signalling	8240S	IF (1:100)	Rabbit
α-Mouse IgG HRP-conjugated	Southern Biotech	1036-05	WB (1:10,000)	Goat
α-Rat IgG HRP-conjugated	Southern Biotech	3010-05	WB (1:10,000)	Goat
α-Rabbit IgG HRP-conjugated	Southern Biotech	4030-05	WB (1:10,000)	Goat
α-Rat-568	Life Technologies	A-11077	IF (1:500)	Goat
α-Rabbit-647	Life Technologies	A-21245	IF (1:500)	Goat
α-B23	Santa Cruz	Sc-32256	IP, 5 μg per IP	Mouse
α-RBM3	Santa Cruz	Sc-390139	IP, 5 μg per IP	Mouse
α-NAP1L1	Santa Cruz	Sc-81328	IP, 5 μg per IP	Mouse
α-DDX3	Santa Cruz	Sc-81247	IP, 5 μg per IP	Mouse
α-p54/nrb	Santa Cruz	Sc-376804	IP, 5 μg per IP	Mouse
α-Smarca4	Abcam	ab110641	WB (1:1000)	Rabbit
α-Smarcc1	Cell Signalling	D7F8S	WB (1:1000)	Rabbit
α-RYBP	Abcam	ab185971	WB (1:1000)	Rabbit
α-Trim37	Sigma-Aldrich	SAB2702009	WB (1:1000)	Rabbit
α-Histone 3	Abcam	ab1791	WB (1:1000)	Rabbit

2.7.2 Cloning and mutagenesis oligonucleotides

5' BamHI Smchd1 1683 aa	cgcggatccacacaacagacaactcatattgaagcac
5' BamHI Smchd1 1699 aa	cgcggatccgaacagaacgagcttaagaagaggc
3' EcoRI Smchd1 1899 aa	cggaattcactgctttggtactggagctccgaag
3' EcoRI Smchd1 1881 aa	cggaattcagctttattctgaaggcccccaaac
5' Smchd1 A667E	ctgtgcccattgCaaagctggatagg
3' Smchd1 A667E	cctatccagctttGcaatgggcacag
5' SMCHD1 A667E	gtgcccattgCaaagctggatagg
3' SMCHD1 A667E	cctatccagctttGcaatgggcac
5' Smchd1 V1774G	caaggtcgccagcaagGgttgccccttgattc
3' Smchd1 V1774G	gaatcaaggggcaacCcttgctggcgaccttg
5' Smchd1 G1872A/G1875A/G1876A	gagatagaattagaagtaatgCgaagtttgCggCccttcagaataaagctccgccaatgg
3' Smchd1 G1872A/G1875A/G1876A	ccattggcggagctttattctgaaggGccGcaaacttcGcattacttctaattctatctc
5' Smchd1 R1848A	gcagccaatcattatGCaaaagaggttgttaaaatc
3' Smchd1 R1848A	gattttaacaacctcttttGCataatgattggctgc
5' Smchd1 R1867G	ctagagatggagatGgaattagaagtaatg
3' Smchd1 R1867G	cattacttctaattcCatctccatctctag
5' Smchd1 R1790A	cttccagattggaaaGCacctctacctcatttc
3' Smchd1 R1790A	gaaatgaggtagaggtGCtttccaatctggaag
5' Smchd1 R1796A	cctctacctcatttcGCaaatggtaaattacac
3' Smchd1 R1796A	gtgtaatttaccatttGCgaaatgaggtagagg
5' Smchd1 D1842A	cattattttggataatctggCtgcagccaatcattatag
3' Smchd1 D1842A	ctataatgattggctgcaGccagattatccaaaataatg
5' Smchd1 R1869A	gatggagatagaattGCaagtaatgggaagtttg
3' Smchd1 R1869A	caaacttcccattacttGCaattctatctccatc
5' Smchd1 K1873A	gaattagaagtaatgggGCgtttgggggccttcag
3' Smchd1 K1873A	ctgaaggcccccaaacGCcccattacttctaattc
5' Smchd1 S1870M	ggagatagaattagaaTGaatgggaagtttggg
3' Smchd1 S1870M	cccaaacttcccattCAttctaattctatctcc
5' Smchd1 S1870N	ggagatagaattagaaAtaatgggaagtttggg
3' Smchd1 S1870N	cccaaacttcccattaTttctaattctatctcc
5' Smchd1 S1870Q	ggagatagaattagaCAGaatgggaagtttggg
3' Smchd1 S1870Q	cccaaacttcccattCTGtctaattctatctcc

5' Smchd1 H1856W	gttgttaaaatcacaTGGtgccccacactc
3' Smchd1 H1856W	gagtgtggggcaCCAtgtgattttaacaac
5' Smchd1 L1710_M1884 aa del	cgagcttaaaaagaggcccagaagggacaaacttcggggaatggtcttcgg
3' Smchd1 L1710_M1884 aa del	ccgaagaccattccccgaagtttgtcccttctgggcctctttttaagctcg
5' BamHI SMCHD1 111 aa	cgcggatccacgaaagaacgaattgacttcttacc
5' BamHI SMCHD1 25 aa	atagatcgcggatcccacaggacggtgtacttgtttgatc
3' NotI SMCHD1 580 aa	acacttcgtgcggccgctcaccttaaaaagtgtaaattttatttg
3' NotI SMCHD1 702 aa	acacttcgtgcggccgctcattcatctccttcaggcc
3' SMCHD1 199 aa	ccaaaataggaaatatcactcacggcccagttgttaagctgtttag
5' SMCHD1 230 aa	gtacagcttaacaactgggccgtgagtgatatttcctattttggtgttggg
3' SMCHD1 200 aa	ggaaatatcactatttaaactgcgatacacggcccagttgttaagctgtttag
5' SMCHD1 226 aa	cttaacaactgggccgtgtatcgcagtttaaatagtgatatttcctattttgg
5' BglII MORC2 2 aa	acagaaagatctgctttcacaaattacagc
3' XbaI MORC2 603 aa	ctagtctagatcaagtggaaggtctggtgg
M13 reverse	caggaaacagctatgacc
5' pUC reverse (pProEX HTb)	agcggataacaatttcacacagg
3' pProEX reverse (pProEX HTb)	tcaggctgaaaatcttctctc
5' BacUp (pFastBac HTb)	ccggattattcataccgtc
3' BacDown (pFastBac HTb)	cctctagtacttctcgaca

2.7.3 shRNA oligonucleotides

Non-silencing	tgctgttgacagtgagcgatctcgcttgggcgagagtaagtagtgaagccacagatgtac ttactctctcgcccaagcgagagagtgcctactgcctcgga
SMCHD1 3'UTR_3 Forward	tcgagaaggtatattgctgttgacagtgagcgcttggttgtcactaccttgcaatagtgaagc cacagatgtattgcaaggtagtgacaaccaaatgcctactgcctcgg
SMCHD1 3'UTR_3 Reverse	aattccgaggcagtaggcatttggttgtcactaccttgcaatacatctgtggcttcactattgca aggtagtgacaaccaagcgctcactgtcaacagcaatataccttc

2.7.4 DNA-binding assays oligonucleotides

15mer ssDNA polyC	6-Fam-CCCCCCCCCCCCCCC
HS5-1b sense methylated	6-Fam-ATCTGCCACCTGGTGTTC*GA (* denotes methylation)

Chapter 3

Biochemical and structural characterization of full-length Smchd1

3.1 Abstract

Smchd1 plays important roles in epigenetic silencing and normal mammalian development. Recently, heterozygous substitutions in *SMCHD1* were found to contribute to two distinct conditions: FSHD and BAMS. Despite Smchd1's essential role in epigenetic regulation, its atomic structure and the molecular mechanisms underlying its function in both a healthy and diseased state remain to be elucidated. To provide a better understanding of Smchd1's molecular structure and function, I have successfully expressed and purified the mouse full-length 2007-amino acid protein. Follow-up electron microscopy analyses of the Smchd1 dimer revealed an elongated rod-like structure that displays a high conformational flexibility, similar to that of other structural maintenance of chromosomes (SMC) proteins. I have shown that Smchd1 co-purifies with DNA, which can be substituted for shorter, single-stranded DNA oligonucleotides *in vitro*. Immunoprecipitation and Mass Spectrometry studies revealed known interacting partners of Smchd1, in addition to several new putative interactors. However, these proteins could not be validated as Smchd1 binding partners in orthogonal assays.

3.2 Introduction

Previous studies have focused on investigating the individual domains of Smchd1: the N-terminal region containing the catalytic GHKL ATPase domain [61, 133], and the C-terminally located SMC hinge domain [62, 162]. Expressing the large 2007-amino acid full-length protein remained unsuccessful, furthermore hindered by the uncharacterized central region that connects the two terminal ends. To provide a complete understanding of Smchd1's function at the molecular level, I aimed to express and analyse the full-length Smchd1 protein. A comprehensive understanding of Smchd1 function relies on obtaining a molecular structure of the full-length protein, which will inform how the two protomers assemble into dimers, the dispositions of the component UBL, ATPase and hinge domains, and organisation of the uncharacterised intermediate domain that connects the ATPase and hinge domains. Such information will also help uncover the functional effects of various mutations in *SMCHD1* found in cohorts of patients suffering of FSHD and BAMS, thereby providing a foundation for structure-based drug design as a therapeutic treatment.

3.2.1 Preliminary studies

As part of my Honours research project, I assessed the effects of human *SMCHD1* missense mutations associated with either FSHD or BAMS, using an N-terminal GHKL ATPase-containing fragment of mouse Smchd1 (residues 111-702) to examine the biochemical properties of protein variants. *In vitro* ATPase analysis revealed that FSHD2-associated mutations in *Smchd1* generally perturb the ATP hydrolysis ability of recombinant protein, whereas enhanced ATPase activity was observed in several BAMS-associated variants [146] (Figure 3.1). Our collaborators have subsequently used a *Xenopus laevis* model of craniofacial development to study the functional effects of *SMCHD1* mutations. We previously used this model organism [46], where mRNA transcripts encompassing either wild-type or specific full-length SMCHD1 that include point mutations were injected into the two dorsal animal blastomeres of the eight-cell embryo. Western blot analyses confirmed protein expression of the introduced *SMCHD1* constructs incorporating BAMS-associated mutations G137E, W342S, N524S and R552Q [146] (Figure 3.2). Varying protein expression levels were detected for these variants, which may be explained by the decrease in protein thermal stability

observed for several of these mutants in the context of the SMCHD1 ATPase region [146]. Upon examination of these BAMS-associated mutations in the *Xenopus* model, we observed a dose-dependent small eye phenotype that partially recapitulates the BAMS phenotype in human patients [146] (Figure 3.2). A similar phenotype was achieved by overexpressing wild-type SMCHD1 in these embryos, suggesting these mutations behave as gain-of-function alleles [46]. In contrast, the injection of an FSHD2-associated mutation, Y353C, did not result in any phenotypic change regardless of dosage [146] (Figure 3.2). Because FSHD2-related mutations in *SMCHD1* are well-established to lead to a loss of overall function of the protein, these results further support the idea that BAMS-associated mutations exhibit a gain-of-function phenotype.

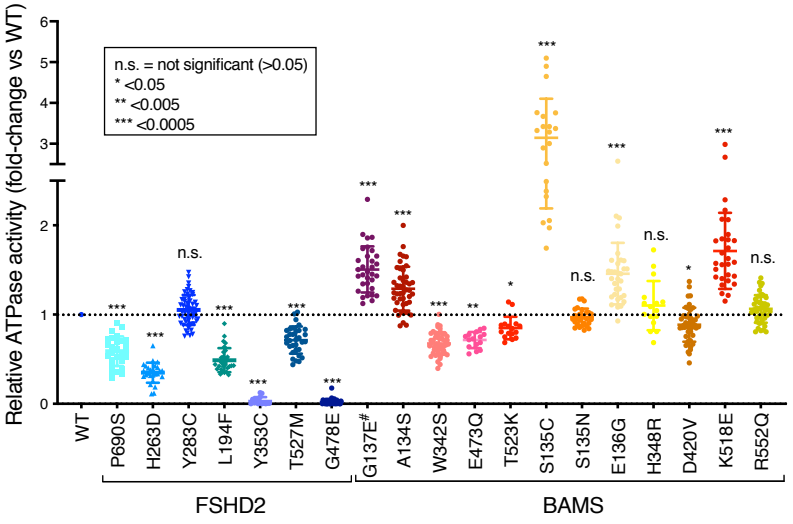


FIGURE 3.1: ***In vitro* ATPase analyses indicate FSHD-associated Smchd1 mutants display a loss of catalytic ability, whereas varied changes are observed across BAMS mutants.** Relative fold-change of in ATPase activity (y-axis) of all studied mutants compared to wild-type (WT) protein, where individual mutations are indicated in the x-axis. G137E is indicated with a number symbol as the mutation was reported in both a BAMS patient and an FSHD2 patient, and is included in the BAMS cohort for analysis purposes. -Fold change values were calculated by direct comparison of 16 points (four ATP and protein concentrations tested per mutant) obtained for each mutant, compared with the corresponding 16 values obtained for the WT protein within each assay performed. Statistical analysis was carried out using the Wilcoxon matched-pairs signed-rank test, additionally applying a correction for multiple testing. p values were obtained, as indicated in the outlined box. This figure is reproduced from [146].

Whether a change in ATPase activity accounts for SMCHD1's dysfunction in both BAMS and FSHD is unclear, particularly for BAMS where approximately half of the studied mutants exhibited a gain in ATPase activity, while the remaining showed either no change, or a loss in activity (Figure 3.1). These discrepancies observed in the

in vitro biochemical assays across BAMS mutants compared with their universal effect on *Xenopus* eye formation indicates they may behave differently in an *in vivo* context. In particular, analysis of the full-length SMCHD1 protein may be required to fully recapitulate mutation-associated changes *in vitro*. One of my aims was therefore to examine the effects of FSHD and BAMS-associated mutations in the context of full-length Smchd1. The main challenge I anticipated was the successful expression and purification of full-length Smchd1. Smchd1 is a large protein of 220 kilodaltons (kDa) that assembles into homodimers, and is predicted to be conformationally flexible. We have recently made advances in studying the structural and biochemical properties of Smchd1's two terminal domains: the C-terminal hinge domain (Chapter 4) [162] and the N-terminal ATPase (Chapter 5). However, the central region that links these two entities together consists of a ~1000 amino acid stretch of unknown function or tertiary structure. Structural studies of the full-length Smchd1 protein will therefore inform what lies in this central region, and overall contribute to our understanding about the molecular mechanisms that underlie its function.

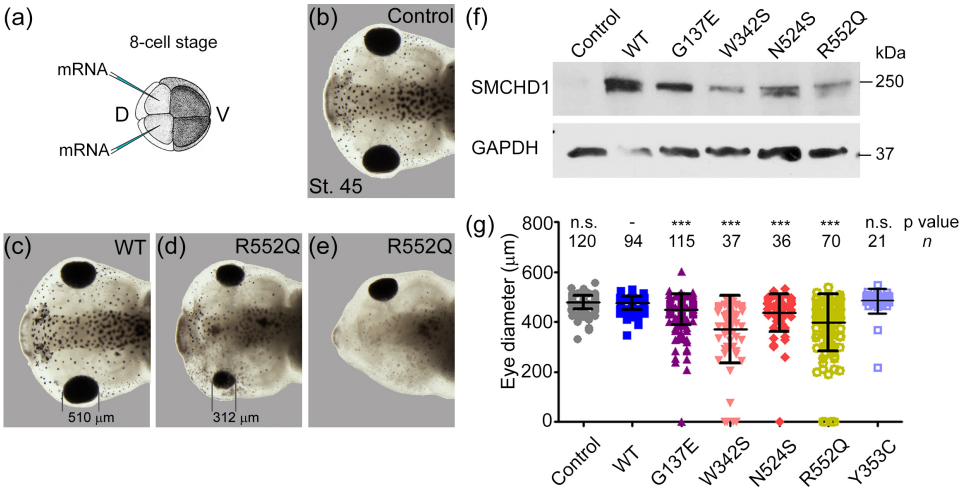


FIGURE 3.2: **BAMS-associated mutations in *SMCHD1* result in a decreased eye diameter in *Xenopus laevis*.** (a) Synthesized mRNA encoding full-length human SMCHD1 was injected into the two dorsal animal blastomeres at the 8-cell stage to target the head structures. (b-e) Representative images of stage 45 tadpoles that were (b) uninjected, (c) injected with WT mRNA, or (d-e) injected with R552Q mRNA. (f) Western blot showing that all injected mRNAs produced full-length SMCHD1 protein. Y353C mRNA has been tested previously [46]. (g) Measurements of eye diameter of tadpoles show that all BAMS-associated mutants cause a significant reduction in eye size. n indicates the number of embryos analyzed. Data are shown as mean values ± standard deviation (S.D.), and p values were calculated by one-way ANOVA followed by Dunn's post-test. n.s., not significant; ***, $p \leq 0.001$. This figure is reproduced from [146].

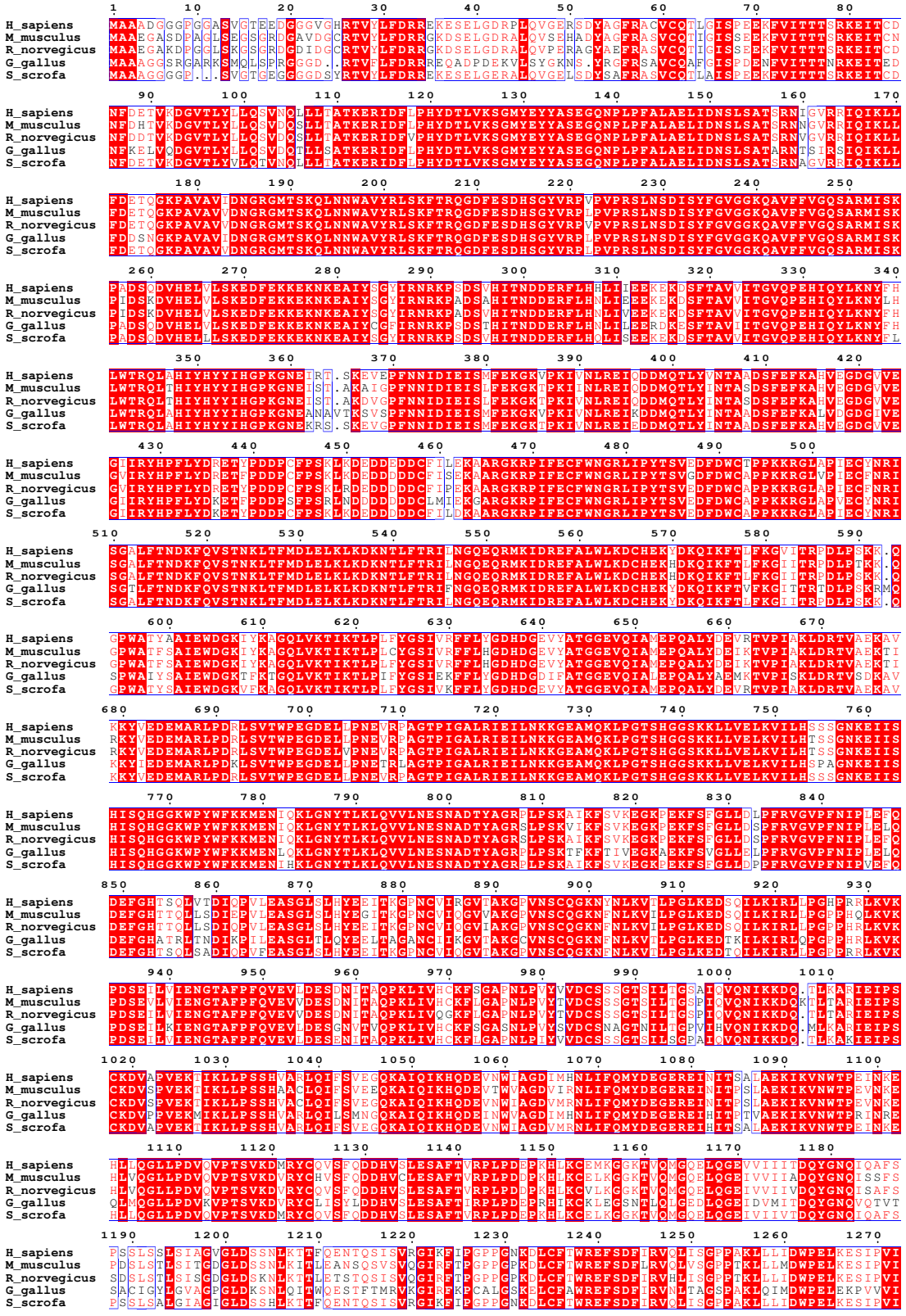
3.3 Results

3.3.1 The *Smchd1* gene is well-conserved among orthologs

The mouse *Smchd1* encodes a 2007-amino acid protein, compared to 2005 residues present in human SMCHD1. This large protein is ubiquitously expressed and well-conserved across mammals, sharing beyond 80% sequence identity (Figure 3.3). In particular, a high sequence conservation is highlighted within the two functional domains of SMCHD1: the N-terminal GHKL ATPase domain (core residues 121-395) and the C-terminal SMC hinge domain (core residues 1710-1884) (Figure 3.3).

3.3.2 Secondary structure prediction of the full-length Smchd1 protein

As mentioned above, the central region of full-length Smchd1 that links the ATPase and SMC hinge domains remains entirely uncharacterised, yet this region constitutes approximately half of the full-length protein. Using the secondary structure protein prediction software, PredictProtein [176], it is suggested the central region of Smchd1 is composed almost entirely of β-strands (Figure 3.4). This differs from canonical SMC proteins, where the central region linking the ATPase and SMC hinge domains is predominantly α-helical, forming long coiled-coil regions that are thought to enable their unique role in chromatin manipulation. In Smchd1, short coiled-coil regions are instead predicted to flank the hinge domain [62, 162], highlighted also by the predicted extended α-helical regions in Figure 3.4. Such a difference in gene architecture and protein arrangement between Smchd1 and canonical SMC proteins likely reflects important functional differences.


H_sapiens
M_musculus
R_norvegicus
G_gallus
S_scrofa

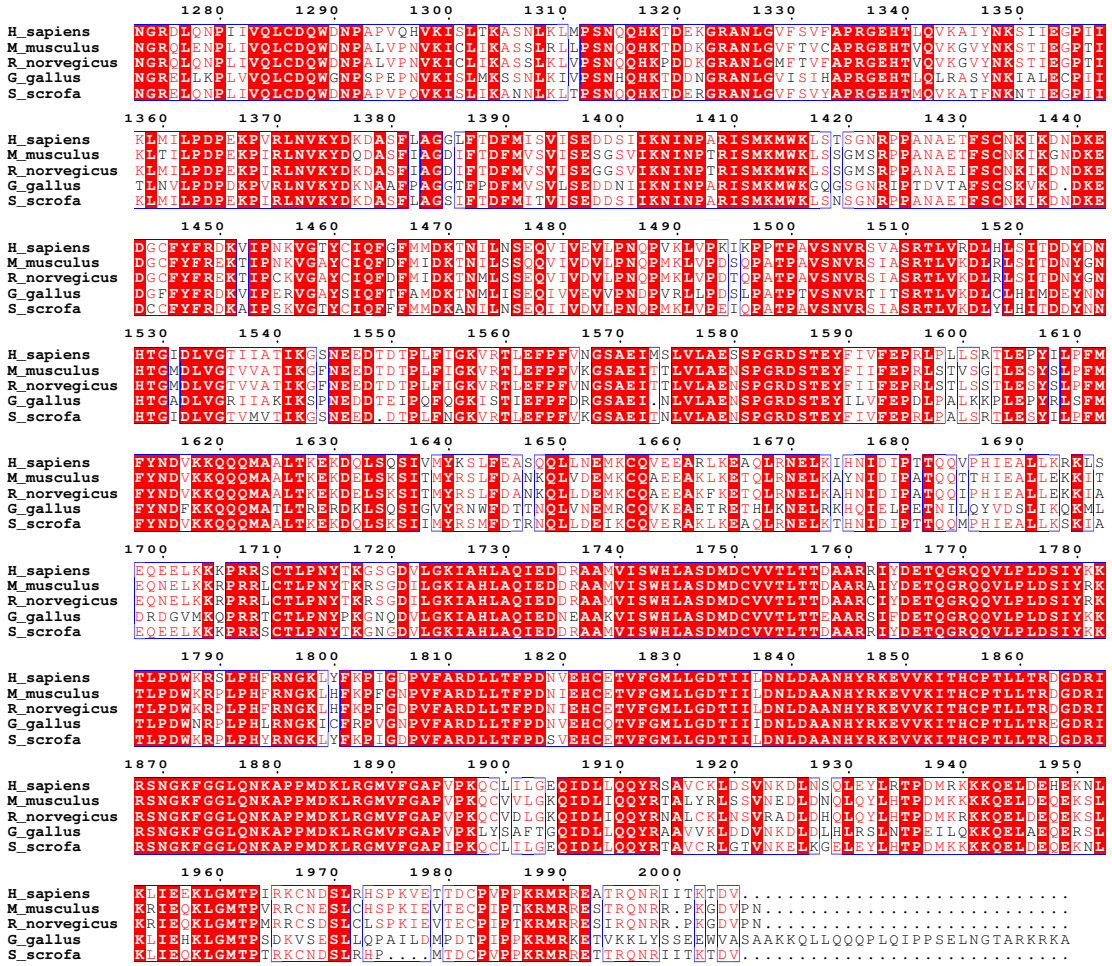


FIGURE 3.3: **Multiple sequence alignment of Smchd1 orthologs from selected species.** Multiple sequence aligment of human (*Homo sapiens*) SMCHD1 with ortholog sequences from mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus*) and pig (*Sus scrofa*). Identical residues conserved across the five species are shown in red boxes, whereas residues conserved in across only three or four species are in red text. The alignment was generated using the program MultAlin [174], and the image was generated using ESPript3.0 [175].

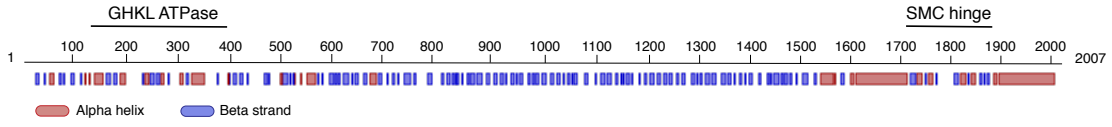


FIGURE 3.4: **Secondary structure prediction of the full-length Smchd1 protein.** Secondary structure of the full-length mouse Smchd1 protein was predicted using PredictProtein [176]. α-helices are shown in red and β-strands are denoted in blue. Amino acid sequence numbers are noted above, where the positions of the core GHKL ATPase (residues 121-395) and the core SMC hinge domain (residues 1710-1884) are indicated by a black line.

3.3.3 Full-length recombinant Smchd1 can be expressed in insect cells

The insect cell-based baculovirus expression vector system (BEVS) is a commonly used eukaryotic cell system largely due to its ability of introducing selected post-translational modifications and the presence of a protein-folding machinery, together yielding the production of highly authentic recombinant proteins. I previously used this system to express Smchd1's N-terminal region, comprising of either residues 111-702 or 25-702, as trials using bacterial cells proved unsuccessful. As this is likely due to the requirement of more extensive folding, and possibly the introduction of post-translational modifications, it indicates the expression of full-length Smchd1 likewise requires a eukaryotic cell expression system.

Initial 6-well plate expression trials of full-length mouse Smchd1 (∼225 kDa) proved successful, yielding optimal protein expression at the 36-hour post-transfection time point, with minimal degradation present in the soluble fraction (Figure 3.5 a). Follow-up harvest times of 36, 48 and 54 hours highlight similar protein expression levels in the soluble fraction, however increasing amounts are present in the cell pellet which indicates an increase in insolubility due to potential protein misfolding or aggregation. This test expression was performed in a non-shaking 6-well plate, whereas large-scale expression occurs under shaking conditions which accelerates the rate of infection and expression. Hence, for large-scale expression of full-length Smchd1, cell pellets were harvested 24-hours post-transfection rather than 36-hours to ensure the majority of protein remains soluble and to minimize protein degradation.

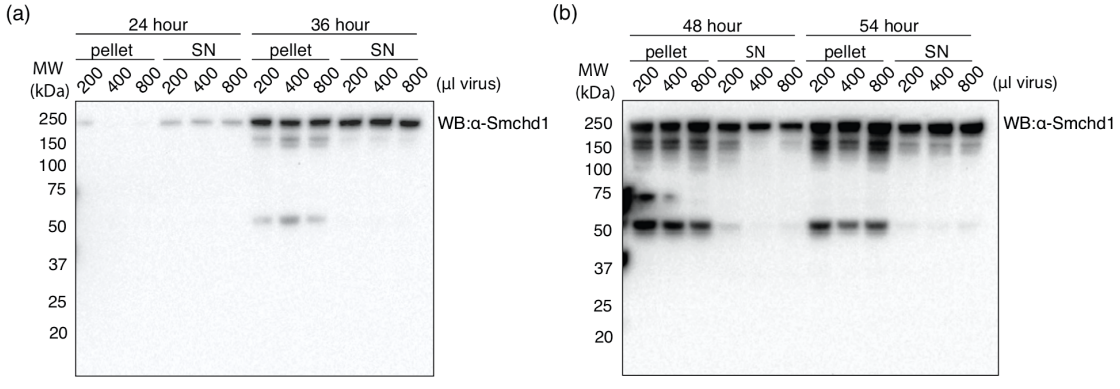


FIGURE 3.5: **Full-length recombinant Smchd1 is expressed in *Sf21* insect cells**. A time-course experiment monitoring the expression of full-length recombinant Smchd1 by Western blot at (a) 24 and 36 hours and (b) 48 and 54 hours post-transfection. Both pellet and supernatant (SN) fractions were analysed from insect cells that were transfected with various amounts of passage 2 (P2) virus, as illustrated. Molecular weight (MW) marker positions are indicated on the left in kilodaltons (kDa).

3.3.4 Purification of recombinant full-length Smchd1 protein

Immobilised metal-ion affinity chromatography (IMAC) was the first purification step performed, which uses the introduced 6x-histidine N-terminal tag on Smchd1 that is able to interact with a nickel-charged resin. As an initial trial, snap-freezing of cell pellets and cell sonication requirements were compared. For the purification of the N-terminal region of Smchd1, we incorporated five sonication cycles in the cell lysis procedure [146]. However, due to the larger molecular weight of full-length Smchd1 and its lower expression level compared to the N-terminal region, I examined the use of two or three sonication cycles instead. The addition of both snap-freezing and three sonication cycles compared to only two appeared to aid cell lysis as more full-length Smchd1 protein was released (Figure 3.6). Subsequent protein purifications of Smchd1 therefore included a snap-freezing step, and three sonication cycles.

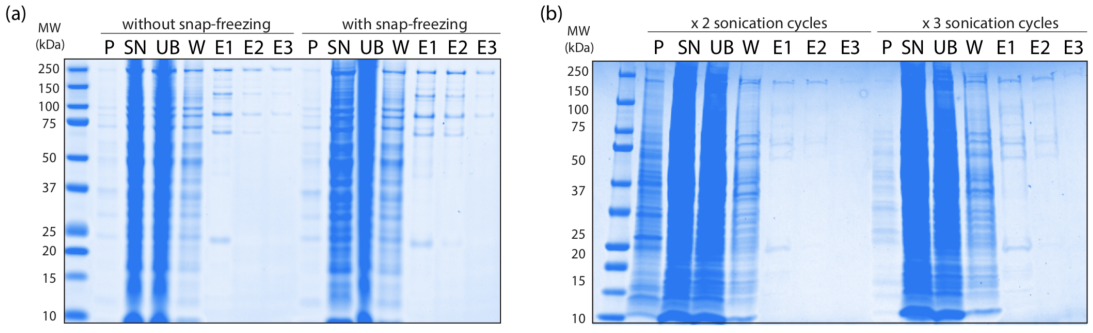


FIGURE 3.6: **Immobilized metal-ion affinity chromatography (IMAC) of full-length recombinant Smchd1 protein.** Pellets of insect cells expressing recombinant His6-Smchd1 protein were harvested and the effect of (a) snap-freezing (with or w/o = without) and (b) number of sonication cycles were varied to test the amount of recombinant Smchd1 protein released. IMAC was performed as a first purification step where samples were obtained at each step: P=pellet, SN= supernatant, UB= unbound, W= wash and E1-3= elution samples (1-3). Samples were analysed by reducing SDS-PAGE and the resulting gels were SafeStain-stained. A molecular weight (MW) marker is shown on the left in kilodaltons (kDa).

Subsequent size exclusion chromatography (SEC) purification increased the purity of Smchd1, where the protein elutes as the second major peak seen in the chromatography profile (Figure 3.7). The elution volume indicates a protein with a molecular weight of approximately 900 kDa, conflicting with the assumption of a dimeric Smchd1 that corresponds to approximately 450 kDa. However, this estimation relies on the overall shape of a protein, also known as the Stokes radius, and is mostly reliable for globular proteins. As Smchd1 is predicted to adopt an elongated structure, the protein would therefore display a larger Stokes radius which would likely cause it to elute in earlier fractions as indicated in Figure 3.7 (b). To further validate the identity of Smchd1,

Western blot analysis was performed on the eluted fractions (Figure 3.7 d). Fractions 27-32 were found to contain most Smchd1 protein and hence were routinely pooled in subsequent purifications. The protein was further concentrated to obtain a final sample in the concentration range of 1-3 mg/ml of protein.

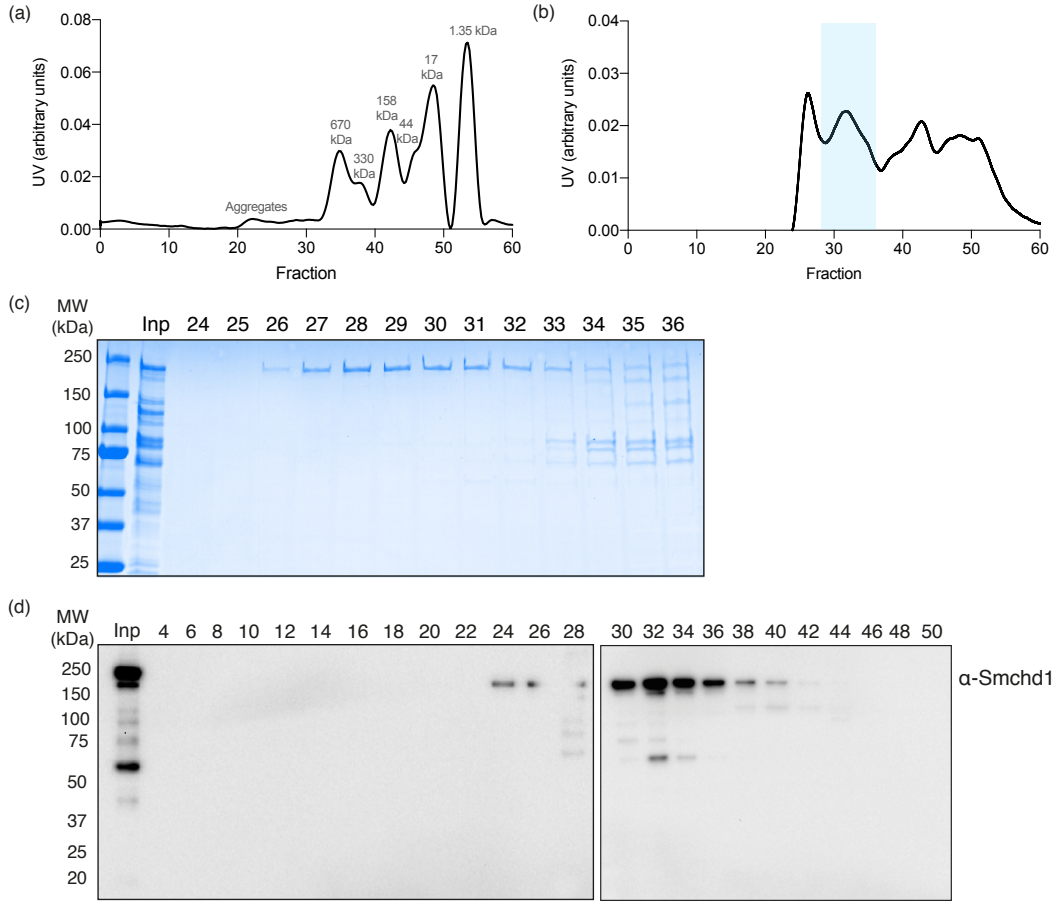


FIGURE 3.7: **Size exclusion chromatography (SEC) of full-length recombinant Smchd1.** SEC was performed on a Superose 6 increase 10/30 GL column for (a) a calibration sample with the protein standards (Thyroglobulin 670 kDa dimer and 330 kDa monomer, Aldolase 158 kDa, Ovalbumin 44 kDa, Myoglobin 17 kDa and Vitamin B12 1.35 kDa) indicated for the corresponding peaks in kilodaltons (kDa) and (b) the purification of full-length recombinant Smchd1, for comparison. (b) Smchd1-containing fractions are highlighted in blue which are also analysed by (c) a SafeStain reducing SDS-PAGE gel and (d) Western blot. (c-d) Molecular weight (MW) marker positions are indicated on the left in kDa.

3.3.5 Circular dichroism spectroscopy indicates purified Smchd1 is a folded protein

Circular dichroism (CD) spectroscopy is a technique that examines the difference in absorbance of right- and left-circularly polarized light of substances. Because different secondary structural elements of proteins carry varying light absorption properties, this

technique is commonly used to assess the presence of secondary structures within a protein sample, acting as an indicative measure of its folded state [177]. I performed CD spectroscopy on the purified full-length Smchd1 protein using an Aviv 410SF CD spectrometer. A spectral signature for α-helical content is indicated by negative bands at 208 and 222 nm. In the obtained results, a clear minimum at 208 nm was observed for the full-length Smchd1 protein, however it is difficult to distinguish whether a negative band is also present at 222 nm as this is very proximal to 218 nm band suggestive of β-strand content (Figure 3.8 a). Nonetheless, the overall CD spectrum obtained for Smchd1 is indicative of a folded protein.

While CD spectroscopy does not provide a quantitative estimation of secondary structure content, various databases can be used to provide comparisons to solved protein structures and their available CD spectra. I used the K2D2 online database [160] to analyse the CD spectra obtained for the full-length Smchd1 protein, which predicted an α-helical content of approximately 84% and β-stranded content of only approximately 1% (Figure 3.8 b). The remaining values to a 100% coverage may correspond to unstructured regions, such as loops. Overall, the experimentally-obtained curve appears comparable to the predicted spectrum obtained via the K2D2 database, as similar negative bands at 208 and 222 nm can be observed (Figure 3.8 b).

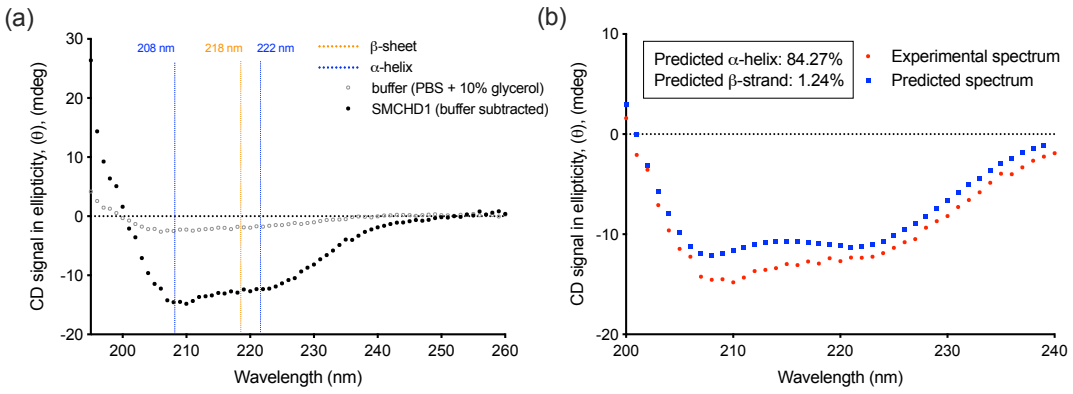


FIGURE 3.8: **Circular dichroism (CD) spectroscopy suggests purified full-length Smchd1 is a folded protein.** (a) Experimental data, indicating expected wavelength absorption for the corresponding secondary structures as per legend. The y-axis denotes the CD output signal, ellipticity (q), in millidegrees (mdeg), while the x-axis indicates the wavelength in nanometers (nm). (b) Experimental CD spectrum (red) compared with the predicted CD spectrum (blue) obtained from the K2D2 online database [160], showing the percentage prediction of secondary structure in the outlined box.

3.3.6 Examining the thermal stability of the recombinant full-length Smchd1 protein

To examine buffer conditions that promote the protein stability of full-length Smchd1, differential scanning fluorimetry (DSF) was performed at the CSIRO Collaborative Crystallisation Centre (C3) [178]. DSF is a means of investigating protein stability by determining the temperature at which a protein unfolds. This process is monitored by an increase in the fluorescence of an added dye with affinity for hydrophobic parts of the protein, which become exposed as protein unfolding proceeds [179]. The expected resulting fluorescence curve is sigmoidal, and a melting temperature (T_m) value can be obtained from the midpoint of the sigmoidal transition.

DSF was performed on purified full-length Smchd1 protein using 14 different buffers over a range of pH conditions, at high and low salt concentrations. The obtained melting curves lacked detectable unfolding transitions in all tested conditions (Figure 3.9). These observations are likely due to the multi-domain property of the full-length Smchd1 protein which is expected to result in multiple transitions during thermal denaturation, with transitions expected for each component domain. Rather than resulting in multiple peaks, the unfolding of distinct domains may coalesce into the observed flat melting curves. Overall, the obtained melting curves yielded unreliable melting temperatures for all tested conditions and DSF seemed an inappropriate measure of protein stability for the full-length Smchd1 protein.

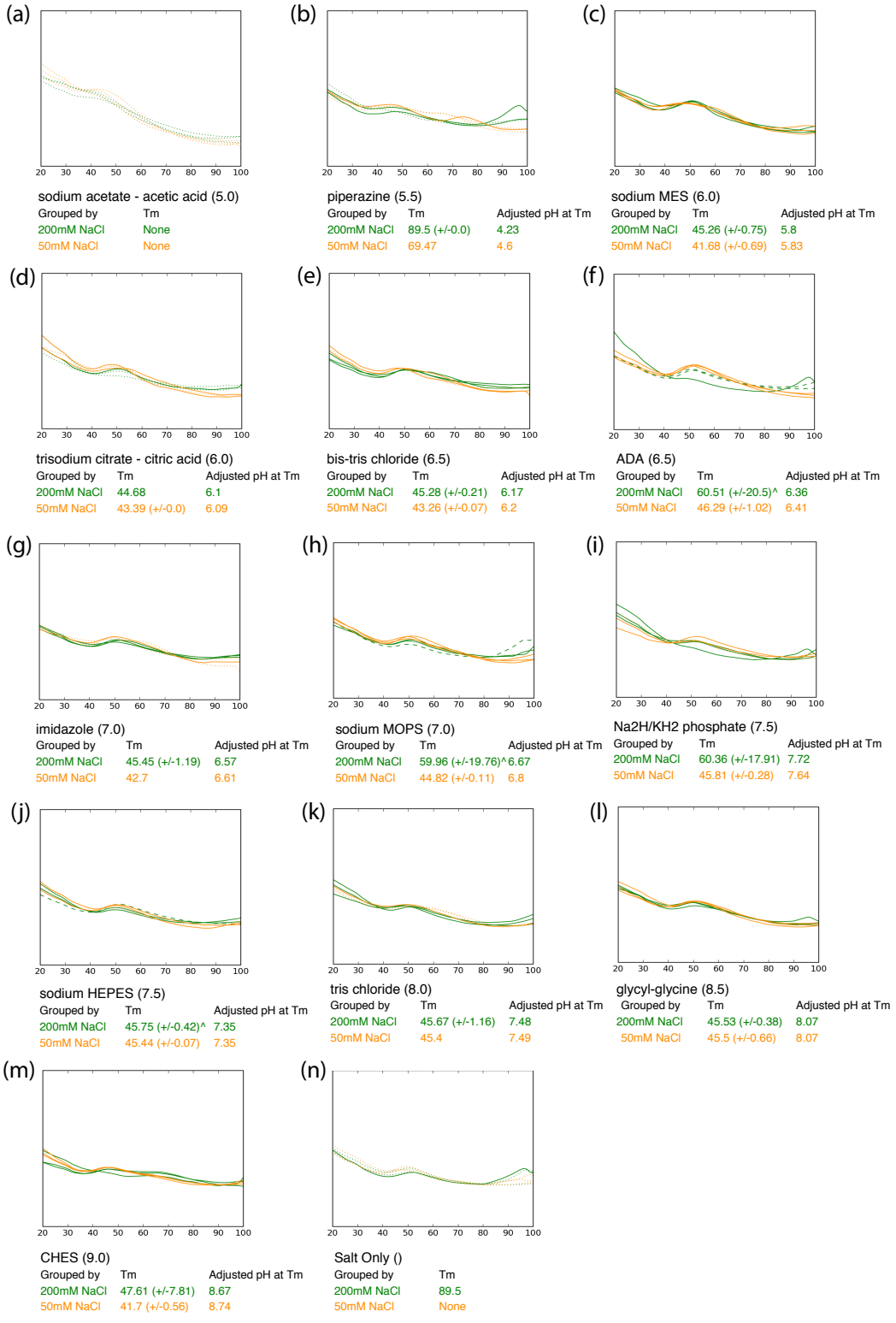


FIGURE 3.9: **Differential scanning fluorometry (DSF) of full-length Smchd1 under different buffer conditions.** (a-n) Melting curves of Smchd1 obtained under the indicated conditions. Temperature is indicated on the x-axis of each graph, in degrees Celsius. Outlier curves are excluded from melting temperature (T_m) calculations. Curves represented by dashed lines have unreliable T_m estimates.

3.3.7 Examining the catalytic activity of the full-length recombinant Smchd1 protein

Smchd1 consists of an N-terminal GHKL ATPase domain that has been previously established as catalytically active [61]. As a way of examining whether the purified full-length Smchd1 protein was likewise functional, I performed an *in vitro* ATPase assay using the same fluorescence polarisation system previously reported for the isolated N-terminal region of Smchd1 (residues 111-702) [61]. The assay was performed in an endpoint mode, where the reaction was terminated following a one-hour incubation of protein with the ATPase reaction substrates. In the presence of magnesium (Mg^{2+}), a co-factor required for the ATP hydrolysis reaction among all members of the GHKL ATPase family [99], full-length Smchd1 exhibits a catalytic rate of approximately 0.05 μM ADP/min/μM protein (Figure 3.10 a). This value is comparable to the catalytic rate obtained for the Smchd1 N-terminal GHKL region alone [61].

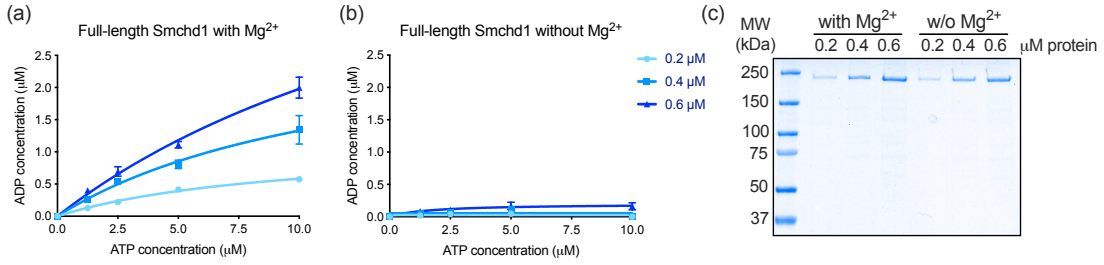


FIGURE 3.10: **Examining the catalytic activity of the full-length recombinant Smchd1 protein.** (a-b) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM) in both graphs, using full-length purified Smchd1 protein (a) with Mg^{2+} and (b) without Mg^{2+}, where reactions took place for 1 hour. Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent ± standard deviation (S.D.) of the mean. (c) Reducing SafeStain SDS-PAGE analysis of prepared Smchd1 protein dilutions used in the assay in (a-b), showing a molecular weight marker position on the left, in kilodaltons (kDa).

As formerly described, substitutions in *SMCHD1* underlie human conditions BAMS and FSHD. We previously examined the effects of several disease-associated variants that localise within Smchd1's ATPase domain by examining their ATP hydrolysis rates *in vitro*. However, these studies were limited to the context of the Smchd1 ATPase domain only (residues 111-702) [146]. The strongest gain-of-function missense variant we identified was the BAMS-associated mutation, Ser135Cys (S135C), which exhibited a ~3-fold increase in the ATPase activity of Smchd1 relative to wild-type protein (Figure 3.1) [146].

An additional missense mutation of interest in *Smchd1* is Ala667Glu (A667E), located in the C-terminal region of the extended Smchd1 ATPase domain. This mutation was identified via an ENU mutagenesis screen, closely following the discovery of Smchd1. A667E was recognised as an enhancer of variegation, thereby eliciting an enhanced repressive function in Smchd1 (manuscript in preparation). However, following *in vitro* ATPase analysis of this mutant in the context of a Smchd1 ATPase region construct, an ATP hydrolysis rate similar to that of wild-type protein was observed (Figure 3.11). These results provided initial evidence that the enhanced function of Smchd1 introduced by the A667E mutation may not be due an increase in ATPase activity.

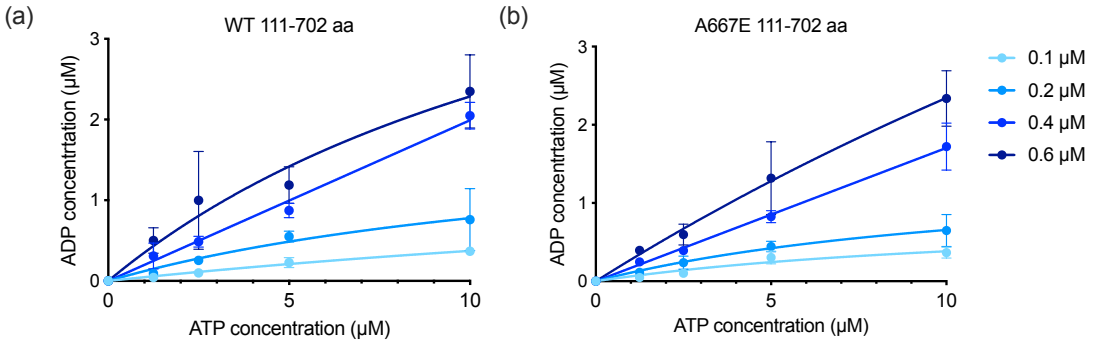


FIGURE 3.11: **Smchd1 variant A667E displays a comparable ATPase activity to wild-type protein in the context of the N-terminal region (111-702 aa).** (a-b) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM) in both graphs, using N-terminal region Smchd1 protein (111-702 aa) in (a) wild-type (WT) form or (b) the A667E variant; where ATPase reactions took place for 1 hour. Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent $\pm$ standard deviation (S.D.) of the mean.

I sought to determine whether SMCHD1 mutants S135C and A667E exhibit similar catalytic rates in the context of the full-length protein. As shown in Figure 3.12 (b), mutant A667E displays a comparable catalytic rate to wild-type protein, consistent with previous studies performed with the isolated ATPase region. Unexpectedly, variant S135C also displays a similar catalytic rate to wild-type SMCHD1 (Figure 3.12 e), in disagreement with the gain in ATPase activity observed for this mutant in the context of the Smchd1 N-terminal region [146] (Figure 3.1).

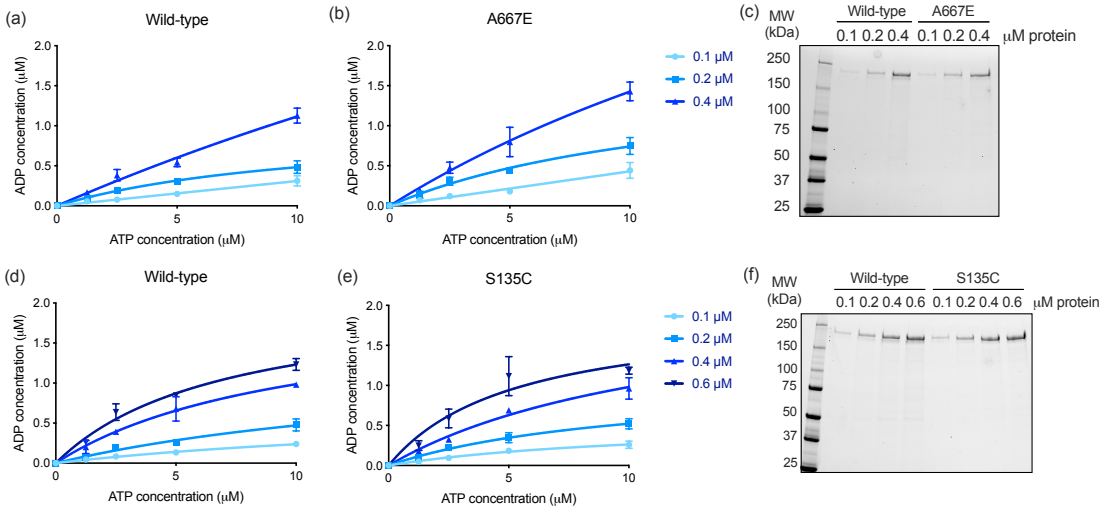


FIGURE 3.12: **SMCHD1 mutants A667E and S135C display a similar ATPase activity to wild-type protein in the context of the full-length protein.** (a,b,d,e) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM), using (a,d) wild-type full-length purified Smchd1 protein or (b,e) full-length SMCHD1 mutants, where all reactions took place for 1 hour. Experiments (a,b) and (d,e) were performed independently, and therefore they are to be compared only to each other. Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent ± standard deviation (S.D.) of the mean. (c,f) Reducing Stain-Free SDS-PAGE analysis of prepared Smchd1 protein dilutions used in the corresponding assays, showing a molecular weight marker position on the left, in kilodaltons (kDa).

In the competition binding fluorescence polarisation (FP) assay initially used, the ADP produced from the hydrolysis reaction can specifically interact with an ADP-antibody. A fluorescently-labelled ADP probe is also added to the reaction, which competes with and can become displaced from the antibody by the ADP produced from the hydrolysis reaction. The displacement of the labelled ADP probe results in a tumbling motion which can be detected as a decrease in FP value. Non-specific binding of the protein of interest to the labelled ADP probe can occur, an event that is more likely to happen in the presence of a larger molecular weight protein, such as full-length Smchd1.

To provide further confidence in the obtained results, I used a different ATPase assay system to examine whether similar trends in activity were observed. Rather than using fluorescence polarisation as a readout of ATPase activity, the ADP-Glo system detects ADP production using an indirect approach, where the ADP is converted to ATP, which is used to generate light in a luciferase reaction. Included in this analysis was an additional SMCHD1 mutant, E147A, which was previously shown to completely abolish the ATPase activity of Smchd1 in the context of the N-terminal region of the protein (residues 111-702) [61, 133]. As depicted in Figure 3.13, similar catalytic rates

were observed across the tested mutants as they all exhibit an ATPase activity comparable to that of wild-type full-length Smchd1. Surprisingly, this includes the E147A mutant where ATPase activity was expected to be completely abolished.

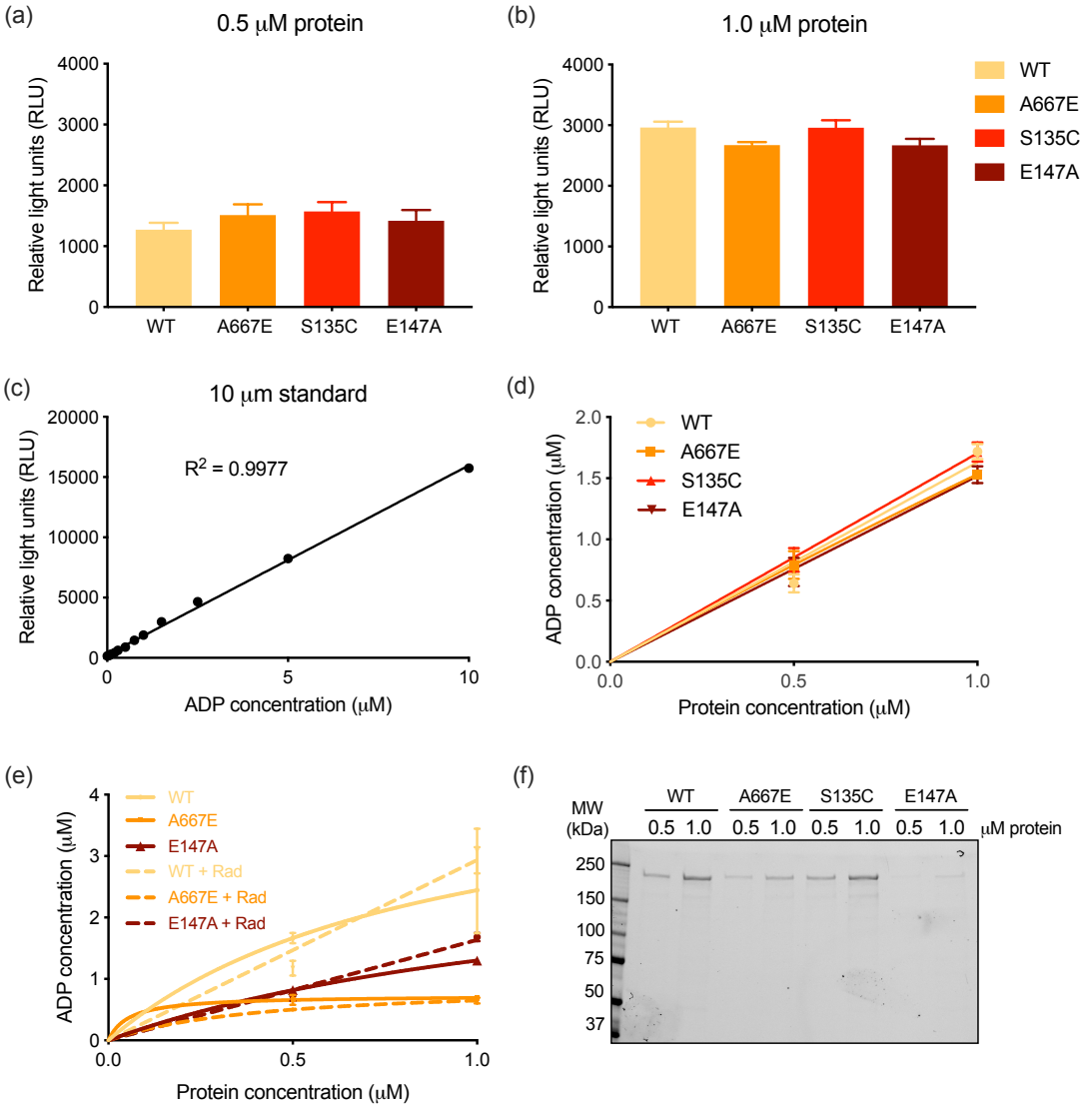


FIGURE 3.13: **SMCHD1 mutant E147A exhibits ATP hydrolysis activity in the full-length protein.** ADP-Glo assay results, using 10 μM ATP across all reactions and testing either (a) 0.5 μM or (b) 1.0 μM full-length Smchd1 protein. (a-b) The y-axis represents the relative light units (RLU) produced by the reactions, whereas the x-axis denotes the full-length Smchd1 protein tested, as depicted by the colour legend on the right. (c) A standard curve for the 10 μM ATP used was plotted for RLU units (y-axis) against amount of ADP produced (x-axis) to obtain (d) a representation of protein concentration tested (x-axis) and ADP concentration produced (y-axis) for all tested Smchd1 samples combined. (e) The same ADP-Glo assay was performed in the presence (dotted lines) or absence (solid lines) of 10 μM Radicicol, an inhibitor of Smchd1 ATPase activity, representing the amount of ADP produced (μM) at varying protein concentrations (μM). (f) Reducing Stain-Free SDS-PAGE analysis of prepared Smchd1 protein dilutions used in the ADP-Glo assay, showing a molecular weight marker position on the left, in kilodaltons (kDa).

To determine whether a protein contaminant may have compromised the assay results, I performed the same ADP-Glo assay on wild-type protein and mutants E147A and A667E, this time in the presence of an established Smchd1 inhibitor, Radicicol [61]. As shown in Figure 3.13 (e), comparable ATPase rates were observed for Smchd1 samples in the presence of Radicicol compared to absence of inhibitor, indicating all protein samples fail to be catalytically inhibited by Radicicol. Additionally, varying mutant-specific ATPase rates were observed which differ to the catalytic trends depicted in Figure 3.13 (d). Despite all protein samples being quantified via a spectrophotometer in preparation for the ATPase assays, it appears that while relatively comparable protein amounts are highlighted by SDS-PAGE for wild-type, A667E and S135C samples, mutant E147A is very faintly visible (Figure 3.13 f). Nevertheless, ATPase activity was still detected for the E147A mutant (Figure 3.13 d,e), suggesting the protein concentration measured for this mutant was overestimated. Taken together, these results suggest the presence of a protein contaminant may interfere with the ATPase analysis of full-length Smchd1 protein samples.

3.3.8 Negative stain electron microscopy (EM) of full-length Smchd1 reveals elongated, rod-shaped particles

Following the successful expression and purification of full-length Smchd1, I sought to perform structural studies on the obtained protein sample. As formerly discussed, Smchd1 is a large ~450 kDa dimeric protein, consisting of two known functional domains: an N-terminal GHKL ATPase and a C-terminal SMC hinge domain (Figure 3.14 a). While the connecting intermediate domain remains entirely uncharacterised to date, it is thought to be conformationally flexible, a property permissive of its anticipated functional role in chromatin manipulation. Considering these factors, X-ray crystallography seemed an inappropriate technique for structural studies on the full-length Smchd1 protein. Instead, I employed electron microscopy, which is a commonly used method for visualising the structure of higher molecular weight species.

I first performed Blue Native PAGE (BN-PAGE) followed by Western Blot analysis, a method that allows the visualisation of migrating native protein complexes to enable the identification of different oligomeric states of a protein. In BN-PAGE, complex migration depends on its tertiary conformation, hence the molecular weight marker is

largely not indicative of an accurate molecular weight of the protein of interest. Instead, by comparing the detected bands for the native Smchd1 protein sample to the denatured Smchd1 protein or lysate, it can be inferred the bottom band is indicative of a monomeric Smchd1, while the top band represents a dimeric Smchd1 (Figure 3.14 b). These results suggest the native full-length Smchd1 purified from insect cells exists as a dimer, thereby it may be assumed the images obtained upon negative stain EM are representative of dimeric Smchd1 particles.

Initial EM trials using full-length Smchd1 protein following IMAC and SEC purification steps revealed small aggregate structures upon negative staining (data not shown). I therefore incorporated an additional purification step, using either cation (Figure 3.14 c (i)) or anion exchange (Figure 3.14 c (ii)), at pH conditions of 9.5 or 5.5, respectively. A relatively poor protein yield was obtained, however the additional purification step provided a more homogeneous Smchd1 protein sample that was suitable for negative stain EM studies. The resulting EM images I obtained revealed elongated, rod-like particles of approximately 50 nm in length with small globular entities at either terminus of Smchd1 sometimes detectable (Figure 3.14 d (i,ii)). However, the visualised structures appeared to adopt varying conformational states, indicating protein flexibility. Because of the observed heterogeneity, follow-up 2D class averaging analysis and cryo-EM studies were unattainable.

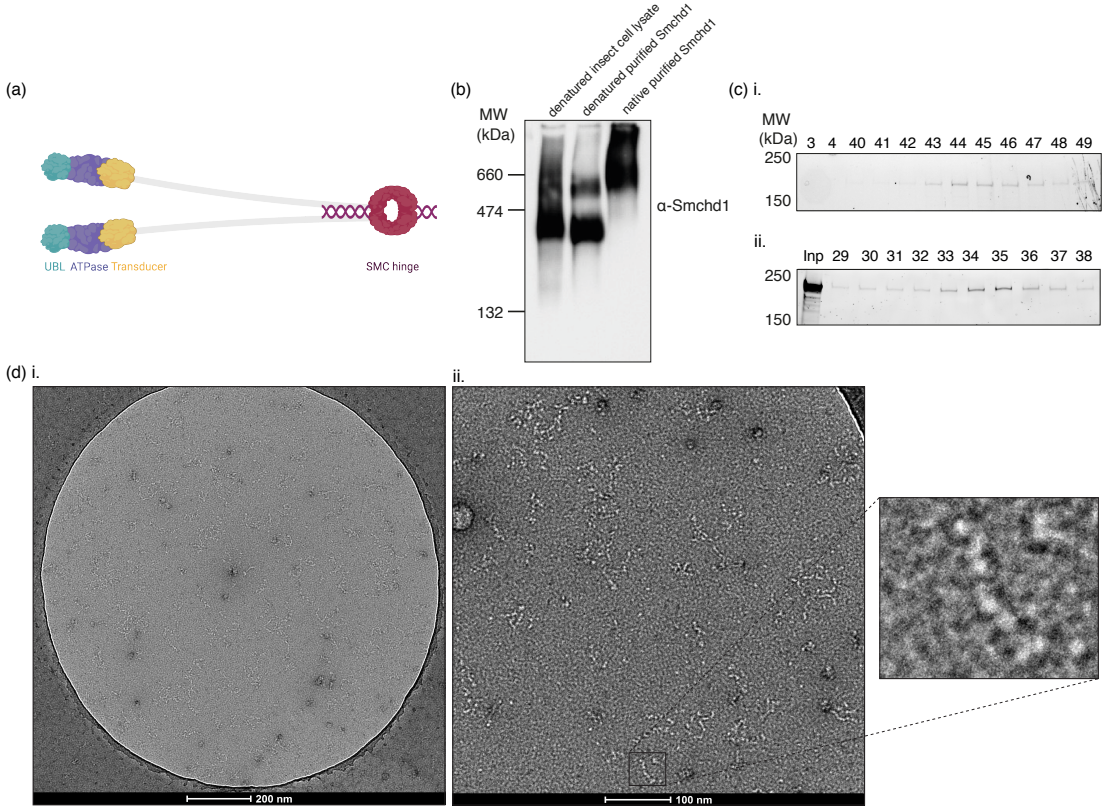


FIGURE 3.14: **Negative stain electron microscopy (EM) of full-length Smchd1 reveals elongated, rod-shaped particles.** (a) Schematic representation of the full-length Smchd1 protein. The ATPase region comprising of a ubiquitin-like (UBL) fold, a GHKL ATPase and a transducer domain, are located at the N-terminus, and the SMC hinge domain flanked by coiled-coil regions is located at the C-terminus of Smchd1. (b) Blue Native-PAGE (BN-PAGE) followed by Western blot analysis of insect cell lysate, compared to both denatured and native purified Smchd1 protein. Molecular weight (MW) marker positions are indicated on the left, in kilodaltons (kDa). (c) Eluting fractions analysed by reducing Stain-Free SDS-PAGE following a third purification step of full-length Smchd1 by either (i) cation exchange (MonoS column) or (ii) anion exchange (MonoQ column) chromatography. (d) (i-ii) Representative negative stain images of full-length Smchd1, with a zoomed-in imaged of one particle shown on the right.

3.3.9 Full-length recombinant Smchd1 co-purifies with DNA and Histone 3

Following SEC purification of full-length Smchd1, the eluting protein was predominantly present in the second peak, further validated by SDS-PAGE analysis (Figure 3.15 a,b). However, faint bands representative of full-length Smchd1 protein were also present in fractions ∼22-25, which correspond to the first peak of the SEC chromatography profile (Figure 3.15 a,b). This initial peak depicts the elution of aggregate proteins or species that are too large to be separated by the Superose 6 column used. Smchd1's presence in fractions ∼22-25 was an indication the protein may form larger, oligomeric complexes, potentially accompanying histone proteins which are the core components of chromatin.

I therefore analysed samples from SEC fractions 22-34 by gel electrophoresis to detect the presence of any co-eluting DNA. I additionally performed Western blot analysis on the equivalent fractions, probing for both Smchd1 and Histone 3 proteins. As revealed in Figure 3.15 (c), there was an abundant presence of DNA in fractions 23-26, which correspond to the first peak of the chromatography profile. Interestingly, the DNA species detected was of a consistent molecular weight of approximately 1.2 kilobases, likely due to sonication and therefore fragmentation of genomic DNA. Among fractions ∼27-31, corresponding to the second SEC peak where the majority of Smchd1 protein elutes, DNA is distributed as a smear of different-sized nucleic acids, indicating a heterogenous population of DNA. It therefore appears that the high salt concentration (500 mM NaCl) buffer used in SEC chromatography is unable to remove the contaminating DNA, indicating a tight interaction and its potential requirement for Smchd1's protein stability. The abundant nucleic acid presence in fractions 23-26 coincided with the high detection of Histone 3 in these samples following Western blot analysis, which was only faintly-detected in later fractions (Figure 3.15 d). These results suggest the first eluting peak following full-length Smchd1 chromatography may represent a high molecular weight species composed of chromatin-bound Smchd1, whereas the second peak comprises of Smchd1 that remains DNA-bound following chromatin shearing upon insect cell lysate sonication.

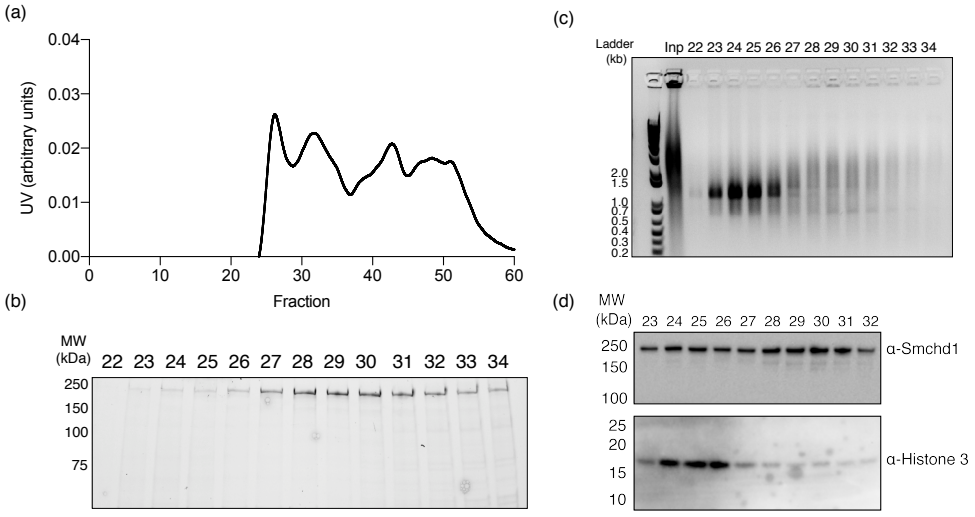


FIGURE 3.15: **Full-length Smchd1 co-purifies with DNA and histone 3.** (a) SEC profile of recombinant full-length Smchd1 and (b) corresponding selected fractions analysed by reducing Safestain SDS-PAGE. (c) Agarose gel (1%) of SEC fractions analysed in panel (b), with a DNA ladder indicating size in kilobases (kb) on the left. (d) Western blot analyses of SEC fractions against Smchd1 (top) and histone 3 (bottom). (b,d) Molecular weight (MW) marker positions are indicated in kilodaltons (kDa) on the left.

To provide a further insight into this hypothesis, I performed mass spectrometry analyses on fractions obtained from the first and second peaks of SEC chromatography on full-length recombinant Smchd1. As expected, Smchd1 was highlighted as the most abundant protein in both peaks (Figure 3.16 a,b), particularly in the second peak where majority of the protein elutes as indicated by SDS-PAGE (Figure 3.7 c). Histone proteins were likewise most abundant in the first peak relative to Smchd1, supporting my hypothesis that this population may represent a chromatin-bound Smchd1. I performed additional Smchd1 immunoprecipitation experiments, using anti-GFP resin to capture Smchd1-GFP from MEFs derived from Smchd1$^{GFP/GFP}$ mice (refer to section 3.3.10). I performed immunoblotting for selected proteins of interest that were identified by mass spectrometry studies (Figure 3.16 a,b). These included Smarcc1 and Smarca4, which are components of the SWI/SNF chromatin remodeling complex [180]; RYBP is a protein that is involved in mediating histone H2A ubiquitylation [181] and was additionally suggested as a Ring1A/B interactor [182]; and Trim37, which has been recognised as a histone H2A ubiquitin ligase [183]. As shown in Figure 3.16 (c-f), none of these proteins are present in the Smchd1-containing (bound, B) fraction, indicating they do not interact with Smchd1 under these conditions. While a faint band was observed for Smarcc1 (Figure 3.16 e), it is not of the expected molecular weight of ∼123 kDa and likely indicates a non-specific detection by the antibody.

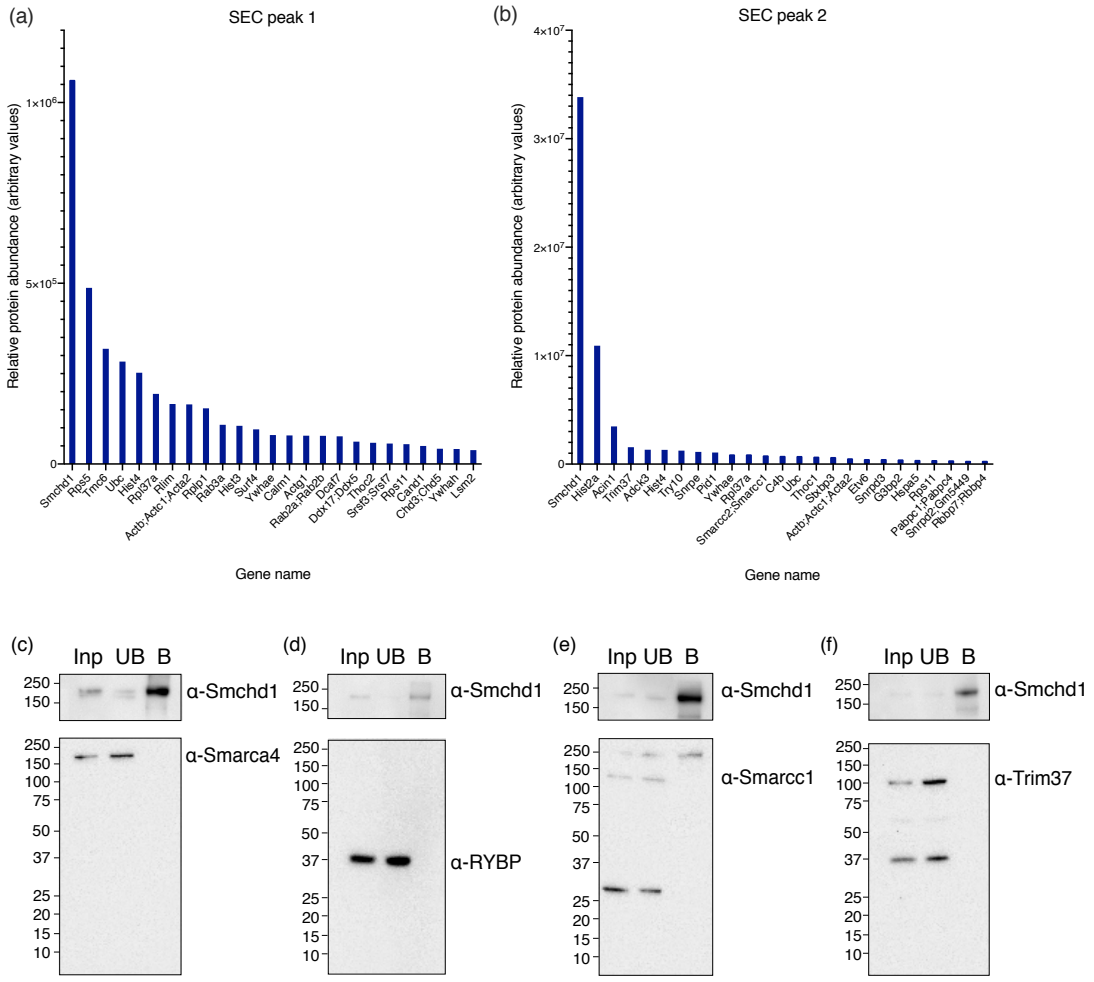


FIGURE 3.16: **Mass Spectrometry analysis of Smchd1 SEC chromatography fractions.** (a-b) A summary of identified peptides upon mass spectrometry analysis of fractions from the first and second peaks obtained upon SEC chromatography of full-length Smchd1, highlighting the 25 most abundant species. The x-axis represents the gene name of the identified peptides, whereas the y-axis denotes relative abundance of the peptides in arbitrary values. (c-f) Immunoprecipitation studies followed by western blot analyses testing a potential interaction between Smchd1 and (c) Smarca4, (d) RYBP, (e) Smarcc1 and (f) Trim37, denoting the input (Inp), unbound (UB) and bound (B) fractions. (c-f) Molecular weight marker positions are indicated on the left, in kilodaltons.

3.3.10 DNA-bound Smchd1 can exchange for shorter, single-stranded oligonucleotides

The finding of Smchd1's co-purification with DNA prompted further investigation into its nucleic acid-binding preference. Because Smchd1 was already associated with DNA, I performed a competitive electromobility shift assay (EMSA) where various concentrations of 6-Fam fluorophore-labelled oligonucleotides were incubated with the purified protein sample. EMSA experiments were previously performed for Smchd1's

isolated hinge domain, which revealed a slight preference for single-stranded DNA (ssDNA), particularly for poly-C oligopnucleotides [58]. I therefore compared the ability of DNA-bound Smchd1 to substitute for and interact with either ssDNA or dsDNA oligonucleotide probes. As shown in Figure 3.17, the full-length protein appears to exchange for a ssDNA oligonucleotide probe but not for dsDNA, as complex migration can be detected in the presence of 2 μM of a 15-base pair poly-C ssDNA probe. As the Smchd1 protein is likely purified bound to genomic dsDNA, these results do not suggest that there is no affinity for dsDNA, but highlight the inability to exchange for the dsDNA probe under these conditions. However, while full-length DNA-bound Smchd1 seems able to exchange for an ssDNA oligonucleotide probe *in vitro*, an almost equimolar amount of DNA relative to Smchd1 protein appears to be required for this event to initiate (Figure 3.17).

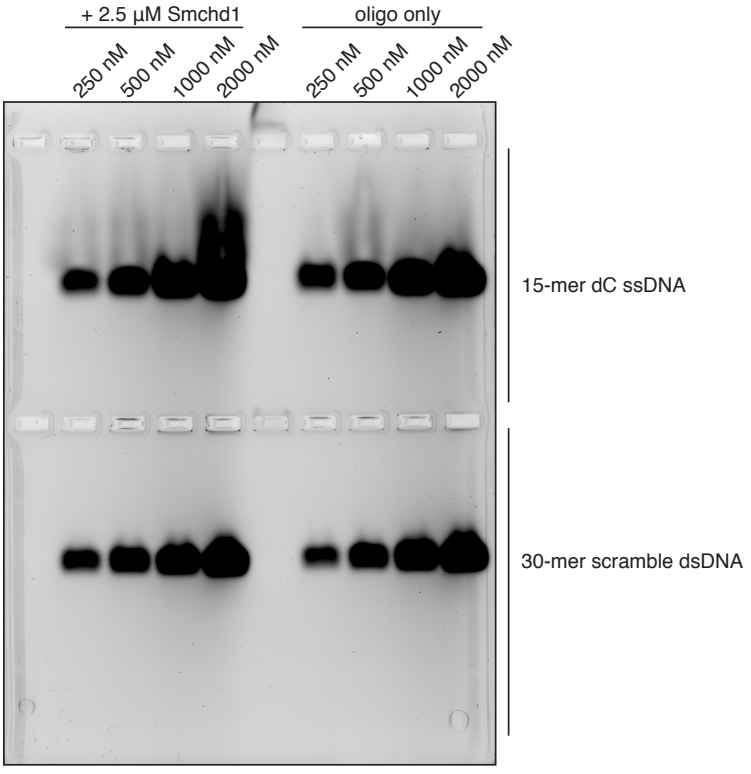


FIGURE 3.17: **Electromobility Shift Assay (EMSA) indicates DNA-bound Smchd1 can exchange for shorter, single-stranded DNA.** EMSA investigating whether full-length purified Smchd1 can interact with either (top) a 6-Fam-labelled 15-base pair poly-C single-stranded DNA (ssDNA) or (bottom) a 6-Fam-labelled 30-base pair scramble double-stranded DNA (dsDNA). The EMSA was performed either (left) in the presence of 2.5 μM of purified Smchd1 protein or (right) without Smchd1, acting as a negative control. Amount of oligonucleotide present in each reaction are listed at the top, in nM. The 0.5% (w/v) agarose gel was scanned on a Typhoon 9410 fluorescence scanner.

3.3.11 Immunoprecipitation followed by Mass Spectrometry studies reveal known interacting protein partners of Smchd1

Previous proteomics analyses revealed HBiX1, a heterochromatin protein 1 (HP1)-binding protein, and its mouse homologue, Lrif1, as protein interactors of Smchd1 [51, 131]. HBiX1 was previously shown to localize on the inactive X-chromosome (Xi), also a property of Smchd1, and both were demonstrated as essential in Xi compaction in human cells [131]. Using a similar experimental approach, I performed immunoprecipitation (IP) studies using female mouse embryonic fibroblasts (MEFs) obtained from Smchd1$^{GFP/GFP}$ mice generated on a C57BL/6 background, with the aim of identifying novel protein interactors of Smchd1. I used GFP-Trap beads (Chromotek), which have been established as highly effective in the pulldown of GFP-fusion proteins and are considered the gold-standard for this procedure, largely due to a high efficiency and binding affinity [184]. As shown in Figure 3.18, Smchd1-GFP (~250 kDa) was largely identified in the first elution (E1) sample following IP from Smchd1-GFP MEFs and was absent in the IP elution samples from wild-type (WT) MEFs, which were used as a negative control as Smchd1 lacks the GFP tag in these cells.

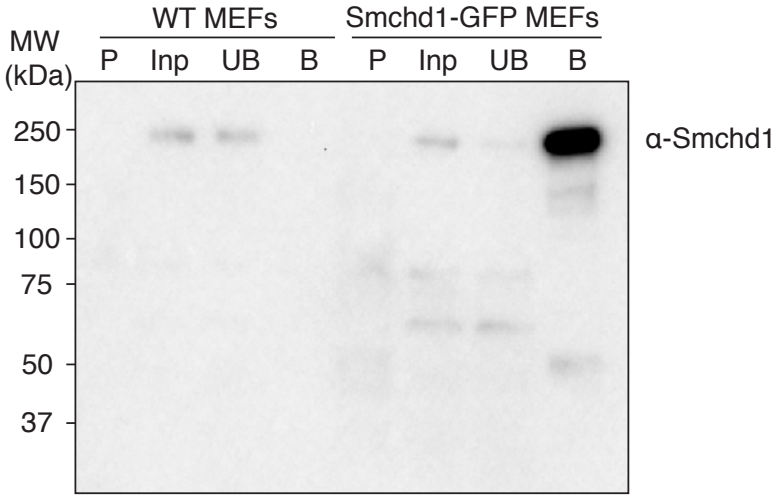


FIGURE 3.18: **Western blot analysis of Smchd1-GFP immunoprecipitation.** Smchd1-GFP immunoprecipitation (IP) experiments were performed for wild-type MEFs as a negative control, and for Smchd1-GFP MEFs as the test sample. Following cell lysis, samples from pellet (P), input (Inp), unbound (UB) and bound (B) fractions were collected and analysed by Western blot using an anti-Smchd1 antibody. Molecular weight (MW) marker positions are indicated on the left, in kilodaltons (kDa).

Following four independent immunoprecipitation experiments, E1 samples from each experiment (Figure 3.18) were trypsinized and analysed by LC MS/MS with the aim of identifying novel potential protein interactors of Smchd1. Peptides corresponding to a total of 2153 proteins were identified (Appendix A), from which 82 were considered of very high significance, and 218 of high significance (Figure 3.19 a,c). Both significant groups represent peptides that are enriched in the test sample compared to the negative control sample, with a p-value below 0.05, and are further distinguished by being more than two-fold (high significance) or four-fold (very high significance) enriched. P-values were obtained from a student's t-test comparing peptides arising from samples following anti-GFP immunoprecipitation of Smchd1-GFP MEFs with ones from WT MEFs, using a Benjamini-Hochberg correction. Highlighted in the volcano plots are Smchd1 and established protein interactors of Smchd1: Lrif1 (*Lrif1*) and HP1γ/α (*Cbx3/Cbx5*) [51, 131] (Figure 3.19 a,c). Displayed in Figure 3.19 (b,d) is the distribution of peptides arising from Lrif1 (*Lrif1*) and HP1γ/α (*Cbx3/Cbx5*). These indicate all identified peptides for the two proteins, despite very few, were found only in the Smchd1-GFP MEFs IP samples and not in the negative control samples from WT MEFs across the four experimental replicates, validating their specificity for Smchd1.

A summary of MS-identified proteins are listed in Table 3.1 as they are of broad interest due to either their involvement in epigenetic processes, or reported interactions with the *Xist* long non-coding RNA which is required for X-chromosome inactivation (XCI), in which Smchd1 plays a critical role. Examples of these are the heterogenous nuclear ribonucleoproteins (hnRNPs), which are a family of RNA-binding proteins. Many of these have been established as *Xist*-associated factors, in particular hnRNPK and hnRNPU [56], both of which are depicted as statistically significant (Table 3.1). In addition to hnRNP proteins, non-POU domain-containing octamer-binding protein (Nono), RNA-binding protein 3 (Rbm3), Splicing factor, proline- and glutamine-rich (Sfpq), RNA-binding protein 14 (Rbm14), RNA-binding protein FUS (Fus) and RNA-binding motif protein (Rbmx) (Table 3.1) represent some of the many components of paraspeckles, a nuclear domain that is formed through multiple RNA–protein and protein–protein interactions. While a complete understanding of their functional role is unclear, paraspeckles have been proposed to sequester proteins into nuclear bodies, which can modulate the local levels of active molecules and

influence gene regulation [185, 186]. Nono, Fus and Sfpq have been established as core components of paraspeckles [187], all of which show a statistically significant enrichment upon MS analysis (Table 3.1).

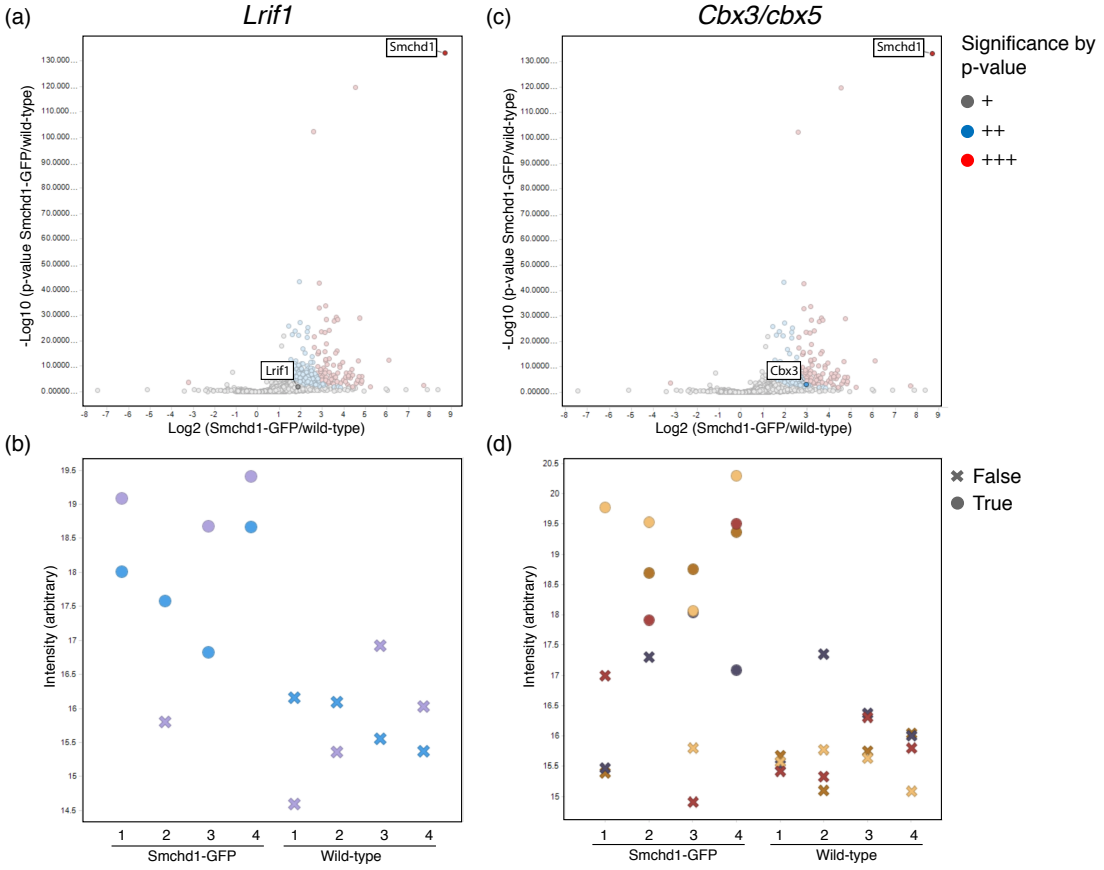


FIGURE 3.19: **Mass Spectrometry studies reveal known interacting protein partners of Smchd1.** Volcano plots highlighting the position of (a) *Lrif1* and *Smchd1* and (c) *Cbx3* and *Smchd1*. Other than Smchd1, the two most abundant identified peptides were Ahnak and Plectin (Appendix A), which are highly expressed proteins involved in cell-cell adhesion and the maintenance of the cellular cytoskeleton structure, respectively. As Smchd1 is found exclusively in the nucleus, the two proteins were therefore considered as non-specific interactors of Smchd1. (a, c) Both the x- and y-axes represent the log of the fold change between output peptides following anti-GFP immunoprecipitation (IP) from Smchd1-GFP MEFs and WT MEFs. The red dots represent the position of most statistically significant peptides, followed by blue and grey dots. (b,d) Plots showing independent peptides identified in the each of the four experimental replicates (x-axis) from Smchd1-GFP MEFs and WT MEFs for (b) *Lrif1* and (d) *Cbx3*. (b,d) Each coloured dot represents an independent peptide, and the crosses are imputed values for the corresponding peptides to allow for an enrichment comparison.

TABLE 3.1: **Selected proteins identified by Smchd1-GFP IP-MS.**

ID	Protein name	Gene name	P-value GFP/WT
37	Structural maintenance of chromosomes flexible hinge domain-containing protein 1	Smchd1	3.80E-137
258	DNA topoisomerase 1	Top1;TOP	1.56E-11
950	ATP-dependent RNA helicase DDX3X	Ddx3x	5.15E-10
992	Heterogeneous nuclear ribonucleoprotein U	Gm28062;Hnrnpu	1.74E-09
178	Nucleophosmin	Npm1;Gm5611	2.20E-09
1165	Heterogeneous nuclear ribonucleoproteins A2/B1	Hnrnpa2b1	9.80E-09
277	Splicing factor, proline- and glutamine-rich	Sfpq	6.52E-08
1166	Heterogeneous nuclear ribonucleoprotein A1	Hnrnpa1;Hnrnpa1l2;Gm5269	2.30E-08
369	Non-POU domain-containing octamer-binding protein	Nono	2.66E-08
562	RNA-binding protein 3	Rbm3	1.58E-07
520	Insulin-like growth factor 2 mRNA-binding protein 3	Igf2bp3	1.78E-07
1384	Nucleosome assembly protein 1-like 1	Nap1l1	3.05E-07
333	Heterogeneous nuclear ribonucleoprotein M	Hnrnpm	1.87E-06
557	Histone H2B	Hist1h2bj;Hist1h2bk;Hist1h2bm;Hist1	4.76E-06
388	Heterogeneous nuclear ribonucleoprotein Q	Syncrip;Hnrnpr;Gm20537	1.56E-05
1047	Heterogeneous nuclear ribonucleoprotein H	Hnrnph1	2.49E-05
613	Polypyrimidine tract-binding protein 1;Polypyrimidine tract-binding protein 3	Ptbp1;Ptbp3	5.31E-05
641	RNA-binding protein 14	Rbm14;p16;Gm21992	7.83E-05
892	Histone H4	Hist2h4;Hist1h4a	8.10E-05
722	Heterogeneous nuclear ribonucleoprotein F	Hnrnpf	0.000102878
851	Chromobox protein homolog 3;Chromobox protein homolog 5	Cbx3;Cbx5	0.000129311
1208	Heterogeneous nuclear ribonucleoprotein A3	Gm6793;Hnrnpa3;Gm9242;Gm17190	0.000209996
1161	Heterogeneous nuclear ribonucleoprotein K	Hnrnpk	0.000219252
1388	Heterogeneous nuclear ribonucleoprotein D0	Hnrnpd	0.000228358
462	Histone H1.4	Hist1h1e	0.000277644
413	DNA helicase;DNA replication licensing factor MCM5	Mcm5	0.000328396
1582	High mobility group protein HMGI-C	Hmga2	0.000621601
666	ATP-dependent RNA helicase DDX6	Ddx6	0.001287268
64	RNA-binding protein FUS	Fus	0.00278297
1550	Ligand-dependent nuclear receptor-interacting factor 1	Lrif1	0.002952603
364	Y-box-binding protein 3	Ybx3;Igf2bp3	0.00302205
1002	ATP-dependent RNA helicase A	Dhx9	0.004231441
1136	RNA binding motif protein	Rbmxl1;Rbmx	0.00608788
397	DNA helicase;DNA replication licensing factor MCM3	Mcm3	0.006811737
2	E3 ubiquitin-protein ligase TRIM32	Trim32	0.006993007
1444	Heterogeneous nuclear ribonucleoprotein A/B	Hnrnpab	0.008230024
464	Histone H1.5	Hist1h1b	0.008711881
381	Histone H2A	Hist1h2al	0.010044539
638	Poly [ADP-ribose] polymerase 1	Parp1	0.013194383

Following up on these results, I aimed to validate the interaction between Smchd1 and selected MS-identified proteins by performing co-immunoprecipitation (co-IP) experiments. I chose to examine the proteins: Nono, Rbm3, Nucleophosmin (NPM1/B23), Nucleosome assembly protein 1-like 1 (Nap1l1) and ATP-dependent RNA helicase (Ddx3). Nono and Rbm3 are paraspeckle proteins, while NPM1/B23 holds diverse roles in nuclear molecular mechanisms, including chromatin remodeling processes via histone protein interaction [188]. Likewise, Nap1l1 has been recognised as a chromatin modulator [189, 190], while Ddx3 has been suggested to function as a tumour suppressor as well as a transcriptional regulator [191, 192]. As depicted in Figure 3.20 (a), co-IP experiments for the selected proteins of interest revealed that while ample Smchd1 was detected in the input samples, none was present in elution samples. These results suggested the examined proteins do not directly interact with Smchd1. I also performed immunofluorescence experiments in Smchd1-GFP MEFs in order to examine the cellular localisation pattern of Smchd1 and the core paraspeckle protein, Nono, as an additional method of testing a potential interaction between these proteins. In this experiment, Smchd1 was endogenously GFP-tagged and mouse Nono protein was detected by an anti-mouse Nono antibody (Dr. Archa Fox, University of Western Australia). As depicted in Figure 3.20 (b), there was no co-localisation detected between Smchd1 and Nono, further suggesting the two proteins do not detectably interact in the nucleus.

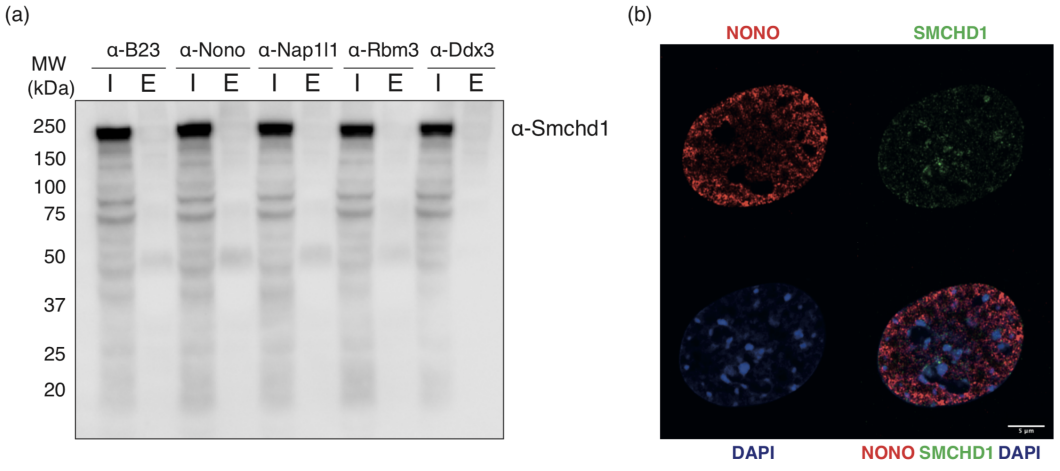


FIGURE 3.20: **Validation of potential Smchd1 protein interactors.** (a) Co-immunoprecipitation (co-IP) experiments using antibodies against B23, Nono, Nap1l1, Rbm3 and Ddx3 as bait, probing for Smchd1 as a potential interactor following Western blot analysis. Input (I) and elution (E) fractions were analysed in each co-IP experiment. Molecular weight is indicated in kilodaltons (kDa) on the left. (b) Immunofluorescence of Smchd1-GFP MEFs depicting Nono in red, Smchd1 in green and DAPI in blue. Images are representative of n=10 imaged nuclei.

3.4 Discussion

Serving as one of my main project aims, I have successfully expressed and purified the full-length mouse Smchd1 protein. I used a eukaryotic insect cell expression system that, in contrast to bacterial cells, is capable of introducing post-translational modifications and provides a complex protein folding machinery that is more similar to that of mammalian cells. Following two purification steps, a relatively pure Smchd1 protein sample was obtained, as indicated by SDS-PAGE analyses. The secondary structure prediction algorithm, ProteinPredict, revealed the intermediate domain of Smchd1 that connects the two functional domains to be predominantly β-stranded, contrasting with canonical SMC proteins that consist of long, α-helical coiled-coil arms. However, CD spectroscopy analysis of the purified full-length Smchd1 protein and follow-up quantitative predictions of secondary structure composition via the K2D2 software estimated an α-helical content of approximately 85%, with only 1% β-stranded content. The K2D2 prediction method relies on obtained CD spectra for proteins with a known tertiary structure, where the prediction accuracy may be estimated by the similarity between the experimental and the predicted CD spectra. Overall, the two curves appeared comparable in my analysis. However, the second minima obtained for the experimental CD spectra does not clearly distinguish whether it is indicative of β-stranded content at 218 nm, or α-helical content at 222 nm. In the obtained predicted spectrum this minima became accentuated towards the 222 nm wavelength, indicative of α-helical content, and likely providing an explanation for the final 85% α-helical predicted estimate of the Smchd1 protein. Recent publications have additionally suggested that protein CD spectra does not yield accurate predictive value for β-turns compared to α-helices [193], likely due to the heterogeneity of their conformation. Nonetheless, while the obtained secondary structure content predictions may not accurately reflect the composition of the full-length Smchd1 protein, the obtained CD spectra indicate the purified protein is folded. Further structural analyses of the full-length protein are therefore required to validate the secondary structure content of Smchd1's intermediate domain.

As a feature of all canonical SMC proteins, the C-terminally located SMC hinge domain is thought to act as the primary dimerisation site of Smchd1, a concept that is further discussed in Chapter 4. On the other hand, the N-terminal region of Smchd1 (residues

111-702) that comprises of the GHKL ATPase domain was established as monomeric in solution [61]. However, a new study that examined a different N-terminal construct of SMCHD1, comprising of residues 25-580, has revealed a dimeric high-resolution structure of the region [133]. These results suggest that dimerisation of full-length Smchd1 occurs at both termini, consistent with the rod-shaped particles I observed upon negative stain EM studies. ATPase assays performed with the N-terminal region of Smchd1 compared to ones performed in the context of the full-length protein are therefore likely to reveal differences in ATP hydrolysis rates due to a potential cooperative mechanism that may be induced by the dimerisation of the ATPase domains. Indeed, ATPase activity rates seemed to differ for various mutants of Smchd1 in the context of the full-length protein compared to the isolated ATPase region. However, rather than being able to attribute these differences to the dimeric state of the full-length Smchd1 protein, it seems likely that other factors are at play.

A feasible explanation is that a potential co-purifying protein contaminant may have interfered with the assay. This was speculated based on the inability of the established Smchd1 inhibitor, Radicicol, to abolish the ATPase activity of either the wild-type or any of the mutants studied. Additionally, the E147A mutant, which was previously shown to completely abrogate the catalytic activity of Smchd1's N-terminal region, remained catalytically active in the context of the full-length protein. Lastly, the ATP hydrolysis rate of Smchd1 is considered slow at $\sim$0.05 μM ADP/min/μM protein. Hence, even the minimal presence of a contaminating enzyme that is likewise able to hydrolyse ATP may easily interfere with the assay and result in a false positive output. I additionally showed that full-length Smchd1 co-purifies with DNA, indicating the *in vitro* ATPase assays I performed were in the presence of contaminating DNA. It is therefore possible that DNA-bound Smchd1 may also alter the ATP hydrolysis rates observed, which were dissimilar to rates obtained for the same mutants in the context of the N-terminal region of Smchd1.

The presence of bound DNA may additionally justify the requirement for a third ion exchange purification step to yield a homogeneous Smchd1 sample prior to negative stain EM. The obtained images revealed elongated, rod-like particles of approximately 50 nm in length that are indicative of dimeric Smchd1 proteins, as highlighted by Blue-Native PAGE analyses of the native protein sample. Unfortunately, the observed conformational flexibility of the sample hindered further analyses, such as

2D-classifications and averaging, where particles are separated into different classes based on their observed orientation. Alongside purifying as DNA-bound, I showed via Western blot analysis that Smchd1 additionally co-purifies with Histone 3, whereas follow-up mass spectrometry analyses confirmed the presence of other histone proteins in both the first and second eluting peaks obtained from SEC chromatography. These results may indicate a direct association of Smchd1 with nucleosomes, a hypothesis that remains to be further investigated. If, however, Smchd1 is a direct interactor of nucleosomes, a protein complex composed of purified nucleosomes and full-length Smchd1 may be subject to further electron microscopy studies. This would not only reveal their mode of interaction, but may stabilise the Smchd1 protein by reducing its conformational flexibility.

Following EMSA experiments, it appeared that a ssDNA oligonucleotide probe was able to substitute for bound DNA that accompanied purified recombinant Smchd1. However, nearly an equimolar amount of the oligonucleotide relative to Smchd1 protein was required to compete off the DNA-bound Smchd1. EMSA experiments are unable to reveal the fraction of the protein population exchanging for the oligonucleotide of interest. However, in order to ensure the full Smchd1 population undergoes this exchange so that a homogeneous sample can be obtained, a concentration exceeding equimolar amounts to Smchd1 protein is likely necessary. While this may be one way of reducing the heterogeneity derived from the co-purifying DNA, it seems unattainable for larger Smchd1 protein requirements such as for structural analysis. Nonetheless, the presence of a homogeneous nucleic acid template would be crucial to obtaining a less heterogeneous Smchd1 protein sample for further electron microscopy. As previously mentioned, one option may be to examine a nucleosome-Smchd1 protein complex, where the ∼147 base-pair nucleosomal DNA could provide a scaffold for Smchd1.

While it appears Smchd1 exhibits some affinity for histones, I sought to understand which other partners it may bind in cells. Paraspeckle proteins were abundantly found and suggested as potential interacting partners of Smchd1 upon immunoprecipitation and mass spectrometry analyses. However, following validation studies via co-immunoprecipitation and immunofluorescence experiments, none were successfully confirmed as interactors of Smchd1. Paraspeckle proteins are RNA-binding proteins, which are often non-specific interactors that arise upon immunoprecipitation studies

due to their 'sticky' nature, offering a potential explanation for their abundance. This family of proteins was recently suggested to undergo liquid-liquid-phase separation (LLPS), a process that describes macromolecules such as proteins or nucleic acids condensing into a dense phase that resembles liquid droplets [194]. LLPS is thought to underlie the formation of membrane-less compartments, such as heterochromatin [195, 196]. Phase separation was additionally suggested to drive the process of X-chromosome inactivation [197], in which Smchd1 holds a critical role. HP1 represents one of the two established protein interactors of Smchd1, for which I provided additional evidence here. As part of its molecular function, HP1 is recruited to histone H3, lysine 9 methylated sites (H3K9me) on chromatin, leading to the formation of compact chromatin, termed heterochromatin. Recently, HP1 was also shown to form phase-separated droplets *in vitro*, where it is proposed to oligomerise in order to compact chromatin into phase-separated condensates [195, 196, 198]. Together, these findings may suggest that Smchd1 may likewise be a component of phase-separated droplets.

In conclusion, I have provided biochemical analyses of the full-length Smchd1 protein, which suggested a requirement for DNA interaction as the bound DNA failed to be removed under the high salt conditions used in SEC chromatography. Follow-up studies will be targeted at ways to try and remove the contaminating DNA, and any potential proteins that may co-elute with Smchd1 and have likely interfered with the *in vitro* ATPase assays. One way of doing so may be to perform chromatography on the full-length Smchd1 protein in the presence of ATP. Across other studied GHKL ATPases, the ATP-binding event has been shown to induce ATPase dimerisation [99, 115]. This is followed by ATP hydrolysis and the subsequent dissociation of the dimer into a monomeric form. While any potential cross-talk between the ATPase and SMC hinge domains of Smchd1 remains to be elucidated, it is proposed that ATP hydrolysis may likewise result in dimer dissociation of the Smchd1 ATPase domain, resulting in an open, V-like conformation of the full-length protein as the dimer remains anchored at the C-terminal hinge domain. The described conformational change may release the bound DNA, facilitating the capture of a DNA-free Smchd1 protein sample.

Chapter 4

Structure and function of Smchd1's hinge domain

4.1 Abstract

The hinge domain of SMCHD1 has previously been implicated in nucleic acid binding and in the pathogenesis of FSHD2 when its DNA-binding capability is reduced. Yet, its underlying mechanism of function as well as overall structure remained unclear. We revealed the first atomic-resolution structure of Smchd1's hinge domain, providing a novel insight into its DNA-binding and dimerisation modes. Using structure-guided mutagenesis, I found that the central pore of the hinge domain dimer is not involved in nucleic acid interaction, and that DNA is not threaded through this region. As an epigenetic regulator, Smchd1 has been directly associated with chromatin association. I was therefore interested in further investigating the requirements of Smchd1's localisation to chromatin in cells. Using immunofluorescence techniques, I found that the hinge domain targets Smchd1 to chromatin, while additionally identifying a functional hotspot within the hinge that is required for chromatin localisation in cells.

4.2 Introduction

At the molecular level, Smchd1 occupies distinct regions across the genome where it can exert epigenetic silencing effects [52, 58]. It was recently proposed that Smchd1 mediates long-range chromatin interactions by acting as an insulator protein that sequesters chromatin from other DNA-binding proteins [52–54]. These functions are thought to be, at least partially, imparted by the hinge domain which has been shown to bind nucleic acids *in vitro* [58]. Additionally, the presence of an FSHD2-associated variant in *SMCHD1*, R1866G which is located within the hinge domain region, was shown to reduce SMCHD1's silencing capacity [199] as well as its ability to interact with DNA *in vitro* [58]. These findings highlight the functional importance of the hinge domain in SMCHD1's overall silencing capacity.

SAXS studies have previously been performed on the hinge domain. These studies indicated that in comparison to canonical SMC proteins, where the heterodimeric hinge domain is flanked by antiparallel intramolecular coiled-coils, Smchd1 adopts an intermolecular parallel coiled-coil arrangement [62]. However, the atomic-resolution structure of the hinge domain and the mechanism that underlies its interaction with nucleic acids remained unclear. We have recently reported the crystal structure of the mouse Smchd1 hinge domain, which represents the first high-resolution structure of a eukaryotic homodimeric SMC hinge domain [162].

4.2.1 The crystal structure of Smchd1's hinge domain reveals differences to canonical SMC proteins

The overall structure of Smchd1's hinge domain is similar to that of other SMC hinge domains, where two protomers assemble into a donut-shape and form a central pore (Figure 4.1 a,c). The *Thermotaga maritima* SMC hinge (*Tm*SMC) is the closest structural homolog of the Smchd1 hinge domain, as unlike other SMC proteins, it also forms a homodimer [162]. Common to all SMC hinge domains, monomers of Smchd1 and *Tm*SMC each comprise of subdomains I and II, which are connected by a flexible linker (Figure 4.1 b,d). In comparison to *Tm*SMC, Smchd1 lacks an N-terminal strand and two helices from subdomain I, and two β-strands from the subdomain interface. In subdomain II, a more extended helix is present at the interface side, alongside an

additional β-strand at the C-terminus (Figure 4.1 b,d). Overall, due to subdomain II of the Smchd1 hinge domain being rotated by 18° or 27° relative to the closed or open *Tm*SMC structure, respectively, the two subdomains II of Smchd1 are brought closer while subdomains I are placed further away (Figure 4.1 a,c).

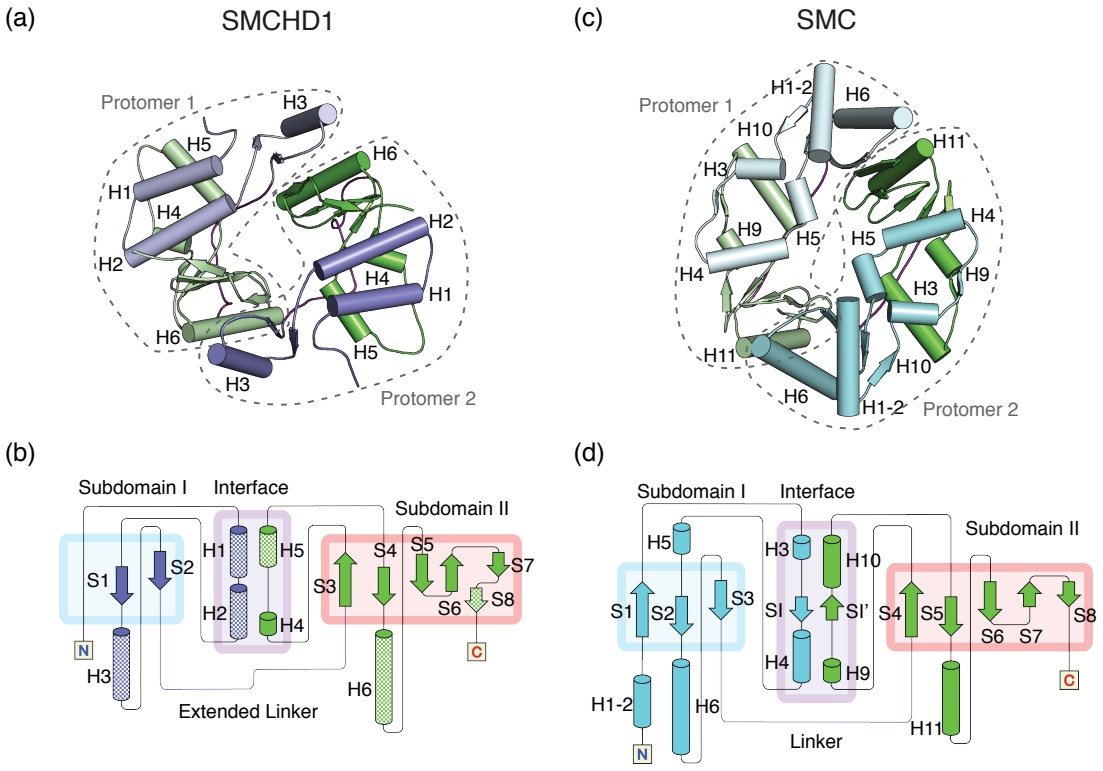
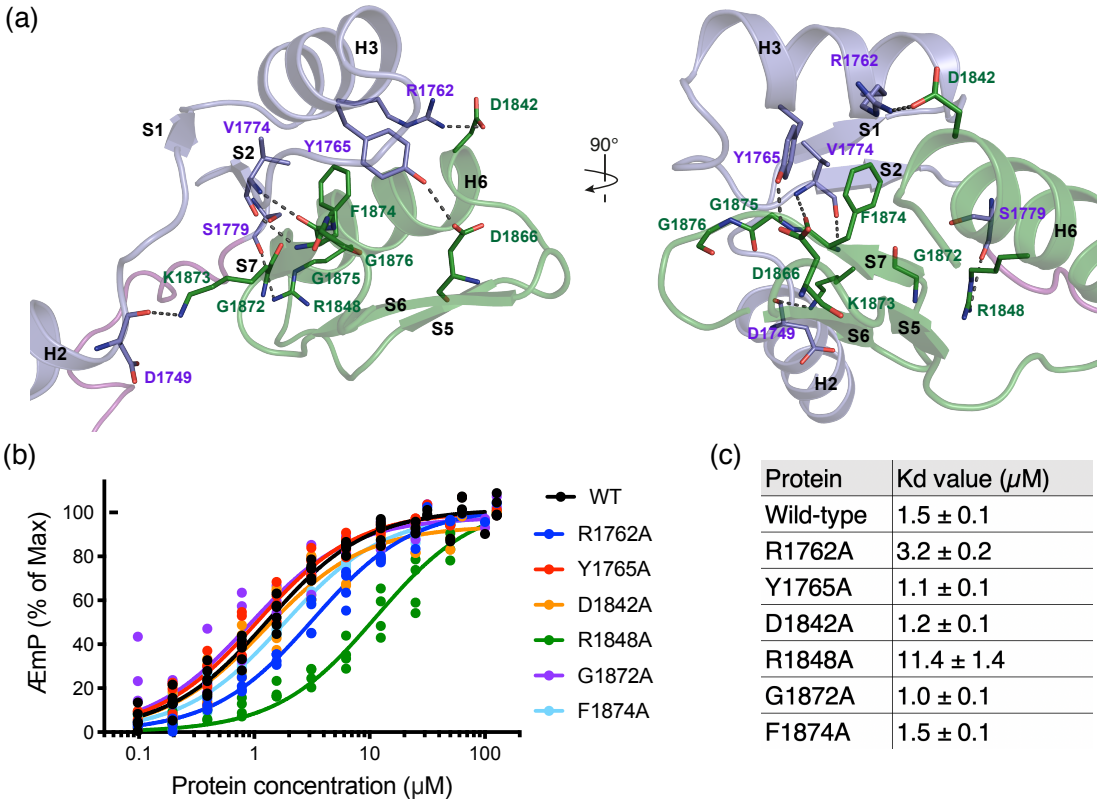


FIGURE 4.1: **The structure of Smchd1's hinge domain differs to canonical SMC proteins.** Comparing the cartoon representations of (a) the Smchd1 hinge domain (PDB:6N64) and (c) the canonical SMC hinge (PDB:1GXK) as seen from above, where individual protomers are highlighted by a dashed line. A comparison of (b) Smchd1 and (d) canonical SMC hinge domain from *T. maritima* topology, using the same colour scheme as in (a) and (c). The figure is reproduced from [162].

4.2.2 The Smchd1 hinge domain interface comprises of canonical and non-canonical interactions

Similar to canonical SMC proteins, dimerisation of Smchd1's hinge domain is mediated by two interfaces from each protomer. In Smchd1, the first interface is formed between H3 of subdomain I and H6 of subdomain II of opposing protomers (Figures 4.1 b, 4.2 a). A second interface is formed by β-strands S1 and S2 of subdomain I from one protomer, and S3-S8 of subdomain II from the opposing protomer. Aside from these, there are contributing non-canonical dimer interface interactions, including a hydrogen

bond between Y1765 in H3 and D1866 of S6, aliphatic interactions between Y1765 of H3 and F1874 of S7, and a salt bridge between K1873 of S7 and D1749 of H2 (Figure 4.2 a). Surprisingly, mutagenesis of dimer interface residues does not affect the overall structure of the hinge dimer but contributes to protein thermal stability [162]. In addition, interface residues R1762 and R1848 also appear to contribute to DNA binding as they display a reduced affinity for DNA *in vitro* upon their substitution to Ala (Figure 4.2 b).

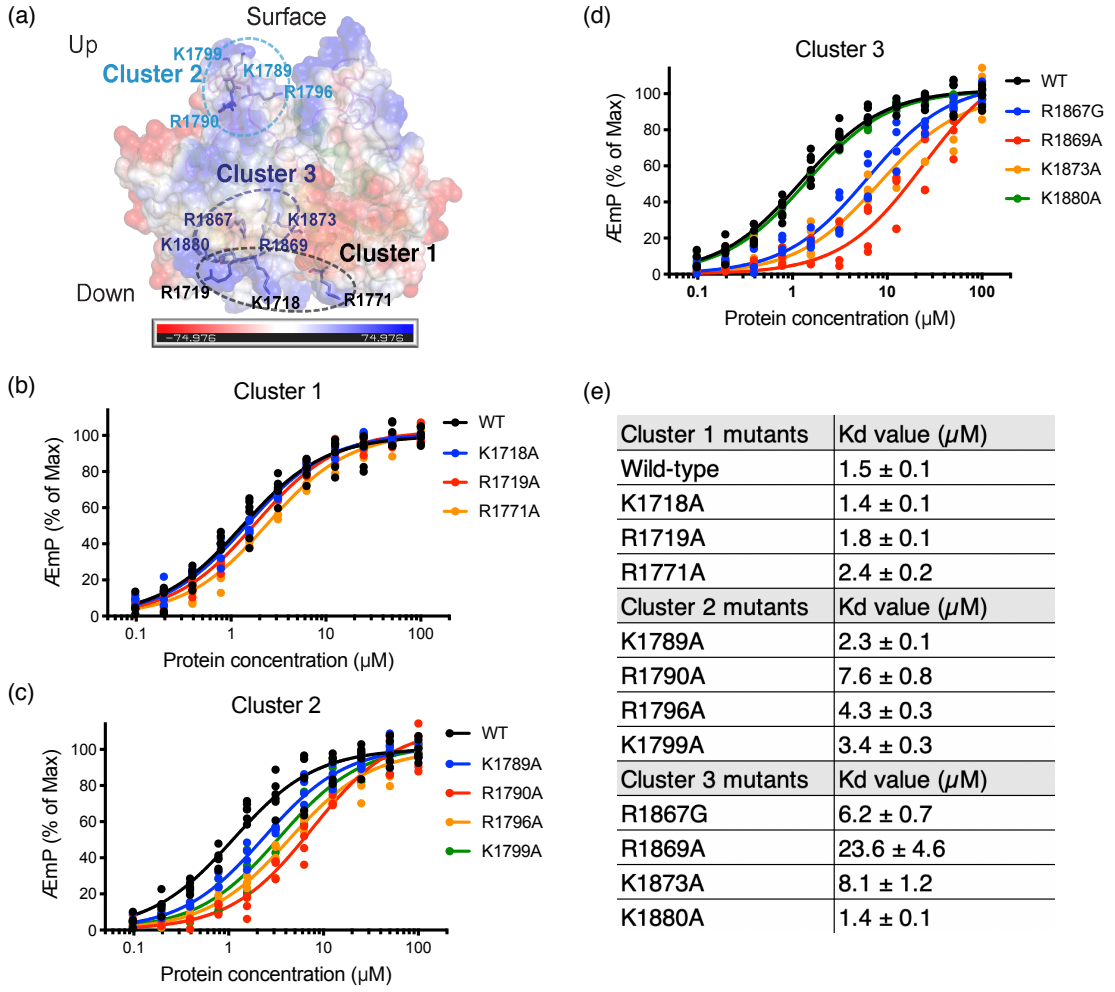


Protein	Kd value (μM)
Wild-type	1.5 ± 0.1
R1762A	3.2 ± 0.2
Y1765A	1.1 ± 0.1
D1842A	1.2 ± 0.1
R1848A	11.4 ± 1.4
G1872A	1.0 ± 0.1
F1874A	1.5 ± 0.1

FIGURE 4.2: **The Smchd1 hinge domain interface comprises of canonical and non-canonical interactions.** (a) A zoomed-in view of the Smchd1 hinge domain dimerisation interface, where subdomain I is shown in light purple and subdomain II is shown in green; a 90° rotation is shown on the right. (b) DNA-binding fluorescence polarisation assay testing mutants of the Smchd1 hinge domain dimer interface. Protein concentrations are indicated on the x-axis (μM) and polarisation values on the y-axis as a percentage of the maximum value. Plotted points represent technical replicates for two independent experiments fitted to a hyperbolic binding curve. DNA-binding assays were performed using a 20-base pair 6-Fam labeled single-stranded DNA corresponding to the HS5-1b sense-methylated enhancer site (2.7.4). (c) K_d and standard deviation values calculated from the fitted curves are outlined for each protein sample, in μM. Figure is reproduced from [162].

4.2.3 Smchd1's hinge domain mediates nucleic acid interactions via two positively charged patches

Upon solving the three-dimensional structure of Smchd1's hinge domain, the electrostatic surface potential of the region was revealed as highly basic in charge. This finding was not surprising as previous studies have shown that Smchd1's hinge domain has the ability to bind nucleic acids *in vitro* [58], a property shared across canonical SMC complexes where the hinge domain holds a key role in DNA interaction [75, 200–203]. The next question was therefore to examine which residues of the Smchd1 hinge domain were involved in nucleic acid interactions. By performing structure-guided mutagenesis, positively charged residues within three clusters of the Smchd1 hinge domain were substituted to Ala to help infer residues involved in nucleic acid binding. These clusters were separated into cluster 1, located on one face of the dimer, cluster 2 which consists of an extended linker, and cluster 3 surrounding the central channel of the dimer (Figure 4.3 a). Examining the DNA-binding affinity of these mutants, it was revealed that residues within cluster 2 and 3 contribute to DNA-binding (Figure 4.3 b-d), alongside dimer interface residue R1848 (Figure 4.2).



(e)

Cluster 1 mutants	Kd value (μM)
Wild-type	1.5 ± 0.1
K1718A	1.4 ± 0.1
R1719A	1.8 ± 0.1
R1771A	2.4 ± 0.2
Cluster 2 mutants	**Kd value (μM)**
K1789A	2.3 ± 0.1
R1790A	7.6 ± 0.8
R1796A	4.3 ± 0.3
K1799A	3.4 ± 0.3
Cluster 3 mutants	**Kd value (μM)**
R1867G	6.2 ± 0.7
R1869A	23.6 ± 4.6
K1873A	8.1 ± 1.2
K1880A	1.4 ± 0.1

FIGURE 4.3: **Smchd1's hinge domain interacts with DNA via two positively charged patches.** (a) A representation of the electrostatic surface potential of the Smchd1 hinge homodimer, with residues of interest highlighted and grouped in clusters based on their location. (b-d) DNA-binding fluorescence polarisation assays testing (b) cluster 1, (c) cluster 2 and (d) cluster 3 Smchd1 hinge domain mutants. Protein concentrations are indicated on the x-axis (μM) and polarisation values on the y-axis as a percentage of the maximum value. Plotted points represent technical replicates for two independent experiments fitted to a hyperbolic binding curve. DNA-binding assays were performed using a 20-base pair 6-Fam labeled single-stranded DNA corresponding to the HS5-1b sense-methylated enhancer site (2.7.4). (e) K_d and standard deviation values calculated from the fitted curves are outlined for each protein sample, in μM. Figure is reproduced from [162].

4.3 Results

4.3.1 DNA-binding residues of the Smchd1 hinge domain are not well conserved across SMC proteins

The crystal structure of Smchd1's hinge domain revealed the presence of highly basic surfaces, of which only clusters 2 and 3, but not cluster 1, were shown to contribute to nucleic acid interaction *in vitro* (Figure 4.3). Surprisingly, these residues are not well conserved across SMC proteins (Figure 4.4 a). Cluster 2 residue R1790 displays charge-conservative substitutions across human SMC2 and SMC4, alongside mouse SMC3 (Figure 4.4 a). Residue R1796, which is also located within cluster 2, is only present in human SMC4. These two residues reside within the extended linker of Smchd1's hinge domain, which exhibits the greatest heterogeneity compared to other SMC proteins. The length of this linker is highly variable across SMC hinge domains, but most abundant in positively charged residues in Smchd1, suggesting a potential role in nucleic acid interaction. Regarding cluster 3 residues, only K1880 appears to be conserved in human SMC4. Finally, the dimer interface residue R1848 that also contributes to DNA-binding in the Smchd1 hinge domain is only conserved in the *Tm*SMC protein. Between orthologs of Smchd1, all residues that contribute to DNA-binding *in vitro* display a high sequence conservation (Figure 4.4 b).

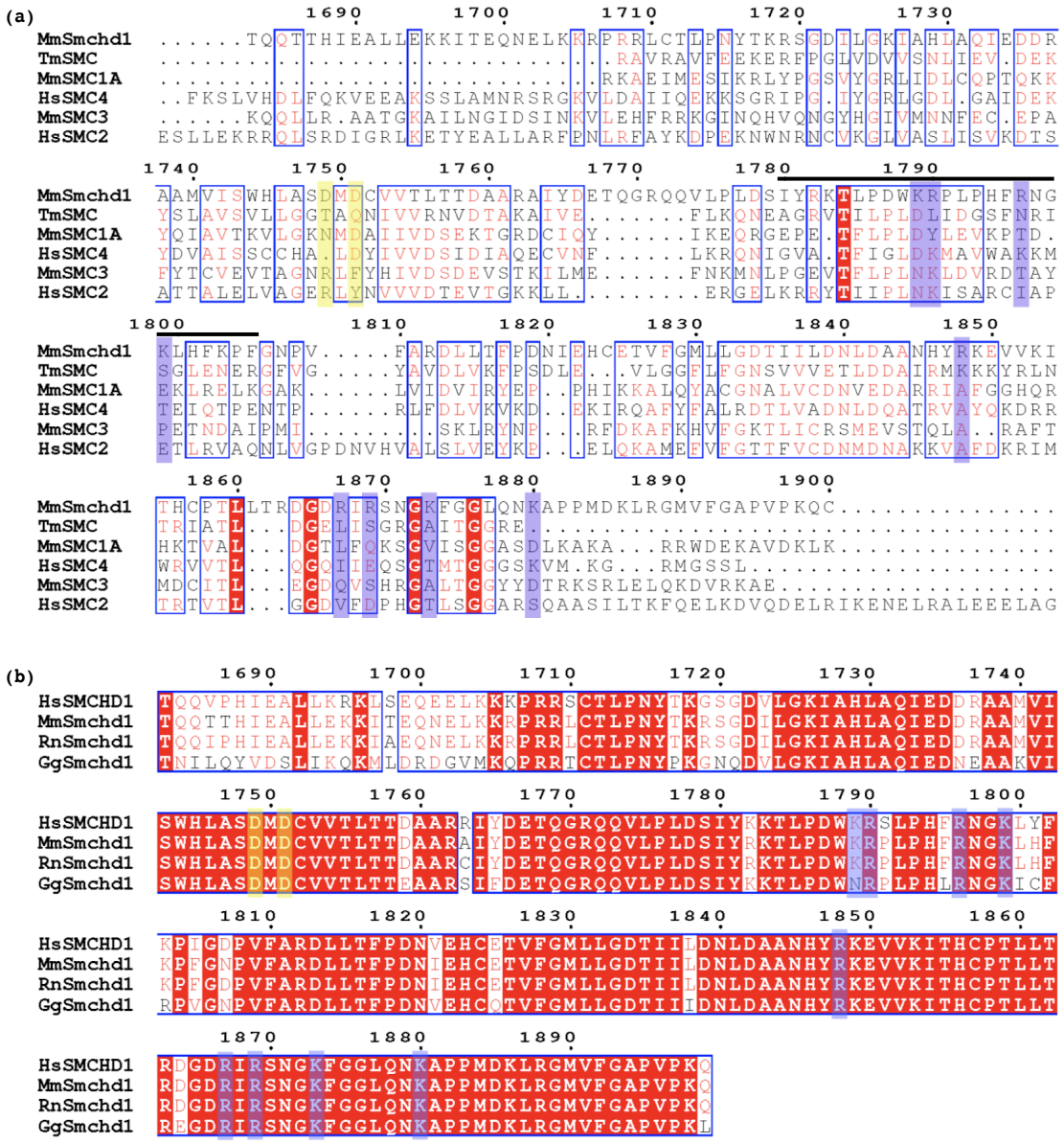


FIGURE 4.4: **Multiple sequence alignment of the Smchd1 hinge domain with orthologs and canonical SMC proteins.** (a) Multiple sequence alignment comparing the *Mus musculus* Smchd1 hinge domain (residues 1683-1899) with *Thermotaga maritima* SMC, *Mus musculus* SMC1A, *Homo sapiens* SMC4, *Mus musculus* SMC3 and *Homo sapiens* SMC2. The linker region of the mouse Smchd1 hinge domain is depicted by a black line. (b) Multiple sequence alignment of the Smchd1 hinge domain (residues 1683-1899) from *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Gallus gallus*. (a-b) Conserved residues are shown in red. Residues that contribute to DNA-binding *in vitro* from clusters 2 and 3, alongside dimer interface residue R1848, are highlighted in purple. Residues D1749 and D1751 that are mutated in FSHD2 are highlighted in yellow. Sequences were aligned using the program MultAlin [174] and images were generated using ESPript3.0 [175].

4.3.2 Expression and purification of Smchd1 hinge domain torus mutants

All SMC hinge domains studied to date have been shown to adopt a donut-like structure with a central cavity formed between two hinge domain protomers [204]. A highly conserved feature of the central channels of SMC hinge domains is the abundance of positively charged residues that line the lumen region [205]. In cohesin's SMC1-SMC3 heterodimer, mutation of three positively charged residues within the hinge domain lumen were shown to abolish DNA loading, but did not affect cohesin's recruitment to chromatin [202]. I therefore sought to examine whether nucleic acid binding might also occur via the torus of the Smchd1 hinge domain homodimer. To do so, I introduced mutations that would occlude DNA from entering the central pore formed by the dimer. Residues selected for substitution were S1870, which does not engage in crucial interactions based on the three-dimensional structure, and H1856, which lies on the rim of the central pore (Figure 4.5 a). Met, Asn or Gln were introduced in place of S1870, and Trp was substituted for H1856 as a means of obstructing the central channel via the presence of bulkier residues. Purification of the four torus mutants resulted in a high protein yield (Figure 4.5 b) and follow-up SEC analysis indicated they all retain a dimeric state, with an elution profile that corresponds to the wild-type counterpart (Figure 4.5 c (i)). Smchd1 hinge domain monomers can be identified as the lower molecular weight band by reducing SDS-PAGE analysis, where the presence of a higher molecular band across fractions ~43-45 are likely representative of contaminating TEV protease which was not entirely removed upon the preceding subtractive IMAC purification step (Figure 4.5 c (ii)).

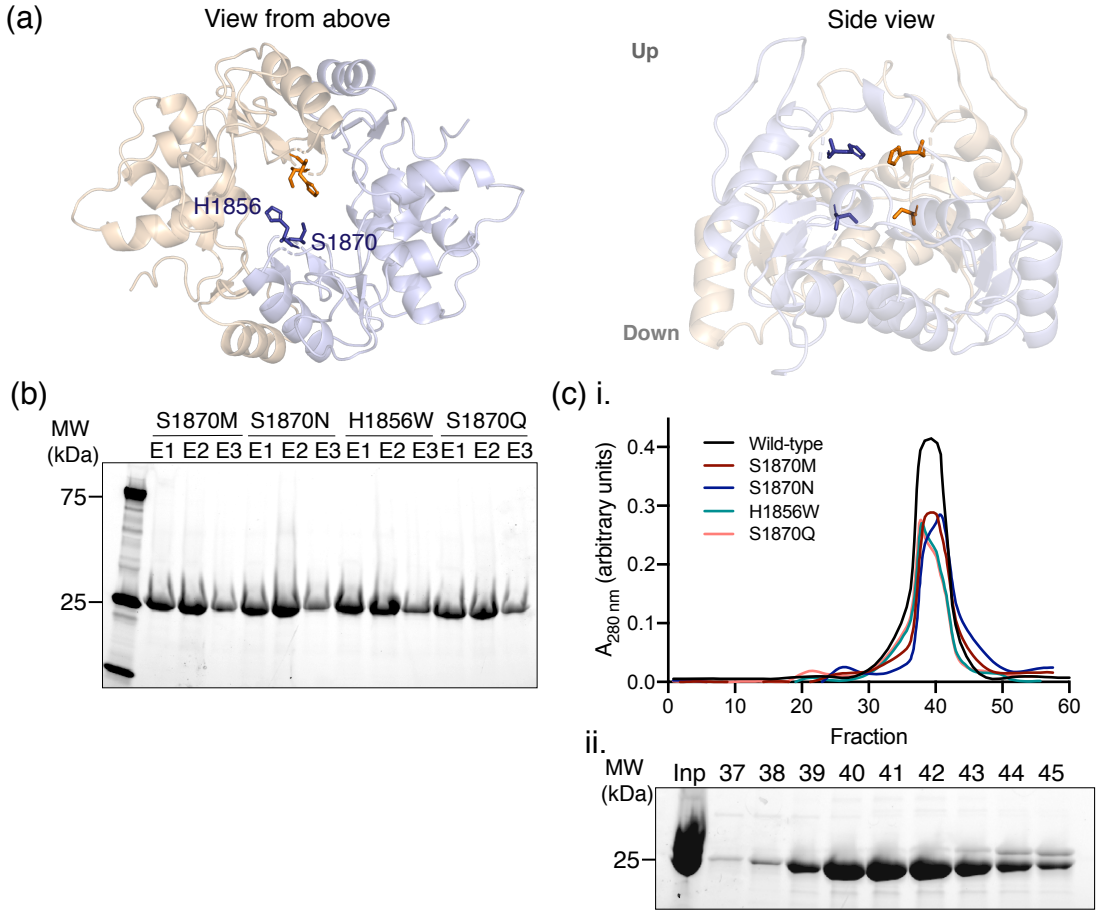


FIGURE 4.5: **Purification of Smchd1 hinge domain torus mutants.** (a) Cartoon representations of the Smchd1 hinge domain (PDB:6N64), where monomers are shown in pale orange and blue. Residues H1856 and S1870 are shown in stick form for each monomer. (b) IMAC purification of each torus mutant, showing eluting samples (E1-E3) analysed by reducing Stain-Free SDS-PAGE. Molecular weight (MW) indicators are shown on the left in kilodaltons (kDa). (c) (i) SEC profiles of wild-type and torus mutants of the Smchd1 hinge domain with (ii) selected fractions analysed by reducing Stain-Free SDS-PAGE.

4.3.3 Mutagenesis of the torus region or the consensus dimerization site of the Smchd1 hinge domain do not alter its overall structure

To confirm the dimerization state and examine any gross structural changes that may be caused by introducing mutations, I performed small angle X-ray scattering (SAXS) analyses on the wild-type Smchd1 hinge domain and all four torus mutants, S1870Q/M/N and H1856W. In parallel, I examined the triple mutant G1872A/G1875A/G1876A that targets dimer interface residues and leads to a reduction in protein thermal stability *in vitro* [162]. Overall, SAXS analyses indicated that none of these mutations disrupted Smchd1 dimer formation (Figure 4.6). The overall size of the protein samples is indicated by the radius of gyration (R_g) value,

which represents the square root of the average distance of each scattering atom from the particle centre, and is calculated by Guinier approximation. R_g values were additionally determined by computing the real-space interatomic distance distribution function ($P_{(r)}$). Both methods yield comparable R_g values for all tested samples, as summarised in Table 4.1.

Comparing torus mutants to wild-type Smchd1 hinge domain, R_g values range between 26.0 and 27.6 Å for mutants, versus 27.0 Å obtained for the wild-type protein (Table 4.1). The triple mutant exhibits a slightly smaller R_g value of 25.1 and 25.8 Å obtained via the two separate calculations. However, maximum dimension (D_{max}) values, which describe the estimated maximum diameter of the protein based on the distance distribution function ($P_{(r)}$), measures at 100 Å consistently across wild-type, torus mutants, as well as the triple glycine mutant sample, indicating that all variants adopt a similar conformation to the wild-type counterpart.

The obtained scattering data also allows the estimation of the intensity of radiation scattered at zero angle $I(0)$, a value that is extrapolated either from the Guinier approximation or from the ($P_{(r)}$) function (Table 4.1). $I(0)$ is equal to the square of the number of electrons in the scattering object and is thus proportional to the molecular mass of a protein at a given concentration. This value, alongside the linearity of the Guinier plot, reflect whether high molecular weight aggregates or inter-particle interference are present, which may contribute measurably to scattering and therefore the quality of the data. $I(0)$ values obtained from the Guinier approximation are consistent with values obtained from the ($P_{(r)}$) function across the protein samples (Table 4.1). Guinier plots depicted in Figure 4.6 (insets in panels i) also highlight a linearity in scattering at low angles for all tested samples, indicating the particles are monodisperse.

Overall, the observation that none of the mutations appear to disrupt Smchd1 hinge domain dimer formation was particularly surprising for the triple glycine mutant. Substitution of this motif in other SMC hinge domains was reported to abrogate dimerization [206], yet SAXS analysis indicates the Smchd1 hinge mutant remains homodimeric in solution. A previous study investigating Smchd1 introduced the same triple mutation to Ala and observed a modest shift in elution time following SEC analysis [131], however they did not further examine this disruption biophysically.

These findings suggest the flanking α-helices present in our Smchd1 hinge domain construct may augment dimer formation.

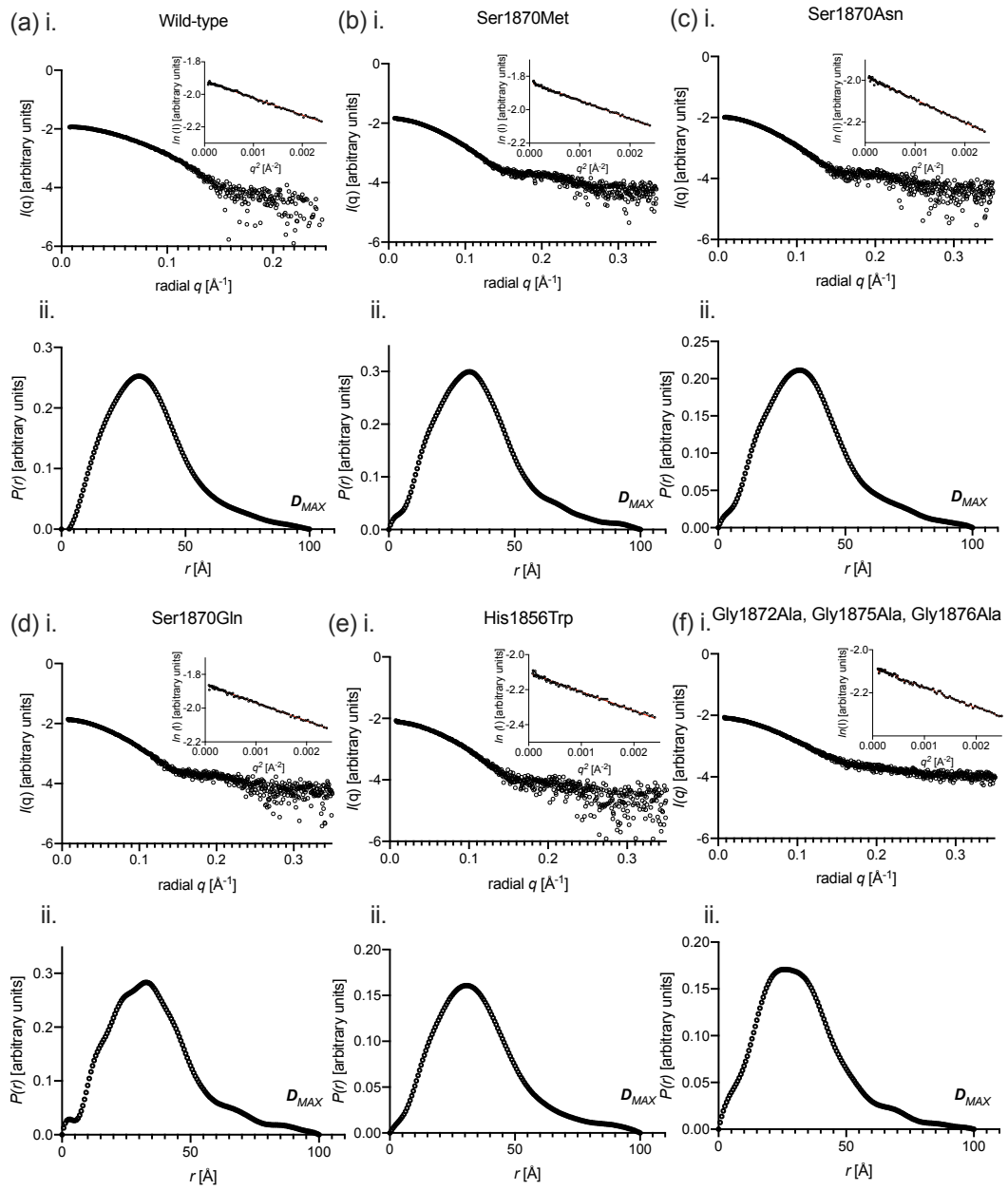


FIGURE 4.6: **Mutagenesis of the torus region or the consensus dimerization site of the Smchd1 hinge domain do not alter its overall structure.** Small angle X-ray analysis of (a) wild-type Smchd1 hinge domain, (b-e) torus mutants and (f) a triple mutant replacing the three consensus Glycine residues required for dimerization in canonical SMC hinge domains. Scattering intensity profiles are shown in (i), where the background-subtract SAXS data are indicated as black circles representing mean intensity I(q) as a function of momentum transfer q in Å^{-1}. Guinier plot for $qR_g \leq 1.3$ are shown as insets in (i), where linearity indicates that particles are monodisperse. The radius of gyration and initial scattering intensity $I(0)$ were approximated using the Guinier equation with PRIMUS, giving values as indicated in Table 4.1. Pair-distribution functions, $P_{(r)}$ plots calculated using GNOM are shown in (ii). The R_g and maximum particle dimension D_{max} calculated from the $P_{(r)}$ analysis are as reported in Table 4.1.

TABLE 4.1: **Data collection and scattering parameters for SAXS analysis.**

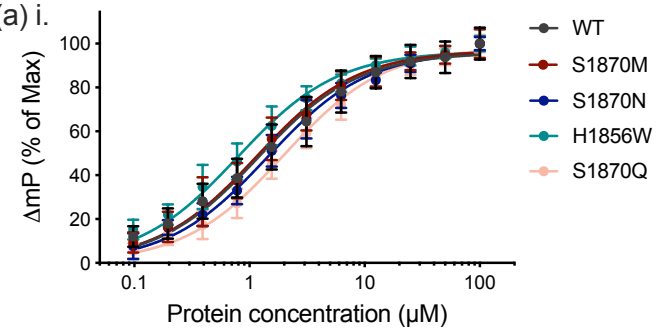
Data collection parameters						
Instrument	Australian Synchrotron SAXS/WAXS beamline					
Beam geometry	120 µm point source					
Beam wavelength (Å)	1.033					
q range (Å^{-1})	0.0114-0.4					
Exposure time (seconds)	2					
Protein concentration	~5 mg/ml sample injected via in-line size exclusion chromatography					
Temperature (°C)	16					
Structural parameters						
Protein sample	WT	S1870M	S1870N	S1870Q	H1856W	G1872A/ G1875A/G1876A
I(0)(cm^{-1}) [from Guinier]	0.06757 ±0.00022	0.01447 ±0.00085	0.01041 ±0.00086	0.01377 ±0.00094	0.07781 ±0.00079	0.08236 ±0.00069
Rg (Å) [from Guinier]	27.0 ±0.141	27.1 ±0.242	27.6 ±0.336	27.2 ±0.283	26.9 ±0.415	25.1 ±0.332
I(0)(cm^{-1}) [from *P(r)*]	0.06799 ±0.00019	0.01455 ±0.00079	0.01043 ±0.00077	0.01384 ±0.00010	0.07850 ±0.00075	0.08284 ±0.00065
Rg (Å) [from *P(r)*]	27.70 ±0.127	27.8 ±0.274	28.0 ±0.356	28.0 ±0.380	27.9 ±0.423	25.8 ±0.457
D_{max} (Å)	100	100	100	100	100	100
Software						
Primary data reduction	ScatterBrain (Australia Synchrotron)					
Data processing	PRIMUS, GNOM					

4.3.4 The Smchd1 hinge domain does not thread nucleic acids through its central pore

To examine the DNA-binding affinity of the torus mutants compared to wild-type Smchd1 hinge domain, I performed an *in vitro* fluorescence polarization assay using a 20 base-pair single-stranded DNA probe containing one methylated cytosine that was previously tested and shown to interact with Smchd1 [58, 162]. This oligonucleotide represents a *Pcdh* cluster-specific sequence, termed the HS5-1b enhancer site, where Smchd1 was shown to localise by chromatin immunoprecipitation and sequencing (ChIP-seq) studies [58]. These studies also showed that while Smchd1 exhibits a slight preference for methylated DNA *in vitro*, differences in binding affinities compared to both double-stranded DNA (dsDNA) and unmethylated DNA probes were within a 2-fold window [58]. I chose to only examine the methylated HS5-1b oligonucleotide probe across all DNA-binding assays in order to enable direct comparisons between test samples. However, it is anticipated that DNA-binding differences between mutants will apply to all oligonucleotide probes.

As seen in Figure 4.7, the presence of bulkier residues at the torus region of the hinge domain does not impact DNA-binding, as the mutants largely exhibit K_d values similar to the wild-type counterpart. There seems to be a small decrease in the DNA-binding affinity of S1870Q for the HS5-1b oligonucleotide, displaying a K_d value of approximately 2.0 μM compared to wild-type protein which shows a value of $\sim$1.2 μM. The mutant H1856W actually reveals a small increase in DNA-binding affinity, with a K_d value of $\sim$0.8 μM, however this difference is not considered substantial and may be accounted for by technical error. Taken together, these results are consistent with the hypothesis that nucleic acids are not threaded through the central pore of the Smchd1 hinge domain.

(a) i.



ii.

Protein	Kd value (μM)
Wild-type	1.2 ± 0.1
S1870M	1.2 ± 0.1
S1870N	1.4 ± 0.1
H1856W	0.8 ± 0.1
S1870Q	2.0 ± 0.1

FIGURE 4.7: **Torus mutants do not alter nucleic acid binding of the Smchd1 hinge domain, suggesting binding does not occur by threading through the central channel.** (a) (i) DNA-binding fluorescence polarization (mP) of 12.5 nM 6-FAM labeled 20 bp single-stranded DNA by wild-type (WT) and torus mutants of the Smchd1 hinge dimer at indicated protein concentration (μM). Data points are plotted for technical duplicates of three independent experiments fitted to a hyperbolic binding curve, where error bars represent the standard deviation. (ii) K_d and standard deviation values calculated from the fitted curves are outlined for each protein sample, in μM.

4.3.5 Interface mutations in conjunction with shorter coiled-coil regions in the Smchd1 hinge domain disrupt the dimer

While our crystallised Smchd1 hinge domain construct comprised of residues 1683-1899, electron density was observed only for residues 1710-1886. The predicted N-terminal α-helices which are thought to form coiled-coils are therefore absent from the three-dimensional structure, and any interactions that may be mediated by these regions are yet to be determined. This also indicates the regions lacking electron density are likely to be conformationally flexible. Previous SAXS analyses of the 1683-1899 amino acid construct revealed a 'drumstick' model where the hinge domain is flanked by an N-terminal protruding region, thought to correspond to the N-terminal α-helices [62]. These data provide reasoning for the flanking coiled-coil domains forming head-to-head dimers, which contrasts with previously reported SMC hinge domains that were shown to form head-to-tail (N- to C-terminal) coiled-coils [204, 207]. Additionally, as the introduction of mutations within the Smchd1 hinge domain interface do not appear to interfere with dimerisation, it was therefore speculated these coiled-coil regions may play a role in hinge domain homodimerisation.

To answer this question, I expressed a shorter hinge domain construct encompassing residues 1699-1883 of Smchd1 which lacks the flanking α-helices (Figure 4.8 c). I additionally introduced mutations V1774G, triple mutant G1872A/G1875A/G1876A

or R1848A in this construct, all of which are located at the dimer interface (Figure 4.8 d) and exhibit a reduced thermal stability *in vitro* [162], therefore they are expected to further interfere with hinge domain dimerisation. While the wild-type protein was successfully purified for this construct, all three mutants were poorly expressed (Figure 4.8) and majority of protein eluted in the void volume upon SEC analysis (data not shown). However, the same mutants were successfully expressed in the context of the longer hinge domain construct that comprises of the flanking α-helices and appeared homodimeric [162], while deletion of these predicted coiled-coil regions compromised their stability. Therefore, my data support the existence of the hinge domain dimer interface observed in the crystal structure. These results are consistent with a role for the flanking α-helices in augmenting the dimeric structure of the Smchd1 hinge domain, undeterred by mutation-induced disruption of the intersubunit hinge domain interface.

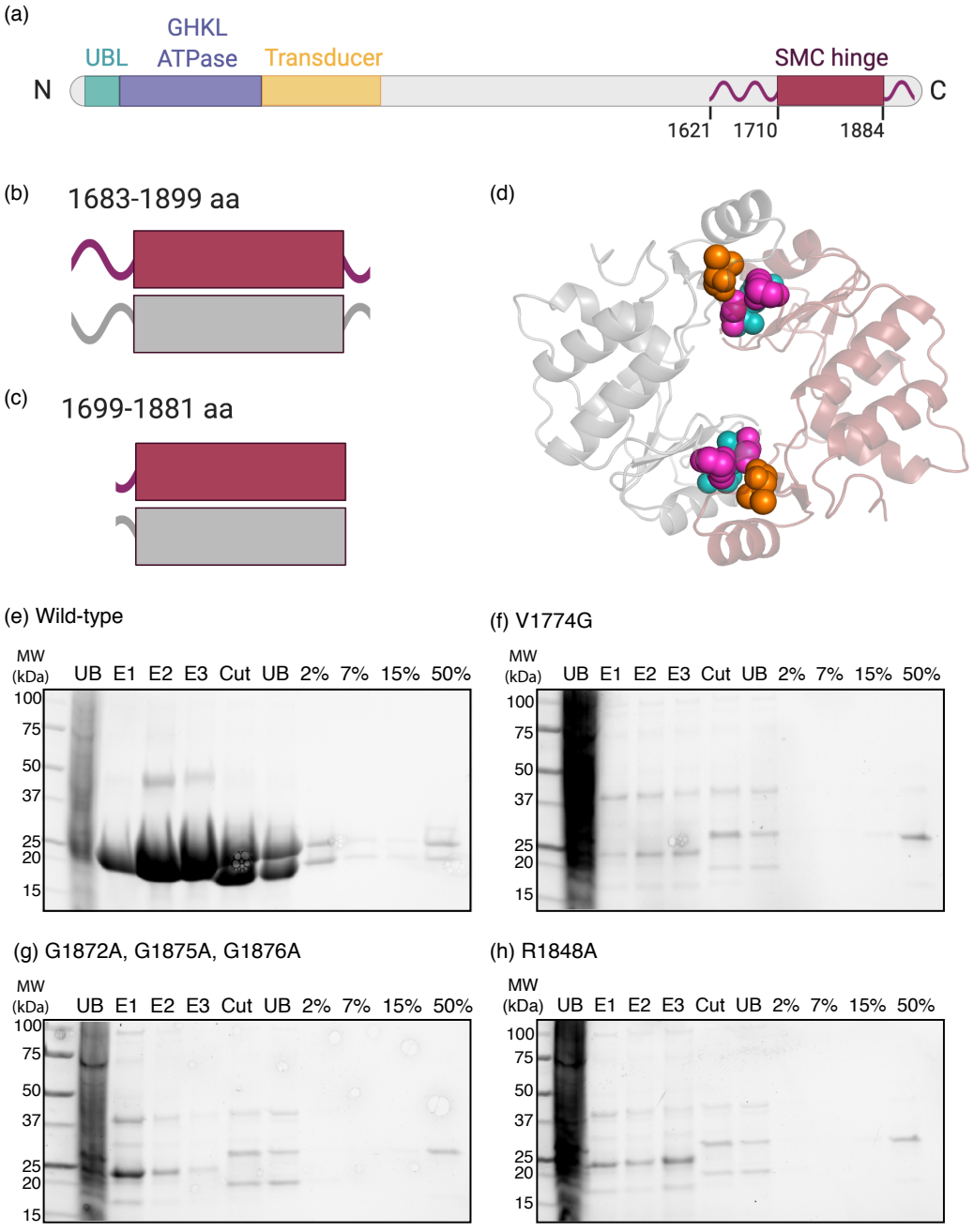


FIGURE 4.8: **Dimer interface mutations in conjunction with shorter coiled-coil regions in the Smchd1 hinge domain disrupt the dimer interface.** (a) Schematic representation of the full length Smchd1 gene architecture, indicating domain boundaries of the SMC hinge domain as amino acid residue numbers. Comparing constructs of (b) the crystallised Smchd1 hinge domain and (c) the Smchd1 hinge domain with a shorter N-terminal coiled-coil region. (d) Cartoon representation of the Smchd1 hinge domain dimer (monomers shown in red and grey; PDB:6N64) highlighting residues at the dimer interface: V1774 in orange, G1872/1875/1876 in magenta and R1848 in cyan, as spheres. (e-h) IMAC followed by subtractive IMAC purification of (e) wild type and (f-h) dimer interface mutants (1699-1881 aa). (e-h) Samples were analysed by reducing Stain-Free SDS-PAGE, where UB=unbound, E1-E3= elution samples, cut= sample following TEV-cleavage, 2%-50%= imidazole concentrations. Molecular weight (MW) indicators are indicated on the left in kilodaltons (kDa).

4.3.6 Mutation of R1848 and R1867 alter Smchd1's nuclear localization pattern

To examine how residues that contribute to Smchd1's nucleic acid binding *in vitro* may affect its cellular function, cluster 2 mutations R1790A and R1796A, interface mutations D1842A and R1848A, and cluster 3 mutations R1867G, R1869A and K1873A, were introduced into full-length mouse Smchd1 by Dr. Tracy Willson (Walter and Eliza Hall Institute). I transfected these constructs into human 293 cells in which endogenous SMCHD1 expression was silenced by a short hairpin RNA (shRNA). Crucially, shRNA-silencing is restricted to human SMCHD1 in this system, therefore enabling the examination of the introduced full-length mouse Smchd1 constructs containing point mutations of interest. Successful knockdown of SMCHD1 is highlighted by Western blot analysis in Figure 4.9, which was performed by Ruoyun Wang (Walter and Eliza Hall Institute), in addition to depletion of SMCHD1 nuclear foci in SMCHD1-knockdown (KD) 293 cells (Figure 4.10). Bright nuclear foci are known to correspond to SMCHD1's localisation to the inactive X chromosome in wild-type cells, as previously shown [51]. Importantly, all transfections were performed in parallel, with an equivalent amount of construct transfected for each of the tested samples, and imaging parameters identical between samples to allow for an accurate comparison.

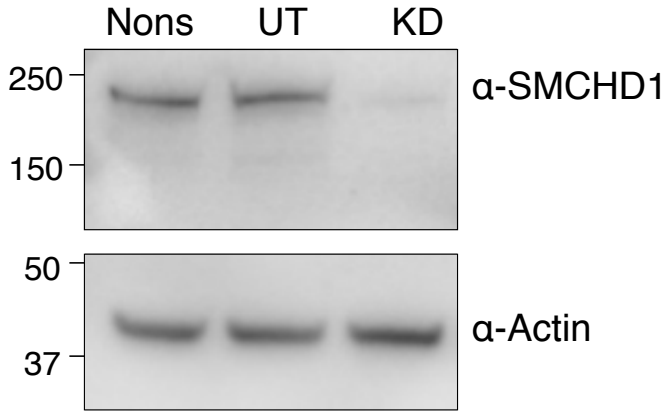


FIGURE 4.9: **SMCHD1-knockdown in HEK293 cells.** Western blot analysis of HEK293 cells transduced with either a non-silencing (nons) shRNA, untransduced HEK293 cells (UT) or cells transduced with a SMCHD1-knockdown shRNA (KD). Actin was used as a loading control. Molecular weight markers are indicated on the left, in kilodaltons.

Transfection of the wild-type mouse Smchd1 construct into KD cells restore Smchd1's enrichment to nuclear foci. However, these appeared distinct from endogenous human SMCHD1 staining as multiple foci are observed (Figure 4.10 c), rather than only the expected two foci corresponding to the two inactive X-chromosomes present in tetraploid 293 cells (Figure 4.10 a). This is likely due to the over-expression of the construct, and hence all mutants of interest are compared to the over-expression of the wild-type Smchd1 construct rather than the non-silencing shRNA control. Across the various mutants examined, a similar cellular localisation pattern to wild-type Smchd1 over-expression was observed for cluster 2 mutants, R1790A and R1796A, and cluster 3 mutants R1869A and K1873A. Likewise, mutant D1872A, which was chosen as a control as it does not alter protein thermal stability or nucleic acid binding *in vitro*, exerted a similar localisation in cells to wild-type Smchd1. Only two tested mutations, R1867G (cluster 3) and R1848A (dimer interface), appeared to abrogate nuclear foci formation (Figure 4.10 g-h). The R1867G mutant exhibited a visible decrease in the number of nuclear foci present compared to the introduction of the wild-type counterpart, where multiple nuclear foci were seen. A similar phenotype was observed for the R1848A mutant, with a very diffuse presence of nuclear Smchd1 that exhibited a complete absence of nuclear foci.

R1867G is an FSHD2-associated substitution, located within cluster 3 of Smchd1's hinge domain (Figure 4.3). Previous *in vitro* studies also investigated this mutation in the isolated hinge domain of Smchd1 and showed this residue is important in nucleic acid interaction, as its substitution to glycine reduced its affinity for DNA [58]. We also found similar results via our *in vitro* fluorescence polarisation assay, where the mutant exhibits a 6.2-fold reduction in DNA-binding affinity relative to its wild-type counterpart (Figure 4.3 d). R1848 is positioned at the central pore region of the hinge domain, indicating a potential role in mediating interprotomer contacts. Additionally, the Ala mutant shows a 11.4-fold reduction in DNA-binding affinity *in vitro* (Figure 4.2 b), suggesting R1848 may also mediate nucleic acid interactions. Taken together, these data indicate that mutations that impact Smchd1's nucleic acid binding *in vitro* do not typically compromise its chromatin binding ability in cells, likely owing to due compensation from adjacent residues. However, it appears that residues R1848 and R1867 contribute to a functional hotspot in the hinge domain that plays a key role in localising Smchd1 to chromatin and their neutralisation cannot be compensated for.

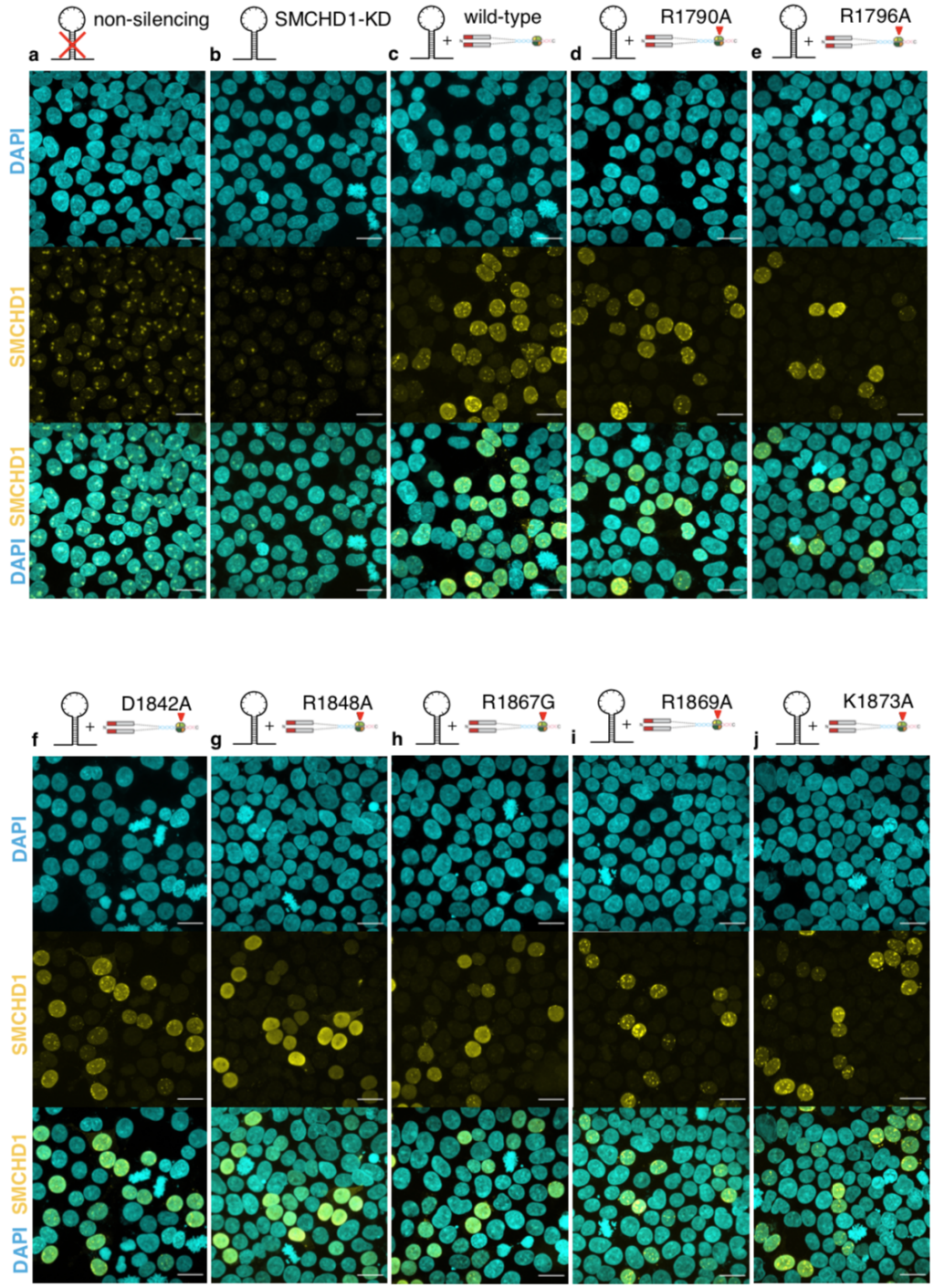


FIGURE 4.10: **Mutation of R1848 and R1867 alter Smchd1's nuclear localization pattern.** Immunofluorescence in (a) control, non-silencing shRNA transduced, or shRNA-mediated SMCHD1-knockdown 293 cells 24 hours post-transfection with (b-j) the indicated full-length Smchd1 variants. DAPI staining is depicted in blue, whereas Smchd1 staining is shown in yellow, and merged channels are displayed below. Maximum intensity projection images are shown as representative of n>150 nuclei positive for Smchd1 overexpression per sample; data shown are representative of 3 independent experiments. All images were obtained with identical settings between controls and all transfected cells, to enable comparison between the images provided in the figure. Scale bars, 20 μm.

4.3.7 The hinge domain targets Smchd1 to chromatin

To further investigate whether the hinge domain contributes to Smchd1's chromatin localisation, additional mutations were introduced in the full-length construct by Dr. Tracy Willson (Walter and Eliza Hall Institute): V1774G and the triple mutant G1872A/ G1875A/G1876A. We previously demonstrated these mutations reduce the thermal stability of hinge domain dimers within the 1683-1899 aa construct that retains an N-terminal coiled-coil region, therefore compromising the integrity of Smchd1's hinge domain [162]. Structurally, these residues reside at the dimer interface (Figure 4.8 d), while the G1872A/G1875A/G1876A mutant additionally targets a triglycine motif located at Smchd1's C-terminus which is highly conserved and necessary for dimerization across SMC hinge domains [206] (Figure 4.4). I also examined the cellular localisation of a full-length Smchd1 construct that lacks the hinge domain altogether (Δ L1710-M1884). Importantly, the nuclear localisation signal (NLS) of SMCHD1 was previously reported to be located within residues 1961-1992, which was shown as sufficient for its localisation to the nucleus [208]. Deletion of the hinge domain in my construct therefore did not interfere with the integrity of the NLS. Upon transfection of these constructs into 293 cells and follow-up immunofluorescence studies, it can be seen that either the introduction of mutations V1774G, G1872A/G1875A/G1876A, or deletion of the hinge domain from full-length Smchd1, lead to abrogation of focus formation (Figure 4.11 d-f). These results support a role for the hinge domain in Smchd1's chromatin localisation.

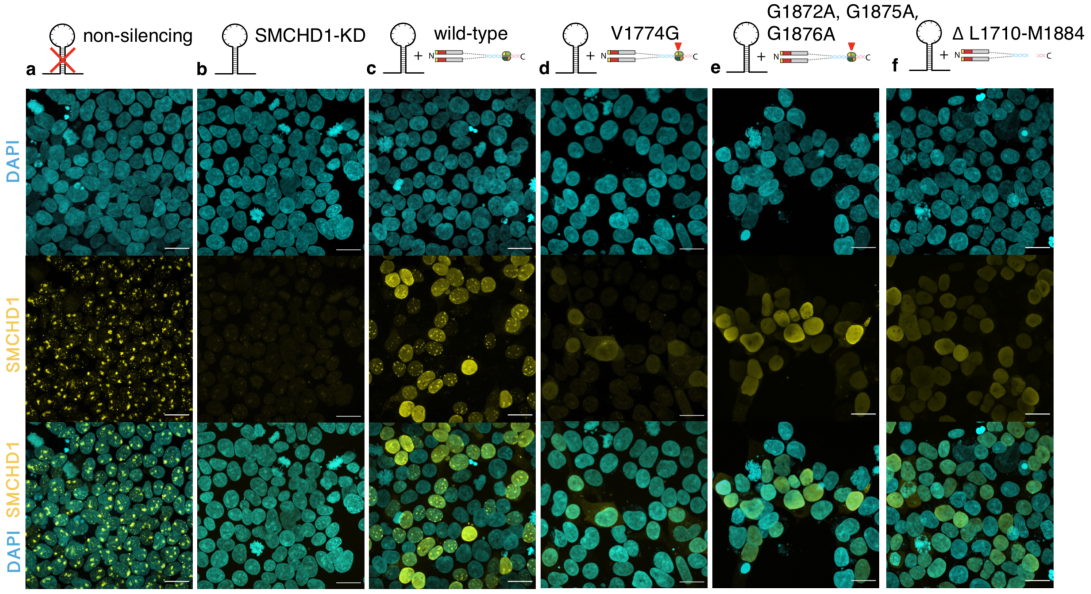
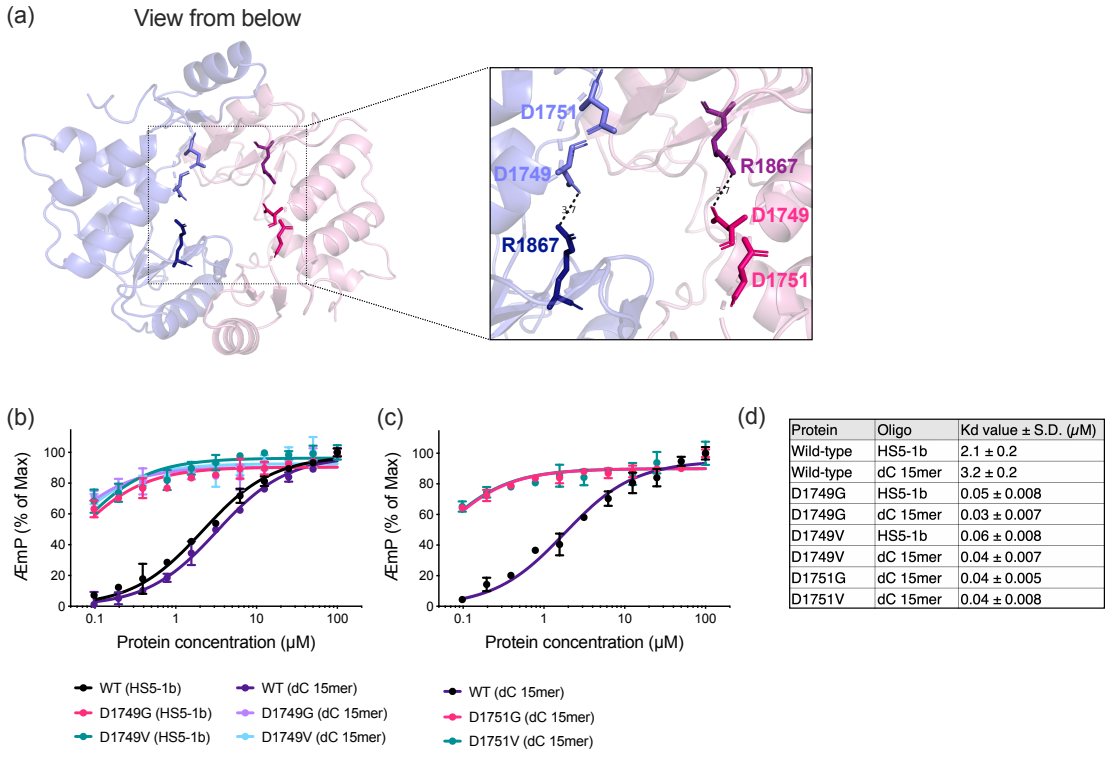


FIGURE 4.11: **Deletion or compromise of hinge domain integrity alter Smchd1's nuclear localisation pattern.** Immunofluorescence in (a) control, non-silencing shRNA transduced, or (b-f) shRNA-mediated SMCHD1-knockdown 293 cells 24 hours post-transfection with the indicated full-length Smchd1 with mutations (d,e) in the dimer interface or (f) deleted hinge domain. DAPI staining is depicted in blue, whereas Smchd1 staining is shown in yellow, and merged channels are displayed below. Maximum intensity projection images are shown as representative of n>100 nuclei positive for Smchd1 overexpression per sample; data shown are representative of 2 independent experiments. All images were obtained with identical settings between controls and all transfected cells, to enable comparison between the images provided in the figure. Scale bars, 20 μm.

4.3.8 Smchd1 hinge domain residues D1749 and D1751 exhibit an enhanced affinity for DNA *in vitro* and an altered cellular localisation when substituted for Gly or Val

Given the FSHD2-associated mutation, R1867G, affects both Smchd1's DNA-binding *in vitro* in addition to chromatin localisation in cells, I sought to investigate whether residue D1749, which is spatially situated adjacent to R1867, contributes to DNA-binding (Figure 4.12 a). When substituted for small neutral residues, D1749G and D1749V appeared to exhibit an enhanced affinity for DNA in the *in vitro* DNA-binding assay (Figure 4.12 b,d). During the course of my study, two new FSHD2-related mutations in *SMCHD1* were added to the LOVD open-source database [209]: D1750G and D1750V, corresponding to substitutions of mouse Smchd1 D1751. While D1748 of human SMCHD1 (D1749 in mouse Smchd1) does not show sequence conservation across SMC hinge proteins, the FSHD2-affected residue in SMCHD1, D1750 (D1751 in mouse Smchd1), is conserved across mouse SMC1A and human

SMC4 proteins (Figure 4.4). Variants D1751G and D1751V also exhibit an enhanced affinity for DNA *in vitro* (Figure 4.12 c,d). This is an unexpected finding as FSHD2-associated mutations in *SMCHD1* are associated with loss of SMCHD1 function.



Protein	Oligo	Kd value ± S.D. (μM)
Wild-type	HS5-1b	2.1 ± 0.2
Wild-type	dC 15mer	3.2 ± 0.2
D1749G	HS5-1b	0.05 ± 0.008
D1749G	dC 15mer	0.03 ± 0.007
D1749V	HS5-1b	0.06 ± 0.008
D1749V	dC 15mer	0.04 ± 0.007
D1751G	dC 15mer	0.04 ± 0.005
D1751V	dC 15mer	0.04 ± 0.008

FIGURE 4.12: **Two Smchd1 hinge domain residues, D1749 and D1751, exhibit an enhanced affinity for DNA when substituted for Gly or Val.**(a) Cartoon representation of the Smchd1 hinge domain dimer (PDB: 6N64), where monomers are depicted in light purple and pink. A zoomed-in image is shown on the right, depicting residues D1751, D1749 and R1867 in each monomer. (b-c) DNA-binding fluorescence polarization (mP) of 12.5 nM 6-FAM labeled 20 bp single-stranded DNA by wild-type (WT), (b) D1749 mutants and (c) D1751 mutants of the Smchd1 hinge dimer at the indicated protein concentration (μM). Data points are plotted for technical duplicates of two independent experiments fitted to a hyperbolic binding curve, where error bars represent the standard deviation. (d) K_d and standard deviation values calculated from the fitted curves are outlined for each protein sample and oligonucleotide tested in μM, rounded to one significant figure.

These results prompted me to further investigate the effect of mutations D1749G and D1749V in the context of full-length Smchd1 by examining their localisation in cells. Similar to the previous experiments, I transfected the full-length constructs into SMCHD1-knockdown 293 cells and immunofluorescence was performed 24 hours later. As can be seen in Figure 4.13 (d-e), these mutations alter Smchd1's localisation to chromatin as a complete absence of nuclear foci is observed. *In vitro* experiments performed in the context of the hinge domain suggested these mutations enhance

Smchd1's affinity for DNA, therefore absence of nuclear foci formation may suggest Smchd1 mutants D1749G and D1749V undergo non-specific DNA interactions that may result in the diffuse staining pattern observed.

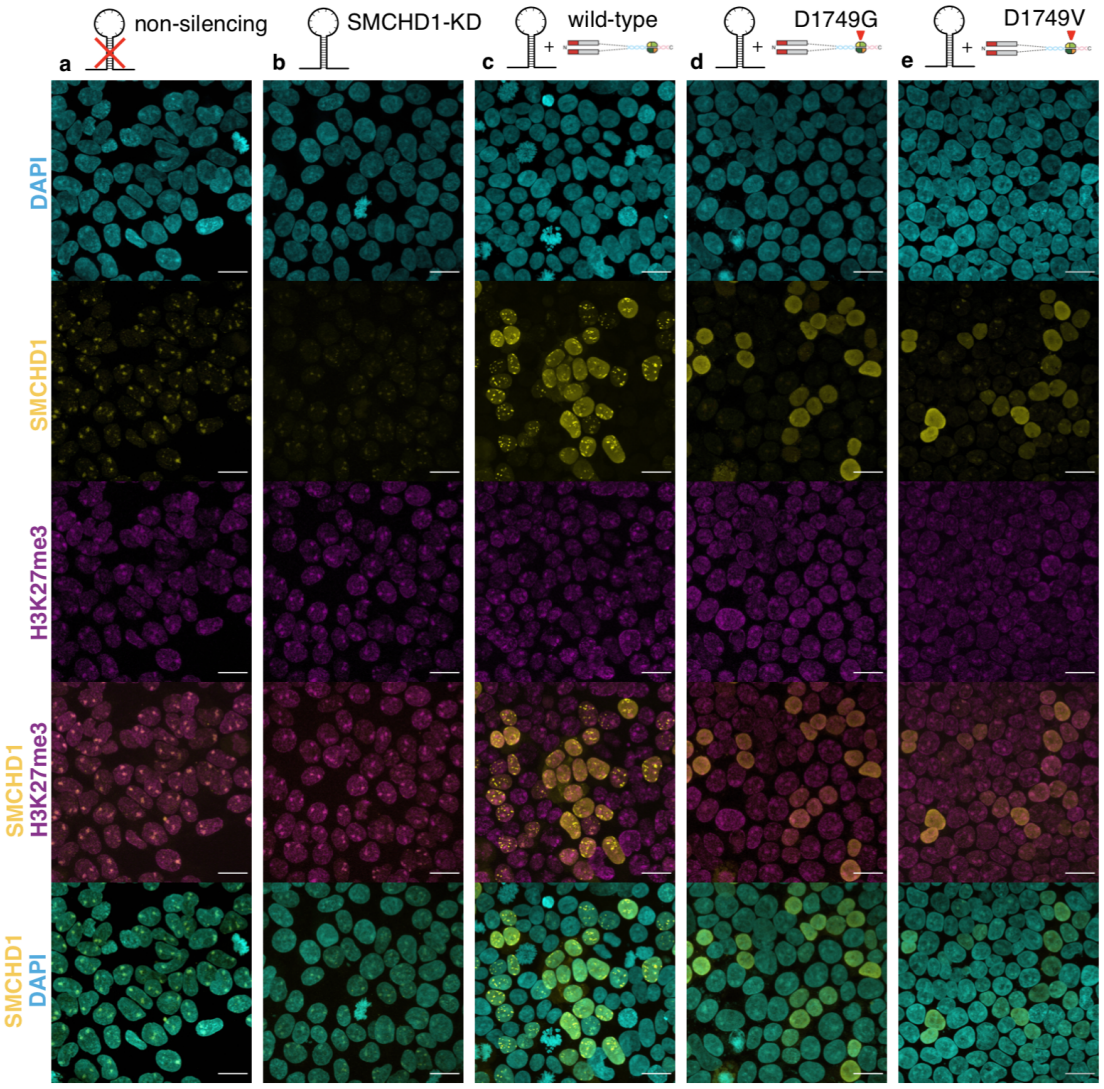


FIGURE 4.13: **Smchd1 hinge domain mutants D1749G and D1749V alter its nuclear localisation pattern.** Immunofluorescence in (a) control, non-silencing shRNA transduced, (b) shRNA-mediated SMCHD1-knockdown 293 cells and (c) SMCHD1-knockdown 293 cells 24 hours post-transfection with the full-length Smchd1 containing point mutations (d) D1749G or (e) D1749V. DAPI staining is depicted in blue, Smchd1 staining is shown in yellow and H3K27me3 is shown in purple, with merged images shown below as indicated. Maximum intensity projection images are shown as representative of n>100 nuclei positive for Smchd1 overexpression per sample; data shown are representative of 2 independent experiments. All images were obtained with identical settings between controls and all transfected cells, to enable comparison between the images provided in the figure. Scale bars, 20 μm.

4.4 Discussion

The availability of the first crystal structure of the Smchd1 hinge domain enabled structure-guided mutagenesis studies to better investigate its mechanistic properties. Of three identified positively charged patches on the hinge domain surface, two were revealed as key sites of nucleic acid interaction. One of these, termed cluster 3, encompasses residues surrounding the central channel therefore highlighting the possibility that DNA may be threaded through the central pore. By introducing bulkier residues as a means of occluding nucleic acids from passing through the torus region, I have provided evidence that disproves this hypothesis as mutations in the hinge torus did not affect DNA-binding *in vitro*. However, residues R1867 and R1848, which were revealed as key sites of nucleic acid interaction both *in vitro* as well as in cellular localisation studies, are located in close proximity to the central pore region. Their position is too distal from the hinge domain surface to interact directly with nucleic acids, suggesting that the hinge domain may undergo a conformational transition from the closed state observed in the crystal structure, to an open hinge conformation that may expose this region and allow direct interactions with DNA and chromatin. A recent study investigating the canonical SMC complex, cohesin, has proposed a similar mechanism for its hinge domain, where a crystal structure of the heterodimeric SMC hinge was obtained in an open conformation while bound to single-stranded DNA [210].

Experimental evidence suggests that the α-helices flanking the Smchd1 hinge region likely play a role in dimer formation. Based on sequence prediction and structural similarities to other hinge domains, the flanking α-helices of the Smchd1 hinge domain are predicted to form coiled-coils. Unfortunately, atomic detail of this region was absent from the electron density of our solved hinge domain crystal structure. Surprisingly, we found that the introduction of interface mutations did not affect hinge dimer formation, but resulted in a reduced protein thermostability [162]. To further test the importance of the flanking α-helices in dimer formation, I introduced dimer-disrupting mutations in a Smchd1 hinge domain expression construct that lacks the surrounding α-helices and I produced the recombinant protein in *E. coli* cells. While the wild-type construct was successfully purified, all dimer-disrupted mutants were poorly expressed and eluted in the void volume upon SEC analysis. Because the same mutants were previously

successfully expressed in the context of a construct that comprises of the flanking α-helices [162], these data suggest a role for the adjacent α-helical regions in augmenting Smchd1 dimer formation. Therefore, interface mutations introduced in the extended construct are being suppressed by dimerisation of the predicted coiled-coil regions.

Previous SAXS studies on the Smchd1 hinge domain suggested a head-to-head arrangement of the dimer based on the observed polarity of the flanking α-helices [62]. Low-resolution electron microscopy images of full-length Smchd1 support this finding, with globular domains observable at either apex, which likely represent the ATPase and hinge domains that segregate at the N- and C-terminus, respectively [131]. However, our crystal structure of the Smchd1 hinge domain reveals the N- and C-termini of the two monomers are on the same face. It may therefore be possible that in response to a stimulus such as DNA interaction, the α-helices surrounding the hinge domain may re-arrange to adopt either a head-to-tail or head-to-head arrangement. It additionally remains unclear whether these helices also play a role in chromatin interactions. Removal of the coiled-coil regions in the presence of an intact hinge domain within a full-length Smchd1 construct and follow-up immunofluorescence studies may help answer this question.

Here, I have shown via cellular studies that the hinge domain is required for Smchd1's localisation to chromatin, and compromising its integrity such as via dimer interface mutations results in an altered nuclear localisation pattern. These findings therefore suggest that Smchd1 dimer stability may also play an essential role in nucleic acid binding. While residues from clusters 2 and 3 were implicated in DNA-binding *in vitro*, their introduction into full-length Smchd1 and follow-up cellular studies revealed only two residues were critical in chromatin localisation in cells: R1867 from cluster 3, and R1848 within the dimer interface. These results indicate most nucleic acid-binding residues are dispensable for chromatin localisation in cells, whereas R1848 and R1867 likely contribute to a functional hotspot in the hinge domain that cannot be compensated for by other adjacent DNA-binding residues. However, the possibility of these two mutations interfering with hinge domain stability cannot be excluded, particularly as I have shown that hinge domain integrity is required for Smchd1's chromatin localisation.

I have additionally identified mutations that enhance Smchd1's DNA-binding affinity *in vitro*, affecting residues D1749 and D1751. Interestingly, residue D1749 is situated only

3.7 Å apart from the FSHD2-affected residue, R1867, based on the crystal structure of the Smchd1 hinge domain. Given the importance of R1867 in nucleic acid binding, these results may suggest a potential interaction between D1749 and R1867, where the negatively charged D1749 may normally act as a brake to nucleic acid binding by interacting with the positively charged R1867. Therefore, upon substitution of D1749 to either Gly or Val, and the removal of the negative charge, it would allow R1867 to freely interact with the negatively charged DNA, providing a potential explanation for the enhanced DNA-binding affinity observed for this mutant. This hypothesis may not apply for residue D1751 as it is located 10.4 Å away from R1867 based on the three-dimensional structure.

An enhanced DNA-binding capability of residues D1749 and D1751 is further translated into an altered chromatin localisation upon their introduction into human SMCHD1-knockdown 293 cells via full-length mouse Smchd1, where the mutants fail to localise to nuclear foci. A possible explanation for this may be that enhancing Smchd1's DNA-binding affinity can result in non-specific interactions that may visually present as a diffuse staining pattern within cellular nuclei. Whether an enhanced interaction with DNA may represent an enhanced epigenetic silencing capability of Smchd1 is unknown. Considering Smchd1 variants D1751G and D1751V are FSHD2-related, where it has been established that the condition arises due a loss of SMCHD1 function, it indicates this might not be the case. Instead, an aberrant or prolonged occupancy of Smchd1 on chromatin may prevent it from performing its original function as it likely requires a specific balance that allows it to engage and disengage with target chromatin sites accordingly.

Relating these findings to the functional properties of another epigenetic regulator, the polycomb repressive complex-2 (PRC2), it has been proposed that its non-sequence specific affinity for nucleic acids acts to retain the complex on chromatin long enough for it to enzymatically modify histone proteins [211]. We have previously shown that Smchd1 is likewise able to associate with RNA without sequence specificity [55]. Additionally, we have found that Smchd1's targeting to chromatin is dependent on the PRC1-mediated ubiquitylation of histone 2A (H2AK119ub) [55]. Therefore, similar to the targeting of PRC2 to chromatin, the nucleic acid-binding capacity of Smchd1's hinge domain may be necessary to stabilise its chromatin-bound state to enable efficient silencing. An enhanced nucleic acid-binding capacity of Smchd1, as observed

for variants of residues D1749 and D1751, would thereby prolong its residence time on chromatin. Although the specific consequences of this behaviour remain unknown, these findings highlight the possibility that either a gain or loss in DNA-binding affinity may lead to a loss of overall function of Smchd1, and the associated FSHD condition in humans.

Chapter 5

Structure and function of SMCHD1's ATPase domain

5.1 Abstract

SMCHD1's ATPase domain has been of recent interest due to the identification of disease-related variants that are frequently located within this region of the protein. However, the mechanisms by which some of these pathogenic variants affect SMCHD1 function are poorly understood. To provide more in-depth information about SMCHD1's ATPase dimerisation properties, I used analytical ultracentrifugation studies to demonstrate that the wild-type SMCHD1 ATPase region undergoes dimerisation in the presence of both the UBL domain and ATPase substrate. *In vitro* ATPase assays revealed that the presence of either the N-terminal UBL domain, or a C-terminal extension of the SMCHD1 ATPase region, do not alter the protein's catalytic activity. Using cellular studies, I revealed that *Smchd1* mutations that alter the catalytic activity of the SMCHD1 ATPase region protein *in vitro* also exhibit an altered cellular localisation upon their introduction in cells, in the context of full-length Smchd1. Via the same cellular assay, I additionally demonstrated that the UBL domain is necessary for the localisation of full-length Smchd1 to chromatin. Together, these studies suggest that either a perturbed dimerisation or catalytic activity of the ATPase region are required for Smchd1's interaction with chromatin.

5.2 Introduction

Until recently, little was known about the N-terminally located GHKL ATPase domain of SMCHD1. Previous *in vitro* assays demonstrated that the ATPase region was catalytically active and was present as a monomer in solution, with further SAXS analyses revealing its low-resolution structure which was suggested to resemble the full-length molecular chaperone protein, Hsp90 [61]. However, the high-resolution structure and biological function of this domain in relation to the overall molecular function of SMCHD1 remained unknown. To provide a better understanding of the structure-function properties of SMCHD1's ATPase region that may underpin the protein's mechanistic level, I sought to determine its three-dimensional structure via protein crystallisation and electron microscopy methods.

5.2.1 FSHD- and BAMS-associated *SMCHD1* mutations do not map to specific regions of the ATPase region

Variants in the human *SMCHD1* gene have been associated with two debilitating conditions: FSHD and BAMS. FSHD-related variants in *SMCHD1* span the full-length of the gene, while BAMS-associated variants are tightly clustered within the extended ATPase region of SMCHD1 (Appendix B). The confined localisation of BAMS-associated *SMCHD1* variants initially suggested that an altered ATPase activity may underlie the disease pathogenesis. We recently showed that BAMS-related variants resulted in altered catalytic rates of the SMCHD1 ATPase region, where several of the variants we investigated exhibited a gain in ATPase activity [146].

The recent publication of a human SMCHD1 ATPase crystal structure provided the ability to spatially map pathogenic variants [133]. As depicted in Figure 5.1, disease variants do not appear to localise to specific regions of the SMCHD1 ATPase. However, there is a BAMS-associated cluster within a loop region that is situated at the dimer interface, suggesting these may potentially interfere with dimerisation (Figure 5.1 a). This hypothesis was briefly explored via native PAGE and cross-linking experiments, which suggested that dimerisation was instead reduced in FSHD-associated variants, but unaffected in BAMS-associated variants [133]. Nonetheless, these findings require further investigation and should extend to more rigorous, biophysical methods that

are able to reveal more specific information about SMCHD1's ATPase self-association properties.

How disease-associated mutations in *SMCHD1* may influence other functions of the protein, such as specific chromatin or protein interactions, remains to be explored. We previously observed a diminished ATPase activity across FSHD-associated SMCHD1 variants, however mutations found in these patients span the full-length of the *SMCHD1* gene and are not limited to the ATPase region, in contrast to BAMS-associated variants. Hence, while defects in ATPase activity or dimerisation properties of the SMCHD1 ATPase region may account for a loss of function in SMCHD1 in some FSHD cases, there are likely other contributing factors. For example, we previously demonstrated that an FSHD-related variant located within SMCHD1's C-terminal hinge domain, R1867G, reduced its affinity for DNA [58]. In Chapter 4, I furthermore revealed that two separate FSHD-related mutations that also localise to the SMCHD1 hinge domain resulted in an enhanced DNA-binding affinity *in vitro*. It is therefore evident that many aspects of SMCHD1 function are associated with human disease if compromised.

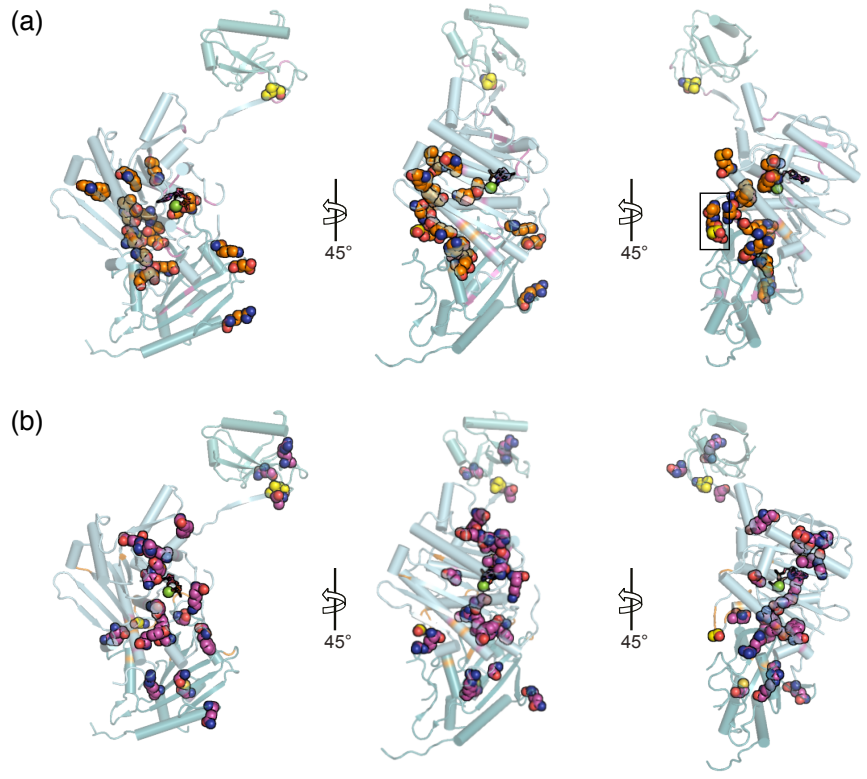


FIGURE 5.1: **BAMS- and FSHD-associated missense variants within SMCHD1's ATPase region.** Cartoon representation of the human SMCHD1 ATPase monomer (25-580 aa) (PDB:6MW7), highlighting (a) BAMS-associated variants in orange, with a box surrounding the location of hotspot residues, and (b) FSHD-associated variants in magenta. (a-b) Residues shown as yellow spheres have been associated with both diseases.

5.2.2 A ubiquitin-like (UBL) domain is present at the N-terminus of the extended SMCHD1 ATPase region

During the course of my studies, the first three-dimensional structure of a human SMCHD1 ATPase region was reported [133]. This structure was of a catalytically inactive point mutant of SMCHD1, E147A, which is unable to hydrolyse ATP. This study additionally identified an N-terminal ubiquitin-like (UBL) domain within SMCHD1's extended ATPase region, which was proposed to be required for dimerisation [133].

Proteins that contain UBL domains have been largely recognised for their role in recruitment of ubiquitinated substrates to the proteasome for degradation, however not in all cases. In the E3 ligase protein, Parkin, the UBL domain inhibits its auto-ubiquitination by competing with ubiquitin for binding to Parkin [212]. In the chromatin-modifying complex polycomb repressive complex 1 (PRC1), the UBL domain of its core component protein, BMI1 (PCGF4), undergoes homo-oligomerisation and thereby contributes to histone H2A Lys119 ubiquitination (H2AK119ub) [213]. Hence, UBL domains can hold various functions that are largely linked with mediating a broad range of protein interactions.

As their name suggests, UBL domains share a high structural homology to ubiquitin, which is best known for its role in mediating protein degradation upon its covalent addition onto target proteins [214]. The sequence conservation between SMCHD1's UBL domain and human ubiquitin or the UBL domain of Parkin are both relatively poor, with a sequence identity of only approximately 16% observed for both comparisons (Figure 5.2 a). However, the UBL region of SMCHD1 and ubiquitin share a similar spatial arrangement, with a Root Mean Square Deviation (RMSD) value of 5.325 Å (Figure 5.2 b,c). As distinguished from the structure of the SMCHD1 ATPase, the newly-identified UBL domain undergoes a domain-swapping event, where the UBL domain from one SMCHD1 monomer crosses over to the monomer on the opposing side (Figure 5.2 b). Yet, apart from contributing to SMCHD1 ATPase dimerisation, the functional role of the UBL domain in terms of SMCHD1's overall mechanism of function and recruitment to specific chromatin sites remains unclear.

UBL domains may possess Lys residues, which in the case of ubiquitin, function as poly-ubiquitination sites and are conserved across orthologues. While human ubiquitin contains seven Lys residues, SMCHD1's UBL domain has only four (Figure 5.2 d). Their position within the SMCHD1 UBL domain also differs compared to human ubiquitin, where only Lys80 (Lys48 in ubiquitin) appears to localise to a similar region in three-dimensional space (Figure 5.2 d). Whether these Lys residues are of functional importance in SMCHD1 remains unknown, however, experimental evidence obtained to date has not yet identified a role for ubiquitination or proteasomal degradation for SMCHD1. Yet, notably, several disease-related mutations in *SMCHD1* localise to the UBL domain, highlighting an overall functional importance (Figure 5.1 a,b).

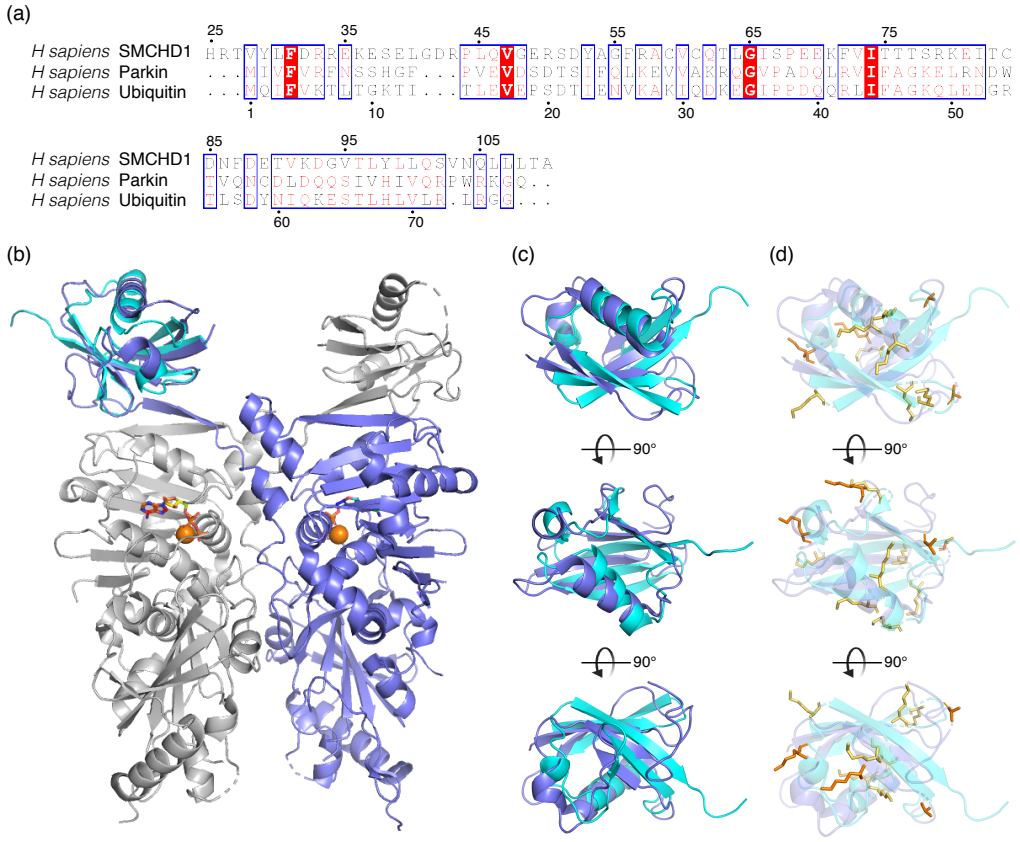


FIGURE 5.2: **A ubiquitin-like (UBL) domain is present at the N-terminus of the extended SMCHD1 ATPase region.** (a) Multiple sequence alignment comparing the *Homo sapiens* SMCHD1 UBL domain (25-110 aa) with the *Homo sapiens* Parkin UBL domain (1-76 aa) and *Homo sapiens* Ubiquitin, where conserved residues are boxed in red. Sequence alignment was obtained using MultAlin [174] and the image was generated using ESPript3.0 [175]. (b) Superimposition of the human SMCHD1 ATPase (25-580 aa; PDB: 6MW7) crystal structure with human ubiquitin (PDB:1UBQ), where two SMCHD1 monomers are depicted, in grey and slate, and ubiquitin is shown in cyan. Magnesium (Mg^{2+}) is shown as an orange sphere, and ATP is depicted in stick-mode. (c) Superimposition of SMCHD1's UBL domain (slate) and ubiquitin (cyan), with (d) lysine residues depicted for each, in orange and yellow, respectively. Images were obtained via PyMOL.

5.3 Results

5.3.1 Co-purification of the SMCHD1 ATPase (111-702 aa) with a fragment antigen-binding (Fab)

Previous crystallisation trials of the SMCHD1 ATPase region (111-702 aa) by other members of our group were unsuccessful, likely due to conformationally flexible regions in the protein. As part of pursuing new experimental approaches to aid protein crystallisation, I generated fragment antigen-binding (Fab) fragments from antibodies raised against the ATPase region of Smchd1 (clone 1D6, WEHI antibody facility), using a Pierce Fab preparation kit (Thermo Fisher). As outlined in Figure 5.3 (a), Fab fragments can be obtained from immunoglobulin G (IgG) antibodies following a digestion step with the cysteine protease, papain, resulting in one Fc portion and two Fab fragments from one antibody. Non-reducing SDS-PAGE analysis identified an obtained Fab population representative of ∼45 kDa, further evident at a molecular weight of ∼25 kDa under reducing conditions due to the removal of the inter-chain disulfide bond (Figure 5.3 b). Upon Papain-mediated digestion, the Fc region can be visualised as a ∼28-30 kDa species under both reducing and non-reducing conditions (Figure 5.3 b). A subsequent purification step using a Protein A column to bind and remove the IgG Fc portion was performed. However, as observed in Figure 5.3 (b), the Fc region remains highly abundant in the eluted samples following this step, failing to interact with the Protein A resin.

As my aim was to obtain a SMCHD1 ATPase-Fab protein complex, I first purified the SMCHD1 ATPase protein (111-702 aa) which was expressed in *Sf*21 insect cells, as previously described [61, 146]. I incubated the purified SMCHD1 ATPase with the Fab-containing sample obtained above, and performed co-purification of the protein complex via size exclusion chromatography (SEC) to remove the unbound Fc portion (Figure 5.3 c (i)). As highlighted in Figure 5.3 (c) (i-ii), the SMCHD1 ATPase region co-elutes with the Fab fragment in fractions 32-37 following SEC, whereas the Fc portion is present in later fractions, depicted by the shoulder peak in the chromatography profile (Figure 5.3 c (i)), and via SDS-PAGE analysis in fractions 39-41 (Figure 5.3 c (ii)). Fractions of interest (32-37) were pooled and concentrated, and subjected to either crystallisation trials or electron microscopy.

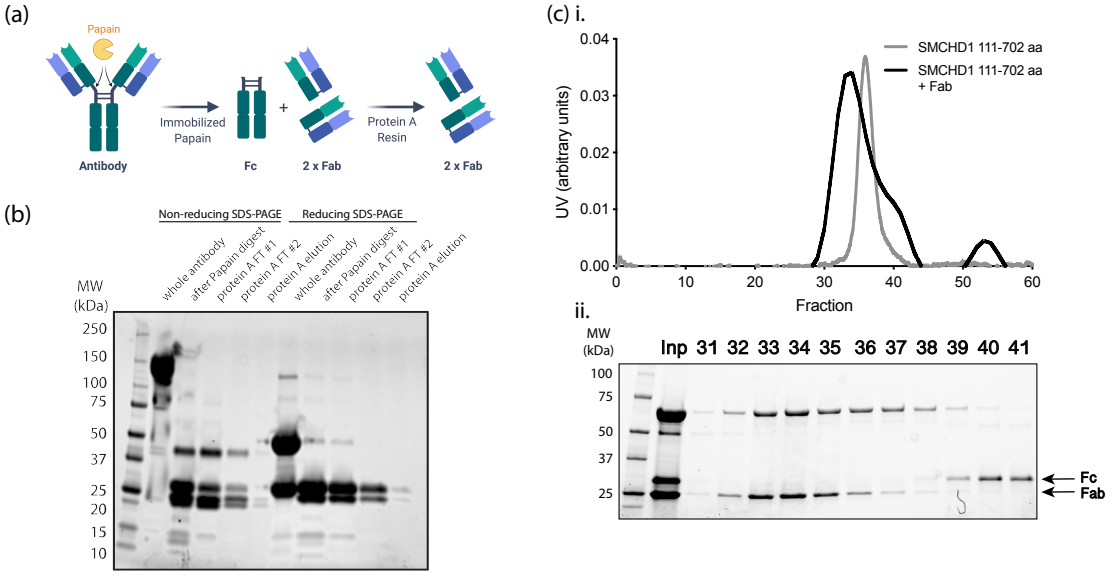


FIGURE 5.3: **Co-purification of the SMCHD1 ATPase (111-702 aa) with a fragment antigen-binding (Fab).** (a) A schematic representation depicting the Papain-mediated digestion step of IgG antibodies to obtain Fab fragments, where digestion of one antibody yields one Fc portion and two Fab fragments. The Fc portion specifically interacts with Protein A, which is used to obtain a purified Fab population. (b) Fractions collected from the Fab purification step were analysed by Stain-Free SDS-PAGE, either under non-reducing or reducing conditions. (c) (i) A SEC profile of SMCHD1 (111-702 aa) alone compared with the co-purification of SMCHD1 (111-702 aa) in complex with a Fab, with (ii) selected fractions analysed by reducing Stain-Free SDS-PAGE, with black arrows distinguishing the representative band for the Fc portion and the Fab. (b, c(ii)) Molecular weight marker position is shown on the left, in kilodaltons (kDa).

I set up crystal trays with either the SMCHD1 ATPase protein (111-702 aa) alone, in combination with an equimolar concentration of the SMCHD1 inhibitor, Radicicol, or, in complex with the Fab fragment obtained above. Crystallisation trials were set up using the Shotgun, Proplex, C3-1, C3-2, C3-3, C3-4 and C3-5 defined screens at the Collaborative Crystallisation Centre (C3), Melbourne. These were prepared using a robotic sitting drop method in 96-well plates, at 22°C and at a protein concentration of 5 mg/ml. From all 96-well plates, only one condition produced protein crystals for all three protein samples, consisting of 1 M diammonium hydrogen phosphate, 0.1 M imidazole and 0.2 M sodium chloride (Figure 5.4). As an initial method of confirming whether the observed crystals were formed from protein and not salt, imaging under UV light was performed to detect the presence of tryptophan residues, which fluoresce when excited by UV light. As observed in Figure 5.4, the obtained crystals were not illuminated by UV light. The presence of salt crystals was further confirmed by diffraction data which was collected by Dr Richard Birkinshaw at the MX2 beamline, Australian Synchrotron.

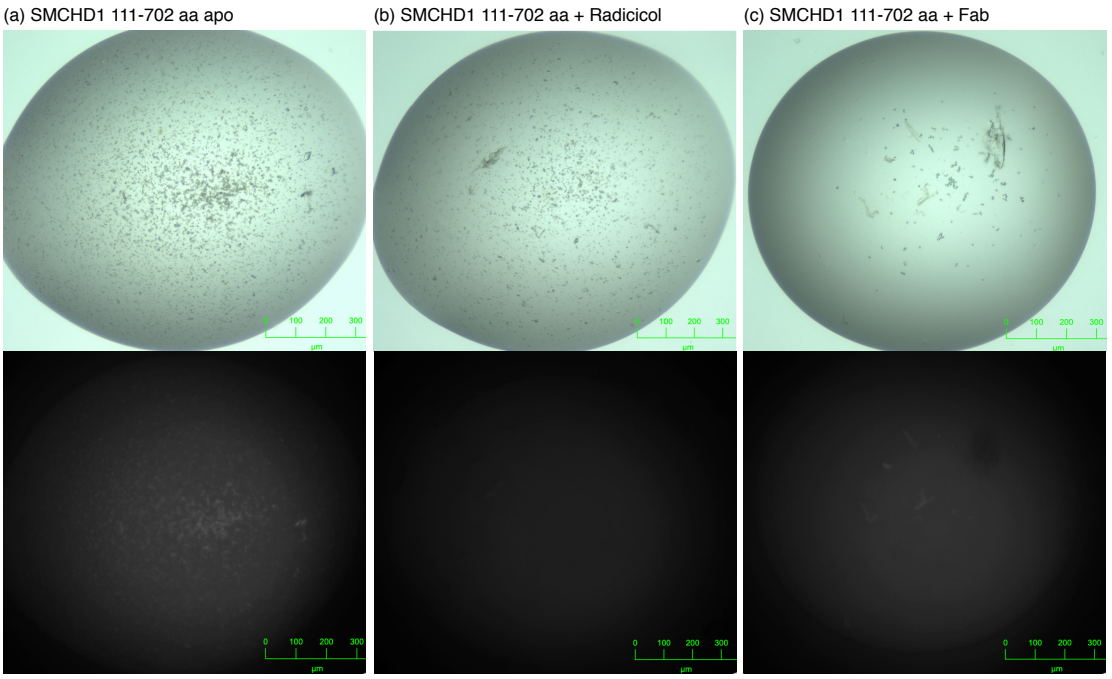


FIGURE 5.4: **Crystallisation trials of the SMCHD1 ATPase region (111-702 aa).** Images obtained upon crystallisation trials of (a) the SMCHD1 ATPase region (111-702 aa) protein alone, (b) with an equimolar amount of Radicicol or (c) in complex with a Fab fragment. (a-c) All crystals were obtained using 1 M diammonium hydrogen phosphate, 0.1 M imidazole and 0.2 M sodium chloride. The same images are shown below, but were taken under UV light. Images were acquired at the CSIRO Collaborative Crystallisation Centre (C3) facility.

I additionally subjected the purified SMCHD1 ATPase-Fab protein complex to electron microscopy analysis, using negative stain as a sample preparation method. The SMCHD1 ATPase region is a ∼65 kDa protein, representing a low molecular weight protein for negative stain electron microscopy studies, which is a technique that is commonly used for proteins larger than 100 kDa. This is mainly due to the absolute resolution being limited to ∼2 nm by the grain size of the negative stain [215]. The addition of a 45-50 kDa Fab fragment would therefore enhance the overall molecular weight of the protein complex, while furthermore providing a defined feature for image alignment. Previous small-angle X-ray scattering (SAXS) analyses of the SMCHD1 ATPase region (111-702 aa) revealed a particle diameter of 105 Å, corresponding to 10.5 nm (Figure 5.5 a) [61, 146]. Combined with a Fab fragment of ∼45-50 kDa, a slightly larger protein complex would be expected to be visualised. The resulting negative stain images of the SMCHD1 ATPase-Fab protein complex revealed a homogeneous protein population, highlighting the presence of boomerang-shaped particles of approximately 15-20 nm in length (Figure 5.5 b), corresponding to a predicted molecular size as indicated by previous SAXS analyses.

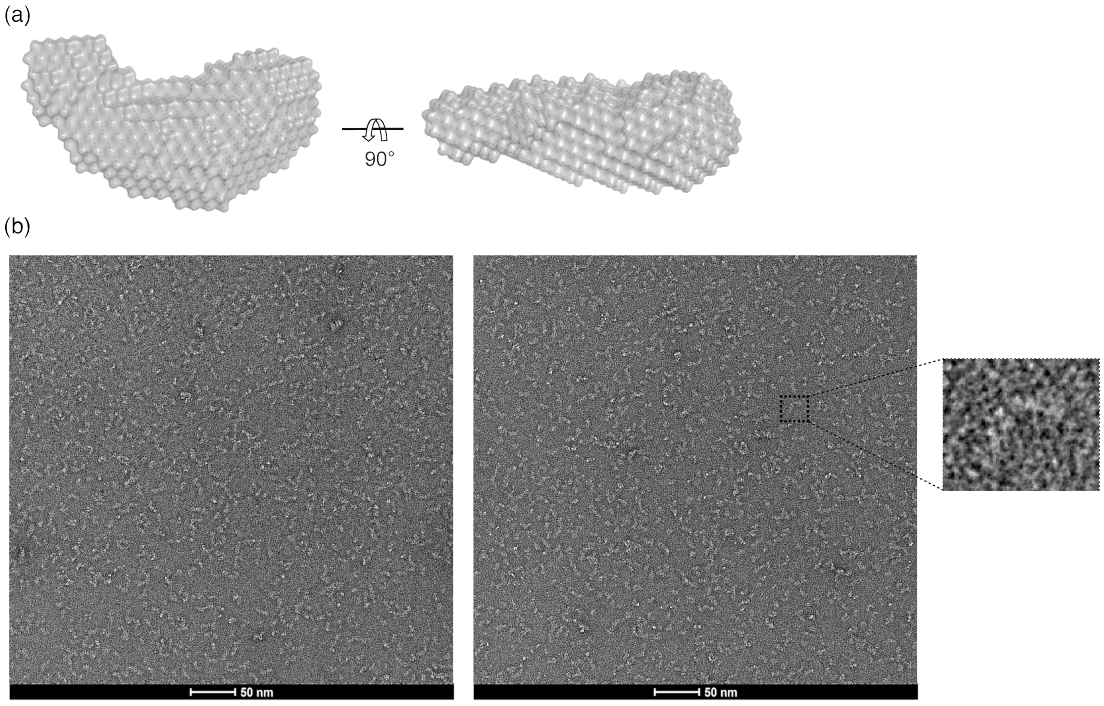


FIGURE 5.5: **Negative stain electron microscopy images of a SMCHD1 ATPase-Fab fragment protein complex.** (a) Bead model of the SMCHD1 ATPase region (111-702 aa) obtained via SAXS analyses. (b) Representative negative stain images of a SMCHD1 ATPase-Fab fragment complex, with a zoomed-in image of one particle shown on the right. Images were obtained via a FEI Talos L120C transmission electron microscope at x150,000 magnification. Samples were prepared using a 1% (v/v) uranyl acetate solution.

5.3.2 Reductive lysine methylation as a protein crystallisation strategy

Reductive surface lysine methylation of protein samples has been used previously to improve the likelihood of protein crystallisation [172, 216]. Surface residues, such as lysines, have long side chains and therefore a high conformational entropy that may prevent the formation of a highly ordered crystal lattice. Reductive methylation can be achieved chemically, where the primary amine groups of lysine residues are modified to tertiary amines [172]. Upon performing the methylation step as outlined by Walter *et al.* [172], the reaction led to a population of precipitated SMCHD1 ATPase protein, which was removed by centrifugation. The soluble SMCHD1 protein was then subjected to size exclusion chromatography (Figure 5.6 a). This step led to the isolation of lysine-methylated SMCHD1 ATPase protein that was free from aggregates, eluting as a single peak that is comparable to the elution profile of unmodified SMCHD1 protein (Figure 5.6 a). However, upon crystallisation trials using the defined conditions in the Shotgun, Proplex, C3-5 and C3-6 screens (CSIRO Collaborative Crystallisation Centre (C3), Melbourne), no crystals were obtained.

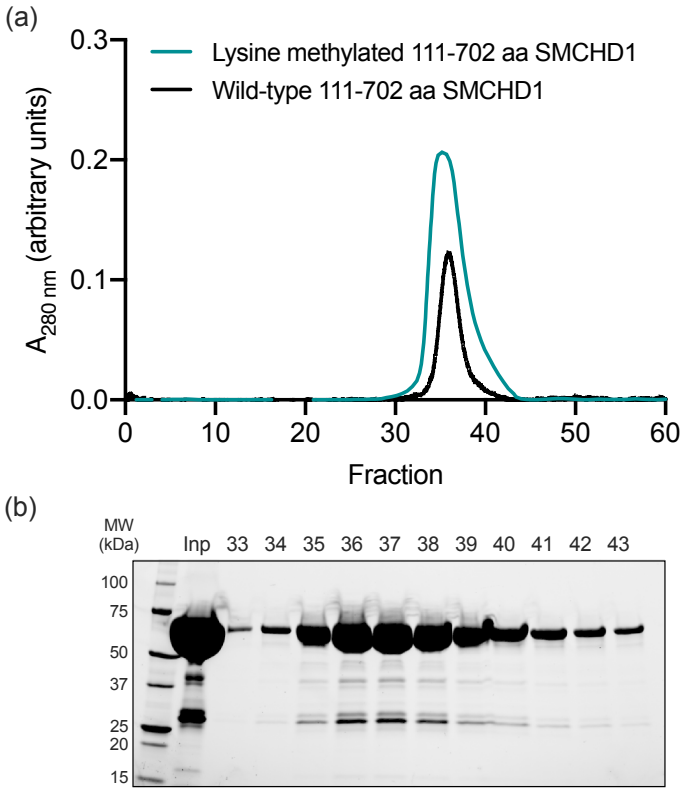


FIGURE 5.6: **Reductive lysine methylation of the SMCHD1 ATPase region (111-702 aa) prior to crystallisation trials.** (a) A size exclusion chromatography profile of wild-type SMCHD1 111-702 aa (black line) and SMCHD1 111-702 aa following a lysine methylation step (green line), showing (b) selected lysine methylated SMCHD1 fractions analysed by reducing Stain-Free SDS-PAGE. (b) Molecular weight marker position is shown on the left, in kilodaltons (kDa).

5.3.3 Limited proteolysis of the SMCHD1 ATPase region

Following many unsuccessful crystallisation trials of the SMCHD1 ATPase region, it seemed likely the 111-702 aa construct was not well suited for crystallisation, possibly due to the presence of unstructured or highly dynamic regions. Limited proteolysis involves the incubation of a purified protein sample with a limited amount of protease that is able to undergo sequence-specific cleavage of the peptide. This method can be used to examine the conformational features of a protein, under the assumption that resistance to proteolytic cleavage denotes a well-folded, sterically unavailable region to the protease. Conversely, sites of limited proteolysis often occur within areas of protein flexibility, such as loops or poorly-folded regions, hence this method may be used to identify sites of conformational flexibility in a protein. Consequently, by removing regions that contribute to protein conformational heterogeneity, a protein sample that is more amenable to structural elucidation by X-ray crystallography may be obtained.

I performed limited proteolysis on the purified SMCHD1 ATPase region (111-702 aa) using the serine proteases: trypsin and chymotrypsin. Trypsin cleaves peptides at arginine and lysine residues, while chymotrypsin has a preference for larger hydrophobic residues, such as tryptophan, tyrosine, phenylalanine, and methionine. Following proteolytic cleavage by either trypsin or chymotrypsin, a protein fragment of a molecular weight of ∼50 kDa was identified by SDS-PAGE (Figure 5.7). I excised this protein band and subjected it to N-terminal sequencing (Australian Proteome Analysis Facility, Macquarie University) to identify the site of proteolysis. The N-terminal 6-amino acid sequence was identified as SLNSDI, encompassing residues 227-231 of SMCHD1, denoting a solvent-exposed and poorly structured region within SMCHD1's ATPase.

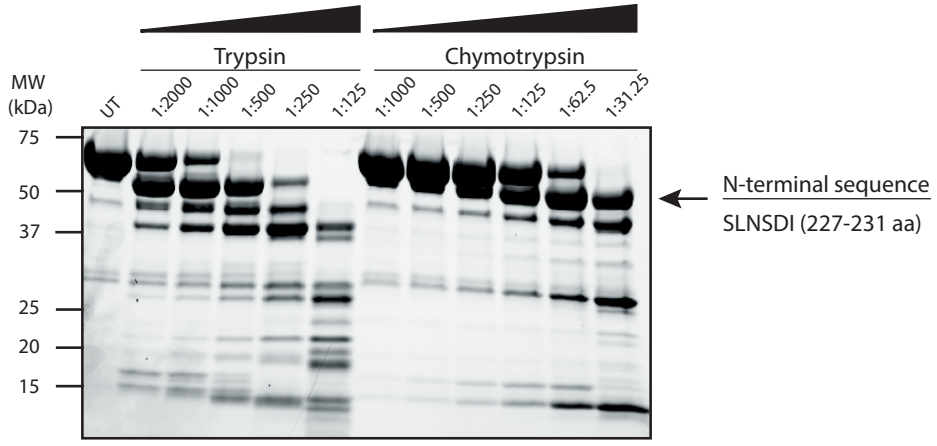


FIGURE 5.7: **Limited proteolysis of the SMCHD1 ATPase region (111-702 aa) with trypsin and chymotrypsin.** Reducing Stain-Free SDS-PAGE analysis depicting detected protein fragments following proteolytic digestion of the SMCHD1 ATPase protein (111-702 aa). Concentrations of trypsin and chymotrypsin used are indicated above each sample as a dilution factor from a starting concentration of 1 μg/μl, where an untreated SMCHD1 protein sample (UT) was used as a comparison. The fragment chosen for N-terminal sequencing and the 6-amino acid sequence subsequently identified is depicted by a black arrow, present as the second band following either trypsin or chymotrypsin-mediated proteolysis. Molecular weight marker position is shown on the left, in kilodaltons (kDa).

5.3.4 Hydrogen Deuterium Exchange (HDX) and mass spectrometry (MS) of the SMCHD1 ATPase (111-702 aa) reveals flexible regions in the protein

To further confirm the results obtained above via limited proteolysis and to identify the exact boundaries of regions within the SMCHD1 ATPase that display large conformational dynamics, I performed hydrogen deuterium exchange (HDX) coupled with mass spectrometry (MS) analyses with Dr Jarrod Sandow (Walter and Eliza Hall Institute). HDX-MS is a powerful tool often used to identify dynamic regions within a protein [217]. HDX relies on the exchange of hydrogen atoms with deuterium, which is incorporated into the surrounding solvent, where the most accessible regions within a protein become more rapidly deuterated. As the deuterium that becomes incorporated into side chain groups is rapidly back-exchanged, this method is sensitive to the backbone amide exchange. A global uptake of deuterium can be monitored by quenching of the HDX reaction and subsequent proteolytic digestion at various time points, giving rise to peptides that can be subjected to MS analysis [217]. In addition, by revealing the most and least solvent-exposed regions of the protein, this technique also provides a global indication of protein fold, thereby facilitating structural interpretation.

For the HDX-MS analysis of SMCHD1's ATPase, I used four different deuterium labeling times, of 30, 60, 600 and 3600 seconds, to ensure sufficient time was provided to allow deuterium uptake of the majority of amides in the protein, while also allowing the capture of fast exchange events with the use of shorter time ranges. A 97.46% sequence coverage was achieved, with obtained results highlighted in Figure 5.8 (a). Regions that were most protected from solvent deuterium uptake can be observed across several locations within the protein, such as where the percentage of deuteration was below 10% after a 30-second pulse, correlating with an involvement in protein secondary and tertiary structure. However, proteins are highly dynamic in nature, exhibiting local and global conformational fluctuations, where even stable secondary structural elements undergo transient breakage and re-formation of hydrogen bonds. Such events are often represented by time-dependent increases in deuteration levels, and are typically indicative of secondary structural elements. As depicted in Figure 5.8

(a), these trends were detected at localised regions throughout the SMCHD1 ATPase protein.

The highest levels of deuterium incorporation, following only 30 seconds of labeling, occurred across residues ∼200-230 aa, 270-300 aa and 670-690 aa (Figure 5.8 a,c). Importantly, these regions exhibited a similarly elevated percentage of deuteration across all time-points and are therefore indicative of solvent-exposed structures within the protein, which do not partake in hydrogen bonding. Interestingly, the rapidly-deuterated region encompassing residues 200-229 (Figure 5.8 b) incorporates the previously identified proteolytic site following both trypsin- and chymotrypsin-mediated limited proteolysis of the protein (Figure 5.7). I was therefore interested in further exploring this potential loop region and examining whether its removal may impact the protein folding or activity of SMCHD1's ATPase region.

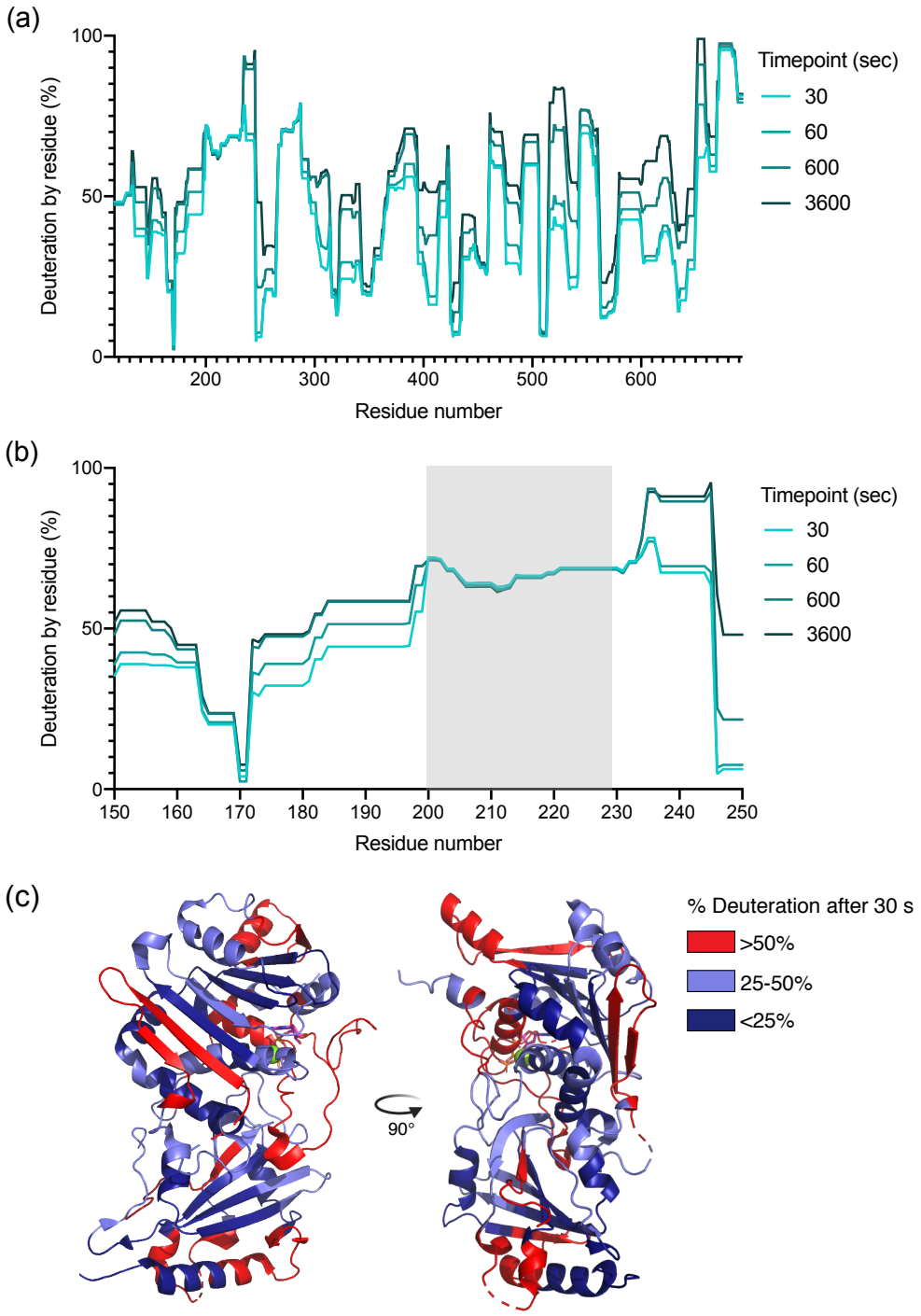


FIGURE 5.8: **Hydrogen Deuterium Exchange (HDX) mass spectrometry of the SMCHD1 ATPase (111-702 aa) reveals flexible regions in the protein.** (a) Full experimental results with (b) a zoomed-in version (150-250 aa) highlighting a region of interest (200-229 aa) in grey. (a-b) The x-axis represents the residue number of the SMCHD1 (111-702 aa) protein and the y-axis depicts residue deuteration as an absolute percentage. HDX labelling was carried out over four time points of 30, 60, 600 and 3600 seconds, as indicated by the legend on the right. (c) A cartoon representation highlighting regions of deuterium exchange within the SMCHD1 ATPase (25-580, PDB: 6MW7) following 30 seconds of labeling. A colour legend is shown on the right, depicting regions that were less than 25% (dark blue), between 25-50% (slate) or above 50% (red) deuterated following the 30 second labeling time.

5.3.5 Removal of a loop region from the SMCHD1 ATPase (111-702 aa) diminishes the protein's catalytic activity

As shown above, both limited proteolysis and subsequent HDX-MS analysis of the SMCHD1 ATPase region (111-702 aa) revealed a flexible region in the protein that corresponds to an "ATP-lid" structure across members of the GHKL family of proteins [99]. While varying conformations and sequence lengths are present among the ATP-lid structures across members of the GHKL family, all hold a conserved role in ATPase activity regulation [99]. I was therefore interested in investigating whether the identified dynamic region in the SMCHD1 ATPase may represent the conserved ATP-lid structure and may therefore also hold a role in regulating the protein's catalytic activity. To do so, I used overlapping PCR to remove the sequence encoding the residues 200-229 from the SMCHD1 111-702 aa construct. These residues encompass the majority of the identified flexible region, leaving only a few extra amino acids to minimise any potential perturbation to the surrounding secondary structural regions that may be induced by the loop deletion. The protein was successfully expressed in *Sf*21 insect cells and purified via a TEV protease-cleavable N-terminal His_6 tag. Subsequent size exclusion chromatography revealed a predominant peak, highlighted by fractions ~37-41 and further confirmed as representative of a ~60 kDa protein by SDS-PAGE analysis (Figure 5.9 a-b). These fractions were pooled and concentrated.

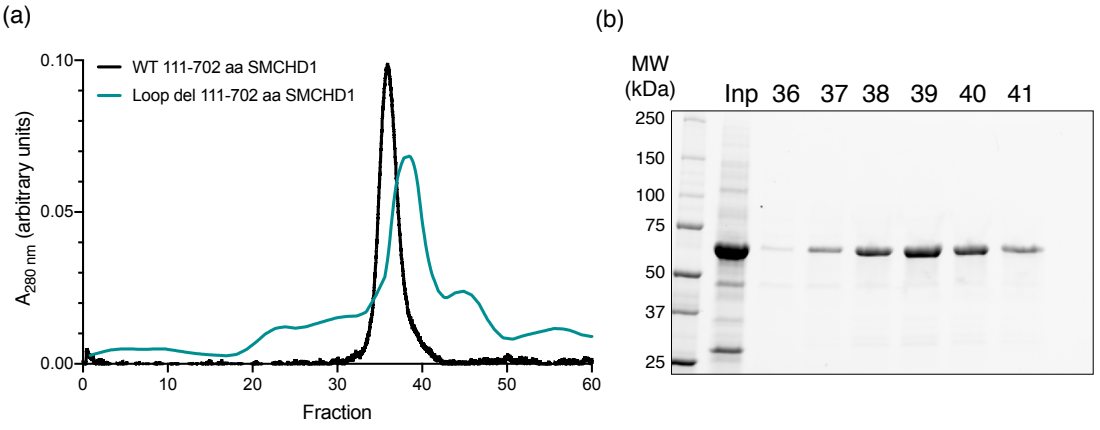


FIGURE 5.9: **Size exclusion chromatography (SEC) of a loop-deleted SMCHD1 ATPase.** (a) SEC profile of a wild-type 111-702 aa SMCHD1 ATPase (black) compared to a loop-deleted version, lacking residues 200-229 (teal). (b) Reducing Stain-Free SDS-PAGE analyses of the input (Inp) sample and selected fractions following SEC of the loop-deleted SMCHD1 ATPase region. Molecular weight (MW) indicators are shown on the left in kilodaltons (kDa).

Upon the release of the three-dimensional structure of a human SMCHD1 ATPase (25-580 aa), which was not available at the time these experiments were performed, I was able to map the position of the flexible loop region. As depicted in Figure 5.10 (a), the identified loop region corresponds to the ATP-lid structure identified among GHKL proteins. The structure is positioned over the ATP-binding site within the GHKL ATPase pocket, and consists of a short α-helical unit. The structural representation of ATP-lid residues 200-229 highlights the small remainder of the loop that is left, allowing contact to the adjacent α-helix without perturbing the local protein fold (Figure 5.10 a). Upon *in vitro* ATPase analyses, an evident reduction in the catalytic rate of the loop-deleted SMCHD1 ATPase was observed compared to the wild-type protein (Figure 5.10 b-c). While wild-type SMCHD1 exhibited a catalytic rate of $\sim$0.03 μM ADP/min/μM protein, the loop-deleted protein displayed approximately a three-fold reduction in ATPase activity, with a catalytic rate of $\sim$0.01 μM ADP/min/μM protein. These results are therefore consistent with a regulatory role of the ATP-lid in SMCHD1's catalytic activity, as established for other GHKL proteins [115, 218].

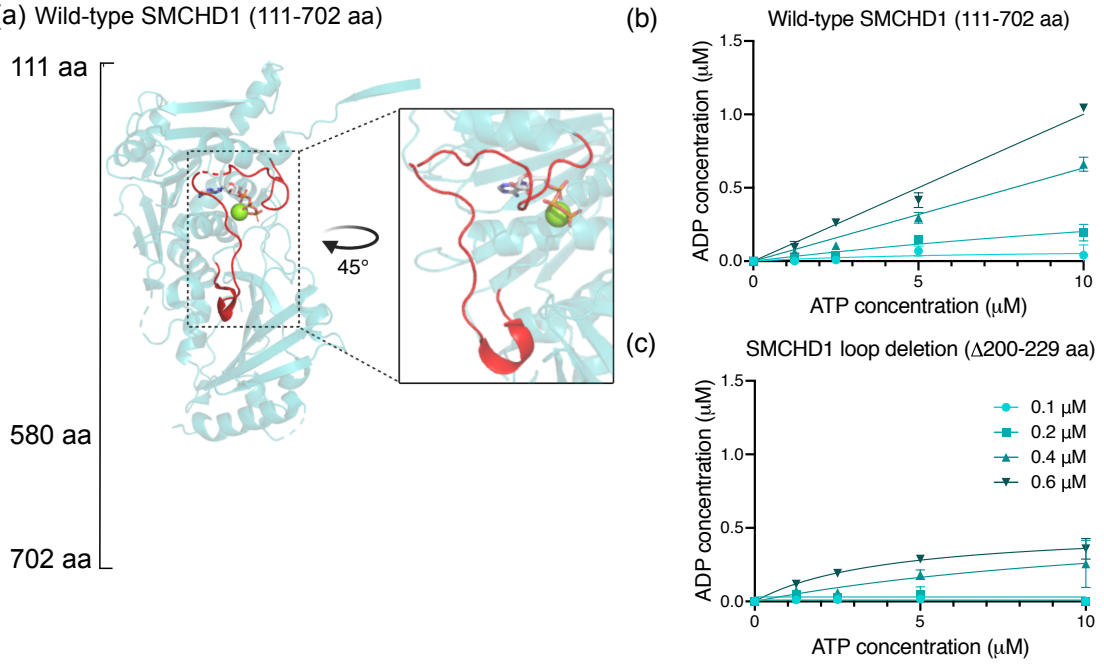


FIGURE 5.10: **Removal of a loop region from the SMCHD1 ATPase (111-702 aa) diminishes the protein's catalytic activity.** (a-b) Showing the three-dimensional structure of the SMCHD1's ATPase (25-580 aa; PDB: 6MW7) as part of the 111-702 aa SMCHD1 construct used in this experiment. (a) Highlighting the loop region of interest that resides over the active site on SMCHD1, where the deleted residues 200-229 are shown in red. ATP is shown in stick form and Mg^{2+} is depicted as a green sphere. ATPase assay results of (b) wild-type SMCHD1 compared to (c) SMCHD1 with the loop deletion. (b-c) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM). Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent $\pm$ standard deviation (SD) of the mean, and are representative of two independent experiments.

5.3.6 Presence of the ubiquitin-like (UBL) domain or a C-terminal extension do not alter the catalytic activity of the SMCHD1 ATPase region

While trialing various protein crystallisation strategies of the SMCHD1 ATPase 111-702 aa construct, the first three-dimensional structure of a human SMCHD1 ATPase (25-580 aa) emerged [133]. Upon identifying various discrepancies in the experimental procedures presented in the study, I aimed to reproduce and expand on selected findings. I was foremost interested in investigating whether the wild-type SMCHD1 ATPase underwent dimerisation in the presence of the newly-identified UBL domain (25-110 aa), or whether the observed dimerisation of the E147A variant presented in the study was potentially induced by the point mutation or was the result of a crystallisation artefact.

I initially expressed and purified two of the SMCHD1 ATPase constructs presented in the published study, comprising of residues 25-580 and 111-580. In order to pursue their experimental procedures, I originally used bacterial cells as an expression system, however, a poor protein yield was obtained (result not shown). I therefore expressed the constructs in *Sf*21 cells instead, a system that I previously used to successfully express the 111-702 aa SMCHD1 construct. Subsequent purification steps included IMAC followed by SEC, for which resulting profiles are displayed in Figure 5.11 (a). Both new constructs eluted at a position most consistent with a monomer form, comparable to the profiles obtained in the published study [133].

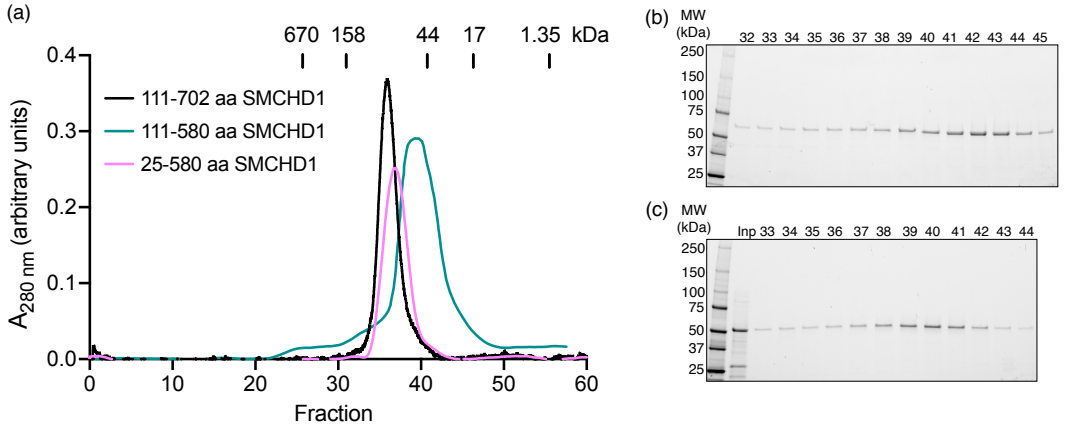


FIGURE 5.11: **Size exclusion chromatography (SEC) of new SMCHD1 ATPase constructs.** (a) Comparing the SEC elution profiles of SMCHD1 constructs 111-702 aa (black), 25-580 aa (pink) and 111-580 aa (teal), with molecular weight standard values indicated above in kilodaltons. (b-c) Reducing Stain-Free SDS-PAGE analysis of selected fractions following SEC of the (b) 111-580 aa and (c) 25-580 aa constructs. Molecular weight marker position is shown on the left, in kilodaltons (kDa).

Following the successful expression and purification of the two new SMCHD1 ATPase constructs, I sought to determine their *in vitro* catalytic rates. Comparing 25-580 aa and 111-580 aa SMCHD1 would therefore elucidate whether the newly-identified UBL domain (25-110 aa) alters SMCHD1's ATPase activity, whereas the 111-702 aa construct I previously investigated excludes the UBL region but extends past the transducer domain (Figure 5.12 a). As shown in Figure 5.12 (b-d), all three tested proteins exhibited comparable catalytic rates. At low protein concentrations (0.1 and 0.2 μM), SMCHD1 constructs 25-580 aa and 111-580 aa exerted a ~three-fold higher ATP turnover than the 111-702 aa SMCHD1 counterpart, however similar trends were observed at higher protein concentrations (0.4 and 0.6 μM) across all three constructs (Figure 5.12 b-d).

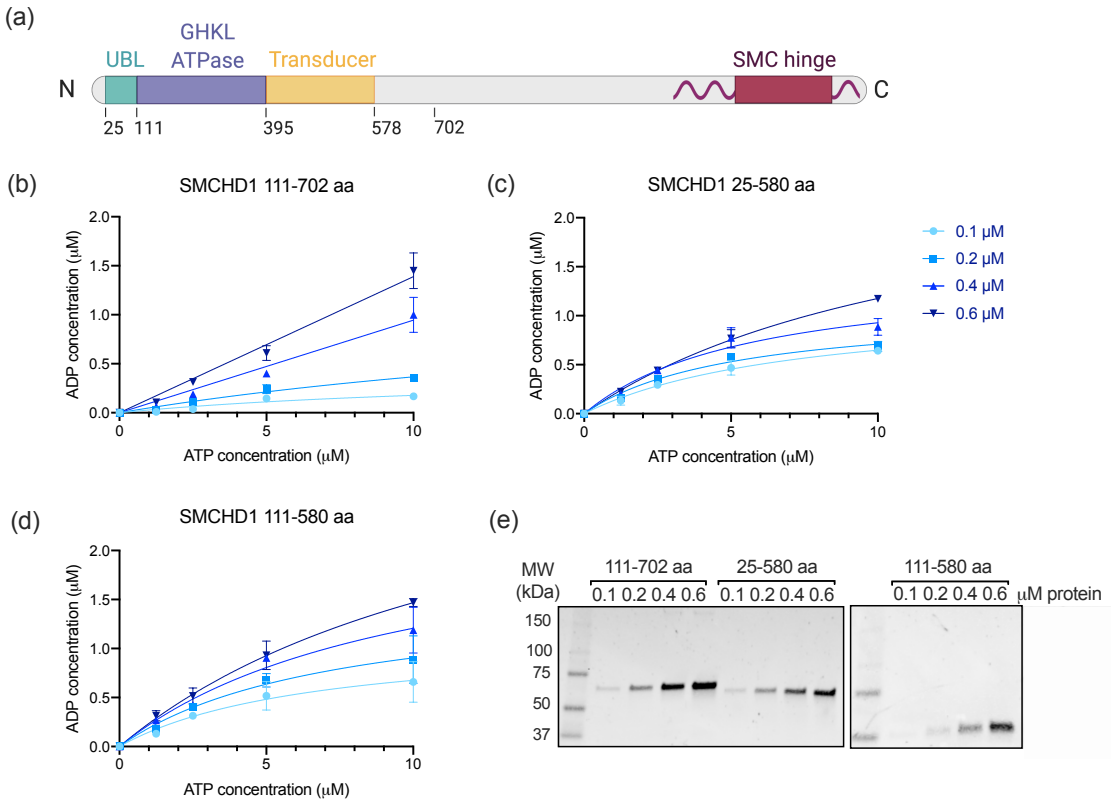


FIGURE 5.12: **Comparing the catalytic rates of various SMCHD1 ATPase constructs.** (a) A schematic representation depicting the gene architecture of full-length Smchd1, with domain boundaries of the ATPase region indicated below (in amino acid residues). (b-d) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM). Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent ± standard deviation (SD) of the mean. (e) Reducing Stain-Free SDS-PAGE analysis of prepared Smchd1 protein dilutions used in the corresponding assays. Molecular weight marker position is shown on the left, in kilodaltons (kDa).

I additionally aimed to express and purify a new construct that encompasses residues 25-702 of SMCHD1, thereby providing the most extended version of a SMCHD1 ATPase region that contains both the N-terminal UBL domain (25-110 aa) and the C-terminal extension following the transducer domain (580-702 aa). As depicted in Figure 5.13, this construct was well expressed and eluted as a single peak following SEC analysis, corresponding to a monomeric species that is analogous to the elution profile of the 111-702 aa SMCHD1.

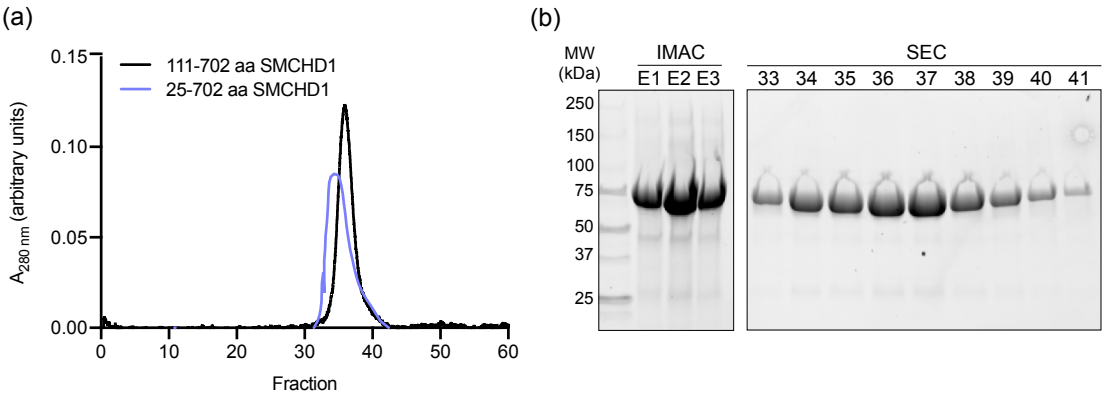


FIGURE 5.13: **Size exclusion chromatography (SEC) of the 25-702 aa SMCHD1 ATPase construct.** (a) Comparing the SEC elution profiles of SMCHD1 constructs 111-702 aa (black) and 25-702 aa (purple). (b) Reducing Stain-Free SDS-PAGE analysis of elutions fractions (E1-E3) following the first IMAC purification step, followed by selected fractions obtained following SEC of the 25-702 aa construct. Molecular weight marker position is shown on the left, in kilodaltons (kDa).

The next step was to analyse the catalytic activity of the 25-702 aa SMCHD1 construct compared to the 111-702 aa construct that lacks the UBL domain. I additionally performed this experiment in the presence and absence of the SMCHD1 inhibitor, Radicicol [61, 131]. As highlighted in Figure 5.14, the two SMCHD1 ATPase constructs exhibit a comparable catalytic rate, of an estimated 0.03 μM ADP/min/μM protein, consistent with the previously examined wild-type SMCHD1 ATPase constructs (Figure 5.12). As expected, both constructs displayed a diminished catalytic rate in the presence of 10 μM Radicicol. While the ATPase activity of the UBL-containing 25-702 aa SMCHD1 was completely abolished (Figure 5.14 b), the 111-702 aa construct displayed very low levels of catalytic activity in the presence of Radicicol. However, the compound's efficacy was previously demonstrated for this construct and shown to fully abrogate its ATPase function [61, 131]. The low activity levels observed are therefore thought to arise from contaminating proteins. Together,

these results therefore demonstrate that the presence of both the N-terminal UBL domain and the C-terminal extension do not alter the catalytic activity of SMCHD1's ATPase, additionally displaying an equal inhibition by Radicicol.

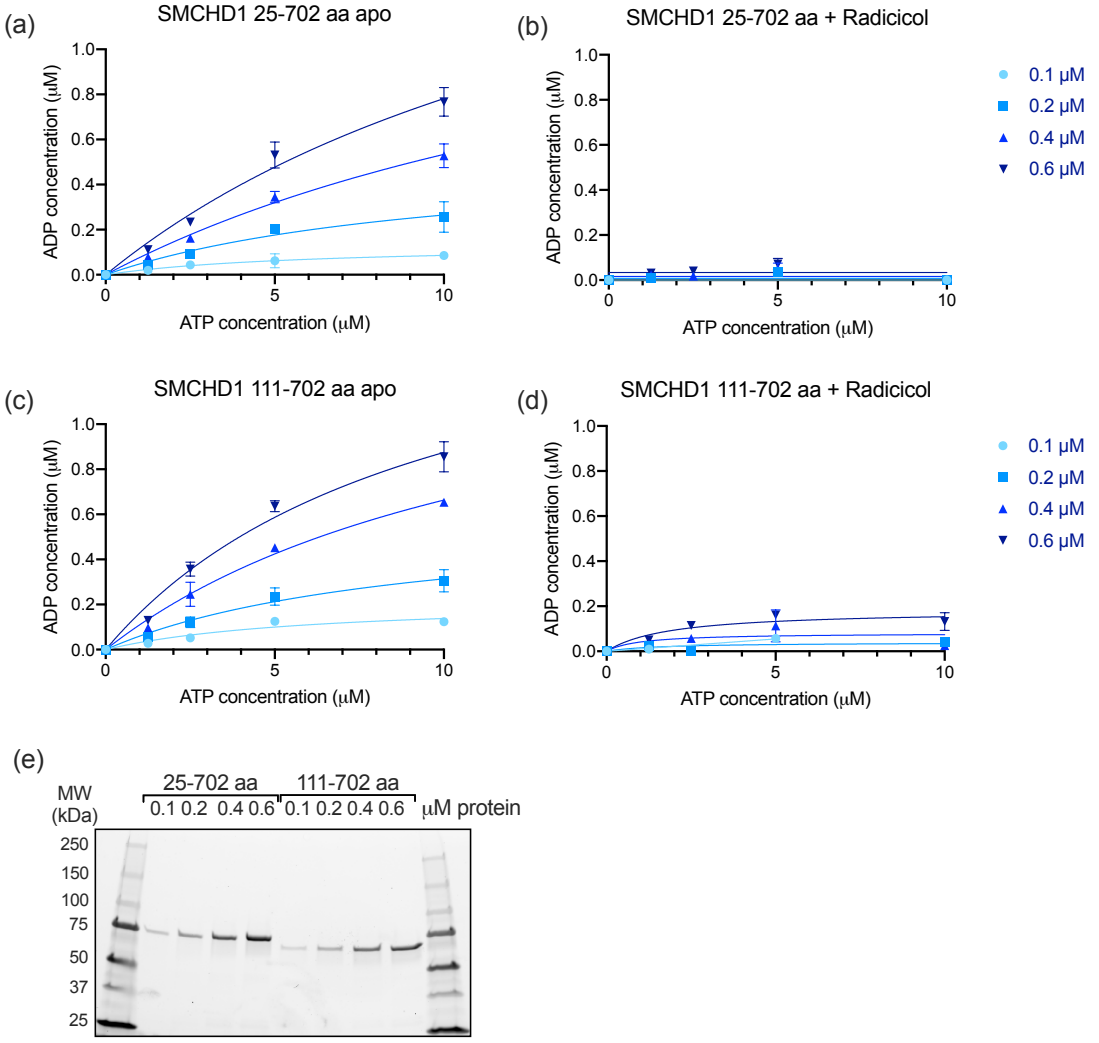


FIGURE 5.14: **The presence of the ubiquitin-like (UBL) domain does not alter the catalytic activity of the SMCHD1 ATPase region.** Comparing the *in vitro* catalytic rates of SMCHD1 ATPase constructs (a) 25-702 aa and (c) 111-702 aa, (b,d) in the presence of 10 μM Radicicol, respectively. (a-d) The x-axis denotes the ATP concentrations used: 1.25, 2.5, 5.0 and 10 μM and the y-axis indicates ADP produced (μM). Protein concentrations tested are displayed in the different shades of blue as indicated in the legend on the right. Each measurement was performed in triplicates, where error bars represent $\pm$ standard deviation (SD) of the mean, and are representative of two independent experiments. (e) Reducing Stain-Free SDS-PAGE analysis of prepared Smchd1 protein dilutions used in the corresponding assays. Molecular weight marker position is shown on the left, in kilodaltons (kDa).

5.3.7 Analytical ultracentrifugation (AUC) studies reveal dimerisation of the wild-type SMCHD1 ATPase

Despite providing valuable insights into the molecular structure of SMCHD1, the recently published structure of the human SMCHD1 ATPase represents the catalytically inactive point mutant of the ATPase, for which preferential dimerisation was observed over the wild-type counterpart [133]. These results therefore prompted an interest in establishing whether the wild-type SMCHD1 ATPase adopts a similar dimeric conformation to the E147A mutant, alongside determining the domain's dimerisation parameters via biophysical experiments such as analytical ultracentrifugation (AUC).

AUC is a commonly used method for determining the self-association properties of proteins by monitoring their sedimentation velocities in a centrifugal field [219]. As I was interested in investigating the dimerisation properties of the wild-type SMCHD1 ATPase region, I performed sedimentation velocity analyses for two wild-type SMCHD1 ATPase constructs, 111-702 aa and 25-702 aa, to validate the reported involvement of the UBL domain in SMCHD1's dimerisation [133]. I additionally examined the requirement of AMPPNP, a non-hydrolysable form of ATP, and Mg^{2+}, which is a co-factor for the ATP hydrolysis reaction, for the dimerisation of the two constructs. As outlined in Figure 5.15 (a-b), the 111-702 aa SMCHD1 ATPase region that lacks the UBL domain is present as a monomeric species both in the absence and presence of 1 mM AMPPNP/Mg^{2+}, with sedimentation coefficient (S) values of 4.09 and 4.10, respectively, and a predicted molecular weight of ∼65 kDa which is consistent with the calculated value of 68 kDa for this construct (Table 5.1). The UBL-containing SMCHD1 ATPase encompassing residues 25-702 was also identified as a monomeric species in the absence of ligand, with a sedimentation coefficient value of 4.45 and an indicated molecular weight of ∼74 kDa, corresponding to the calculated value of 78 kDa (Figure 5.15 c, Table 5.1). In the presence of 1 mM AMPPNP/Mg^{2+}, however, a higher molecular weight species emerged, depicted by a second peak that coincides with a higher sedimentation coefficient of 7.01 and is representative of a molecular weight of ∼126 kDa (Figure 5.15 d, Table 5.1). These results therefore confirm the dimerisation of the wild-type SMCHD1 ATPase, an event that is dependent on the presence of both the N-terminal UBL domain and ligand.

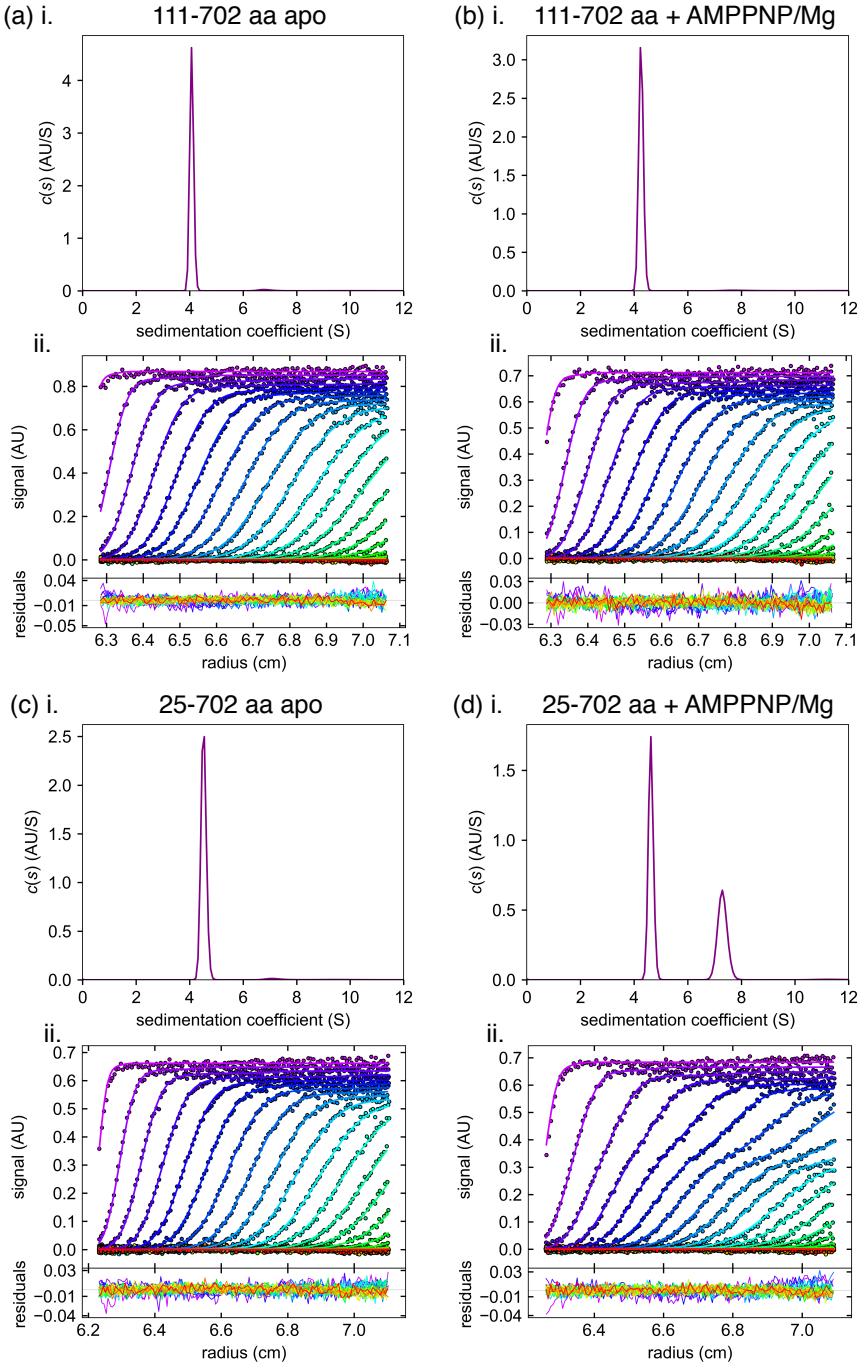


FIGURE 5.15: **Analytical ultracentrifugation (AUC) studies reveal dimerisation of the wild-type SMCHD1 ATPase.** Sedimentation velocity analyses demonstrate that the 111-702 aa SMCHD1 ATPase construct is a monomer in solution in both (a) in the absence and (b) the presence of 1 mM AMPPNP and Mg^{2+}. The SMCHD1 25-702 aa construct that contains the UBL domain is (c) monomeric in the absence of ligand but (d) undergoes AMPPNP and Mg^{2+}-dependent dimerisation. (a-d) Panels (i) represent a continuous size, c(s), distribution plotted as a function of the sedimentation coefficient (S), whereas panels (ii) (top) represent the raw sedimentation profiles of absorbance at 290 nm versus cell radius and (bottom) the residuals for the continuous size, c(s), distribution best fits plotted as a function of radial position (cm) from the axis of rotation. Analyses were performed using the program SEDFIT [166].

TABLE 5.1: **Summary of sedimentation velocity analysis of SMCHD1 ATPase constructs.**

	Wavelength (nm)	Sedimentation coefficient (S)	Molar mass (kDa)	Calculated molar mass (kDa)	Frictional ratio (f/f_0)	Fit RMSD
SMCHD1 111-702 aa apo	290	4.09	65.1	68.0	1.33	0.007
SMCHD1 111-702 aa + 1mM AMPPNP/$MgCl_2$	290	4.10	65.3	68.0	1.33	0.007
SMCHD1 25-702 aa	290	4.45	73.8	78.0	1.39	0.006
SMCHD1 25-702 aa + 1mM AMPPNP/$MgCl_2$	290	7.01	126.0	78.0	1.21	0.006
Fixed parameters						
Buffer density (g/cm^3)	1.004					
Buffer viscosity (cP)	1.027					
Partial specific volume	0.7369					

5.3.8 Mutations within Smchd1's ATPase region alter the cellular localisation of the full-length protein

To assess the cellular effects of *SMCHD1* ATPase mutations on full-length Smchd1 function, I transfected wild-type or mutant forms of full-length Smchd1 constructs into SMCHD1-knockdown (KD) HEK293 cells, a system that I previously described in Chapter 4. Human SMCHD1-KD cells were transfected with mouse full-length Smchd1 constructs to restrict the shRNA-mediated KD to the human SMCHD1, thereby confining the investigation to the introduced mouse Smchd1 constructs. As previously mentioned in Chapter 4, transfection of wild-type full-length Smchd1 into SMCHD1-KD HEK293 cells results in the formation of multiple nuclear foci, a localisation pattern that differs from the non-silencing control cells where two bright nuclear foci are observed, which correspond to SMCHD1's localisation to the two inactive X-chromosomes present in tetraploid 293 cells (Figure 5.16 a,c). Smchd1's abundant presence and multiple nuclear foci formation upon transfection in SMCHD1-KD cells is likely due to the over-expression of the construct. All mutants of interest are thereby compared to the over-expression of the wild-type Smchd1 construct rather than the non-silencing shRNA control.

I was interested in three *Smchd1* mutations that localise to the ATPase region of Smchd1: E147A, A667E and S135C. E147A represents an established catalytically inactive mutant that is unable to hydrolyse ATP, representing a loss-of-function variant [61, 131, 133]. A667E represents a neomorphic mutant that was initially shown to result in an enhanced Smchd1 function *in vivo*, however more recent studies suggest it does not always result in a gain-of-function phenotype (manuscript in preparation). Lastly, S135C is a BAMS-associated variant that exhibited the strongest gain in

ATPase activity following an *in vitro* ATPase assay, of approximately three-fold compared to wild-type Smchd1 (Figure 3.1) [146]. Immunofluorescence results following transfection of either three full-length Smchd1 ATPase variants into SMCHD1-KD HEK293 cells all indicated a distinct nuclear localisation pattern (Figure 5.16 d-f). All introduced mutations resulted in a failure of Smchd1 localisation to nuclear foci, exhibiting a very diffuse Smchd1 staining pattern that differs from wild-type Smchd1 transfected cells (Figure 5.16 c). Overall, these results suggest that mutations that either abolish or enhance Smchd1's ATPase activity *in vitro* also alter the protein's ability to localise to chromatin in the transient transfection cellular assay.

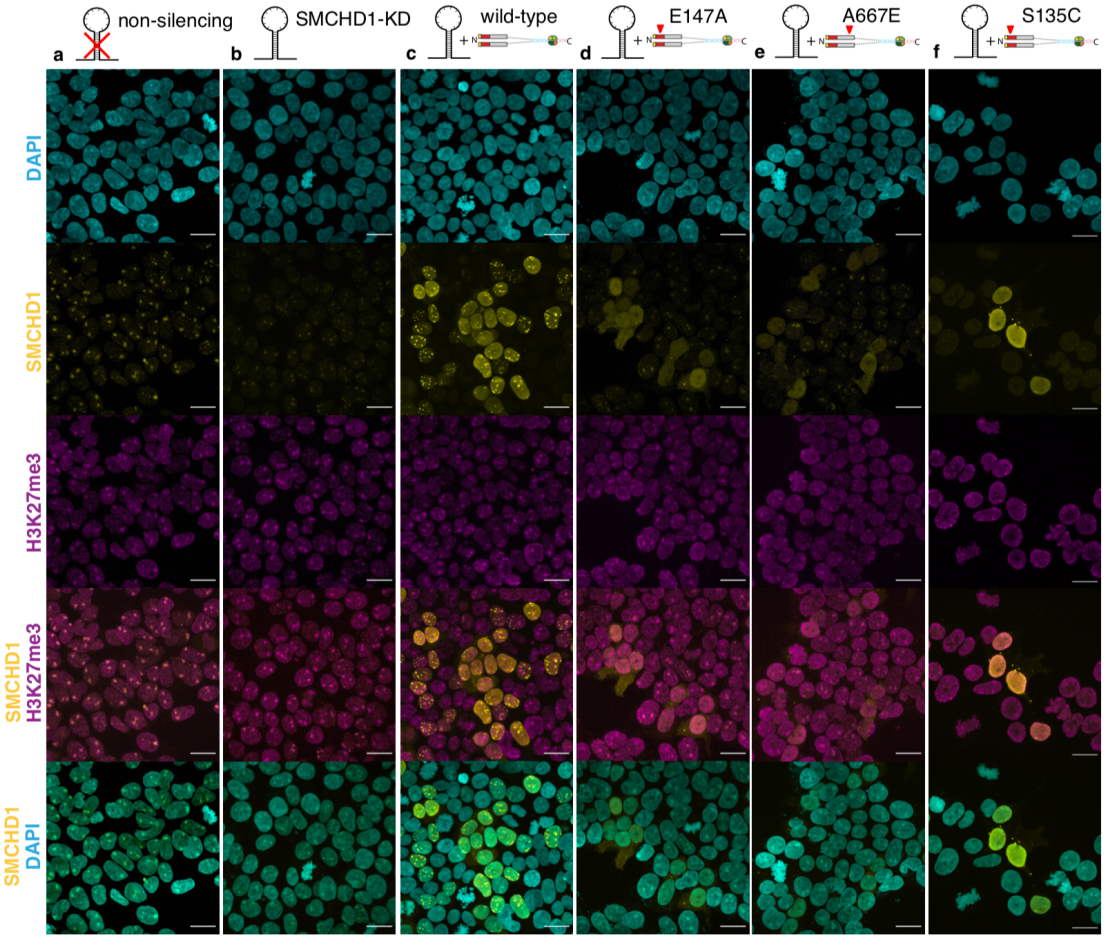


FIGURE 5.16: **Mutations within Smchd1's ATPase region alter the cellular localisation of the full-length protein.** Immunofluorescence in (a) control, non-silencing shRNA transduced, (b) shRNA-mediated SMCHD1-knockdown 293 cells and (c) SMCHD1-knockdown 293 cells 24 hours post-transfection with the full-length Smchd1 containing point mutations (d) E147A, (e) A667E or (f) S135C. DAPI staining is depicted in blue, Smchd1 staining is shown in yellow and H3K27me3 is shown in purple, with merged images shown below as indicated. Maximum intensity projection images are shown as representative of n>100 nuclei positive for Smchd1 overexpression per sample; data shown are representative of 2 independent experiments. All images were obtained with identical settings between controls and all transfected cells, to enable comparison between the images provided in the figure. Scale bars, 20 μm.

5.3.9 Smchd1's UBL domain is required for its localisation to chromatin

Following the identification of an N-terminal UBL domain in SMCHD1, I was interested in investigating its functional role in the context of the full-length protein. Using the same cellular assay described above, I transfected either a full-length wild-type Smchd1 or a full-length Smchd1 construct that lacks the UBL domain (Δ1-110 aa) into SMCHD1-KD HEK293 cells. Transfection of the wild-type mouse Smchd1 construct into SMCHD1-KD cells restored Smchd1's enrichment to nuclear foci, but as previously discussed, this results in a distinct Smchd1 staining pattern compared to endogenous human SMCHD1 due to the overexpression system used (Figure 5.17 c). However, deletion of the UBL domain from Smchd1 led to abrogation of focus formation, resulting in a diffuse Smchd1 staining pattern instead which does not resemble wild-type Smchd1-transfected cells (Figure 5.17 d). These findings suggest that Smchd1's UBL domain is required for its localisation to chromatin.

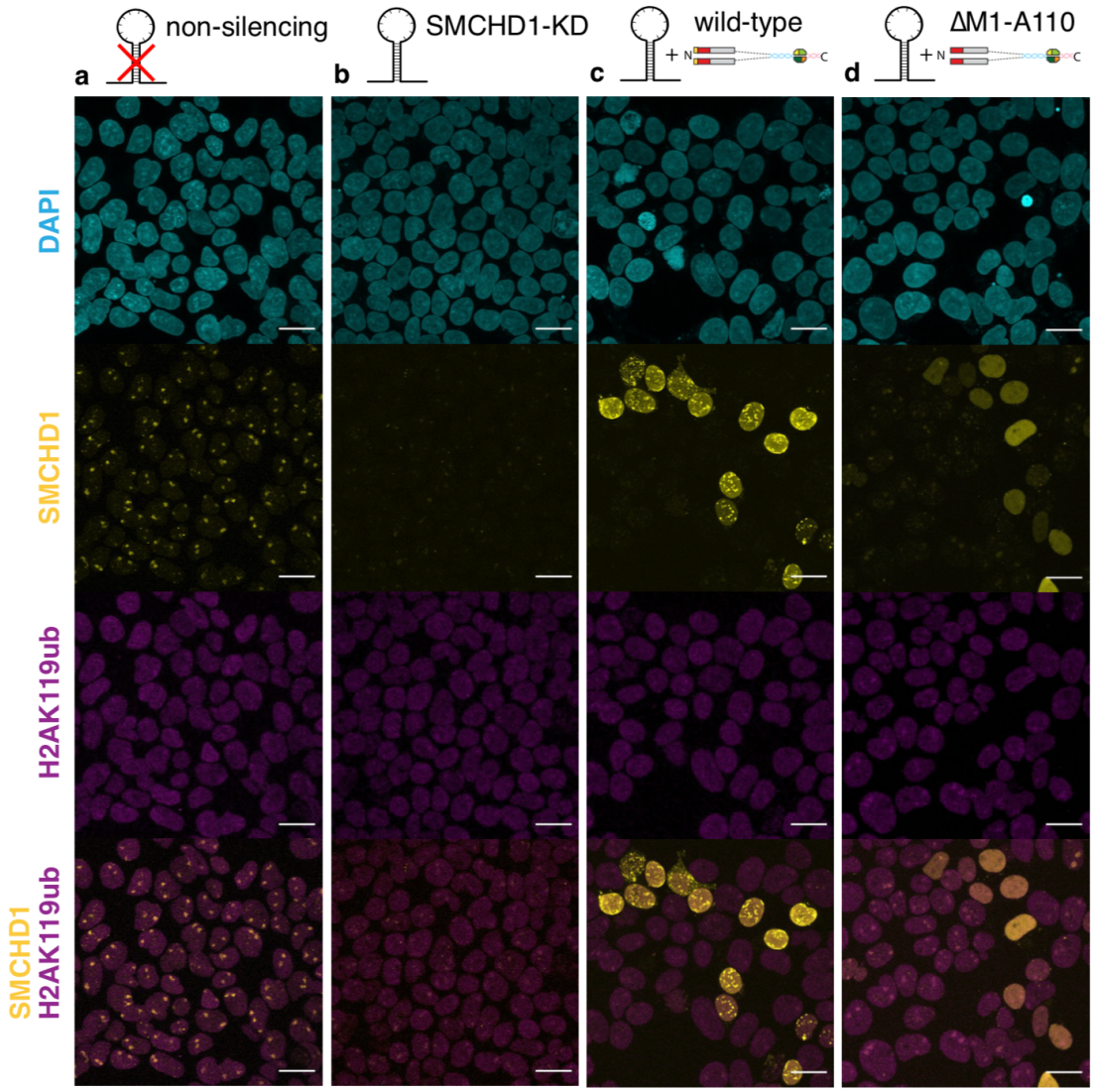


FIGURE 5.17: **Smchd1's UBL domain is required for its localisation to chromatin.** Immunofluorescence in (a) control, non-silencing shRNA transduced, (b) shRNA-mediated SMCHD1-knockdown 293 cells and (c) SMCHD1-knockdown 293 cells 24 hours post-transfection with the full-length Smchd1 or with (d) a Smchd1 construct that lacks the UBL domain (Δ1-110 aa). DAPI staining is depicted in blue, Smchd1 staining is shown in yellow and H2AK119ub is shown in purple, with merged images shown below as indicated. Maximum intensity projection images are shown as representative of n>100 nuclei positive for Smchd1 overexpression per sample; data shown are representative of 2 independent experiments. All images were obtained with identical settings between controls and all transfected cells, to enable comparison between the images provided in the figure. Scale bars, 20 μm.

5.4 Discussion

Upon trialing various protein crystallisation strategies of the SMCHD1 ATPase 111-702 aa construct, no protein crystals were obtained. Following the release of the first structure of a human SMCHD1 ATPase region and the identification of a UBL domain that was proposed to be required for its dimerisation, potential explanations for unsuccessful crystallisation trials became more evident. The 111-702 aa SMCHD1 construct I was investigating lacked the UBL domain (25-110 aa) and was always present as a monomeric species in solution. A dimerisation event would likely reduce the dynamics within SMCHD1's ATPase, providing a greater overall stability and propensity for crystal formation. However, the published structure was obtained using the catalytically inactive SMCHD1 mutant, E147A, in the presence of ATP. Therefore, it may be inferred that the E147A mutant retains the ability to bind ATP but fails to hydrolyze it. Interestingly, dimerisation triggered by ATP-binding is a common feature present among GHKL ATPases [115, 119]. This phenomenon may justify the observed preferential dimerisation of the E147A mutant over wild-type SMCHD1 depicted by Native-PAGE studies [133], as it is likely a result of this mutant being trapped in an ATP-bound state.

Despite the high-resolution structure of SMCHD1's ATPase being unavailable at the time, using limited proteolysis followed by HDX-MS analysis, I identified a loop region within the protein that corresponds to the conserved ATP-lid structure present across members of the GHKL family of proteins [99, 101, 115, 119]. To investigate its functional role, I removed this region from the SMCHD1 ATPase construct to produce a stably expressed protein. Subsequent *in vitro* ATPase analysis revealed a reduced catalytic activity of the loop-deleted SMCHD1 ATPase, by approximately three-fold compared to the wild-type counterpart. Following the availability of the published SMCHD1 ATPase structure, it became evident that a residue located within SMCHD1's ATP-lid, Arg207, directly interacts with the ATP nucleotide and helps its positioning for hydrolysis [133]. Hence, upon deletion of ATP-lid residues 200-229, I thereby disrupted the stabilisation of ATP, consequently reducing its catalytic efficiency.

Across the different SMCHD1 ATPase constructs studied, comparable catalytic activities were obtained via *in vitro* ATPase assays, revealing that SMCHD1 catalytic rates were unaffected by either the presence of the N-terminal UBL domain or a

C-terminal extension past the transducer domain. These results suggest that the dimerisation event of the SMCHD1 ATPase region does not contribute to an altered ATP turnover, furthermore inferring the absence of a synergistic behaviour upon dimerisation of the ATPase region. This contrasts with previous findings obtained for the GHKL ATPase, Hsp90, where monomeric constructs exhibited a reduced ATPase activity compared to the wild-type dimeric counterpart [220].

The identification of an N-terminal UBL domain in SMCHD1 raised a great interest regarding its functional role. UBL domains have been commonly shown to interact with the 26S proteasome, stimulating proteasomal activity in a similar fashion to ubiquitin chains [221, 222]. In the E3 ubiquitin ligase, Parkin, its UBL domain inhibits Parkin's auto-ubiquitination as it competes with ubiquitin for the binding site [212]. Parkin's UBL domain was additionally suggested to regulate the activity of its RING1 domain, as the two directly interact [223]. In the case of SMCHD1, the UBL domain may play a role in targeting SMCHD1 to H2AK119ub-marked chromatin, as removal of H2AK119ub chromatin marks were shown to result in a global loss of Smchd1 protein stability and genome-wide changes in gene silencing [53, 55]. Such interactions may occur either directly, via SMCHD1's UBL domain, or via specific protein interactions that have not yet been determined, to therefore regulate SMCHD1's residence time on chromatin.

I have shown via AUC studies that the UBL domain is required for the dimerization of the wild-type SMCHD1 ATPase region, which is likely driven by the domain-swapping event of the UBL regions [133]. It is therefore possible the UBL domains may likewise undergo domain-swapping with protomers of adjacent SMCHD1 dimers, resulting in a higher-order assembly. Such a phenomenon may account for the absence of nuclear foci formation upon the transfection of a UBL-less full-length Smchd1 construct into SMCHD1-knockdown cells, failing to oligomerise and to arrange into regions of high local protein concentrations. Immediate future studies will aim to further investigate the dimerisation properties of disease-associated mutants affecting the ATPase region of SMCHD1 via AUC studies, which may reveal differences that might account for an altered molecular function of SMCHD1 in patients suffering from either FSHD or BAMS. I will also perform additional AUC experiments to examine the dimerisation of ligand-bound wild-type SMCHD1 ATPase (25-702 aa) at varying protein concentrations to quantify its equilibrium dissociation constant (K_d) for self-association into dimers. Obtaining a crystal structure of the wild-type 25-702 aa

SMCHD1 ATPase, as opposed to the 25-580 E147A SMCHD1 that was recently published, may also reveal additional contacts at the extended C-terminal end of the region that may contribute to the dimerisation interface. The neomorphic *Smchd1* mutation, A667E, that revealed an altered cellular localisation in cells, is also located within this C-terminal extension of the SMCHD1 ATPase. SAXS analysis of this mutant in the context of 111-702 aa Smchd1 additionally exposed gross conformational changes in the protein compared to the wild-type counterpart (manuscript in preparation). Hence, unveiling its exact location and contact sites within the protein may elucidate how the substitution mutation might alter both SMCHD1's activity and its three-dimensional arrangement.

Using immunofluorescence studies, I additionally revealed that mutations that display either a gain or loss of ATPase activity of Smchd1 *in vitro* exhibit an altered cellular localisation in cells. These results suggest that Smchd1's localisation to chromatin is dependent on its catalytic ability, and that a specific "sweet-spot" of ATPase activity is required for Smchd1's optimal function. This interpretation is further supported by the altered *in vitro* catalytic rates observed across FSHD- and BAMS-associated Smchd1 variants. We can therefore further utilise this cellular-based assay as a supporting method to assess the effects of disease-associated mutations upon Smchd1 function, which will aid in providing a better understanding of its molecular mechanism. We additionally plan to develop a stable cellular transfection system with the use of fluorescently-labelled SMCHD1 to enable live imaging studies and the investigation of SMCHD1's dynamic behaviour.

GHKL ATPase-containing proteins, such as Hsp90 or the MORC family of proteins, have commonly been described as 'molecular clamps', where N-terminal dimerisation dictates the opening and closing of the ATPase dimer and is directly coupled to the ATPase cycle [99, 109, 115]. MORC proteins were recently proposed to capture chromatin via a clamping or compaction mechanism, resulting in chromatin looping [123]. In contrast, SMC proteins comprise an ATP-binding cassette (ABC)-type ATPase domain, which have most commonly been recognised across diverse families of transporter proteins where they are involved in translocating solutes across a biological membrane [224]. However, they are also implicated in chromosome segregation and DNA repair, as described for canonical SMC proteins, cohesin and condensin [225]. Both were recently demonstrated to also induce chromatin looping, via an

ATP-dependent loop extrusion mechanism [80, 87]. Considering that SMCHD1 contains both a canonical SMC hinge domain as well as a GHKL-type ATPase domain, its molecular mechanism of function may incorporate functional features from both the SMC and GHKL protein families. If SMCHD1 acts as a molecular clamp, the lifetimes of its open and closed states, and therefore its residence time on chromatin, would likely be determined by the dimerisation dynamics of its ATPase region. Further examination of the full-length Smchd1 protein, such as via single-molecule imaging studies as recently performed for cohesin and condensin [80, 87, 88], may best reveal how Smchd1 interacts with DNA or chromatin, the ATP-driven molecular functions it possesses, and its overall mobility mechanism along DNA.

Chapter 6

Discussion

There has been a growing interest, and consequently, many recent advances made in our current understanding of Smchd1's molecular structure and function. Contributing to this field, I conducted preliminary structural studies investigating the full-length Smchd1 protein which revealed low-resolution images of conformationally flexible, rod-like particles. Using immunofluorescence microscopy, I demonstrated the requirement of Smchd1's GHKL ATPase activity, the N-terminal UBL domain, and of the C-terminal hinge region, in Smchd1's localisation to chromatin and nuclear foci formation. I have additionally provided an insight into Smchd1's mode of interaction with nucleic acids and the role of the coiled-coil regions that border Smchd1's hinge domain in augmenting its homodimerisation. Furthermore, I was able to demonstrate that the wild-type Smchd1 ATPase region undergoes dimerisation in the presence of the newly-identified UBL domain and the GHKL ATPase subtrate, ATP.

Based on experimental evidence presented in this thesis and on related published studies, I propose a model for how Smchd1 may be recruited to chromatin to result in gene expression silencing. I hypothesize that Smchd1's functional cycle is initiated upon the ATP-induced dimerisation of the N-terminal region, a process that results in a UBL-domain swapping event and a compact arrangement of the dimeric ATPase (Figure 6.1 a). Following ATP hydrolysis, I propose the dimeric ATPase reverts back to a monomeric form, resulting in the transition of full-length Smchd1 from a rod-like conformation to a V-like structure that remains anchored at the C-terminally located hinge domain (Figure 6.1 a). Smchd1's hinge domain has been established to hold a role in nucleic acid interaction, with *in vitro* studies suggesting it has a mild preference for single-stranded DNA as opposed to double-stranded DNA, and inappreciable sequence specificity [58]. It is therefore likely that nucleic acid interactions via the hinge domain do not confer selectivity, but instead increase Smchd1's residence time on DNA to facilitate further specific interactions, such as with H2AK119ub-marked chromatin which holds the highest degree of cellular co-localisation with Smchd1 reported to date [55].

Smchd1's hinge domain may undergo conformational changes, potentially triggered by its initial interaction with DNA (Figure 6.1 a). This was recently demonstrated for human cohesin, which was captured in crystal form as an "open" SMC1-SMC3 heterodimer at either the south- or the north-facing end. For cohesin, this mechanism may denote an initial interaction with DNA at the north-end of the hinge domain,

followed by its threading through to the south-end. This event may then allow DNA to permeate the adjacent coiled-coil region of the SMC1-SMC3 heterodimer that separates the hinge domain from the ABC-type ATPases, which adopts a ring-shaped structure where DNA is thought to become entrapped upon loop extrusion. A similar process may therefore underlie Smchd1's mechanism of DNA entrapment (Figure 6.1 a). Furthermore, a conformational change within Smchd1's hinge domain that would result in buried residues becoming exposed would account for the demonstrated requirement of the two buried hinge domain residues, R1848 and R1867, in Smchd1's localisation to chromatin in cells (Chapter 4). To best reveal the precise DNA interaction sites within Smchd1's hinge domain, and potentially expose whether the hinge dimer may obtain an open conformation, obtaining a crystal structure of a nucleic-acid bound Smchd1 hinge would be most favourable.

Whether Smchd1's interaction with DNA results in chromatin loop formation via an ATP-dependent loop extrusion mechanism, as described for SMC proteins cohesin [87, 88] and condensin [80, 81], or via chromatin compaction, which involves entrapping DNA loops via a process that is not driven by ATP and previously demonstrated for the MORC1 GHKL protein [123], remains to be investigated. Notably, the intermediate region of Smchd1 that connects its two functional domains is predicted as predominantly β-stranded based on sequence prediction softwares, potentially indicating a different functional role to canonical SMC proteins, whose central regions comprise of long α-helical coiled-coils. However, short α-helical regions that are likewise predicted to arrange into coiled-coils immediately flank the hinge domain of Smchd1 on either side. Remarkably, similar short coiled-coil regions are present across all members of the MORC family of proteins, where their role has not been clearly determined but are thought to be involved in their dimerisation, protein-DNA interactions, and gene transcription [226]. The described similarity in gene architecture, alongside the presence of a GHKL-type ATPase that is shared between Smchd1 and MORC proteins, may indicate a closer functional resemblance of Smchd1 to the MORC rather than the SMC family of proteins. Nonetheless, further functional assays are required to examine this hypothesis. The removal of bound DNA from purified full-length Smchd1 protein would allow single-molecule imaging studies, enabling the investigation of Smchd1's behaviour on DNA or chromatin, with the added ability of incorporating associated nucleosomes in the experimental setup.

Based on my findings outlined in Chapter 3, the co-purification of Smchd1 with DNA and histone proteins may indicate a requirement of nucleosomal chromatin for Smchd1's protein stability, a hypothesis that prompts further investigation.

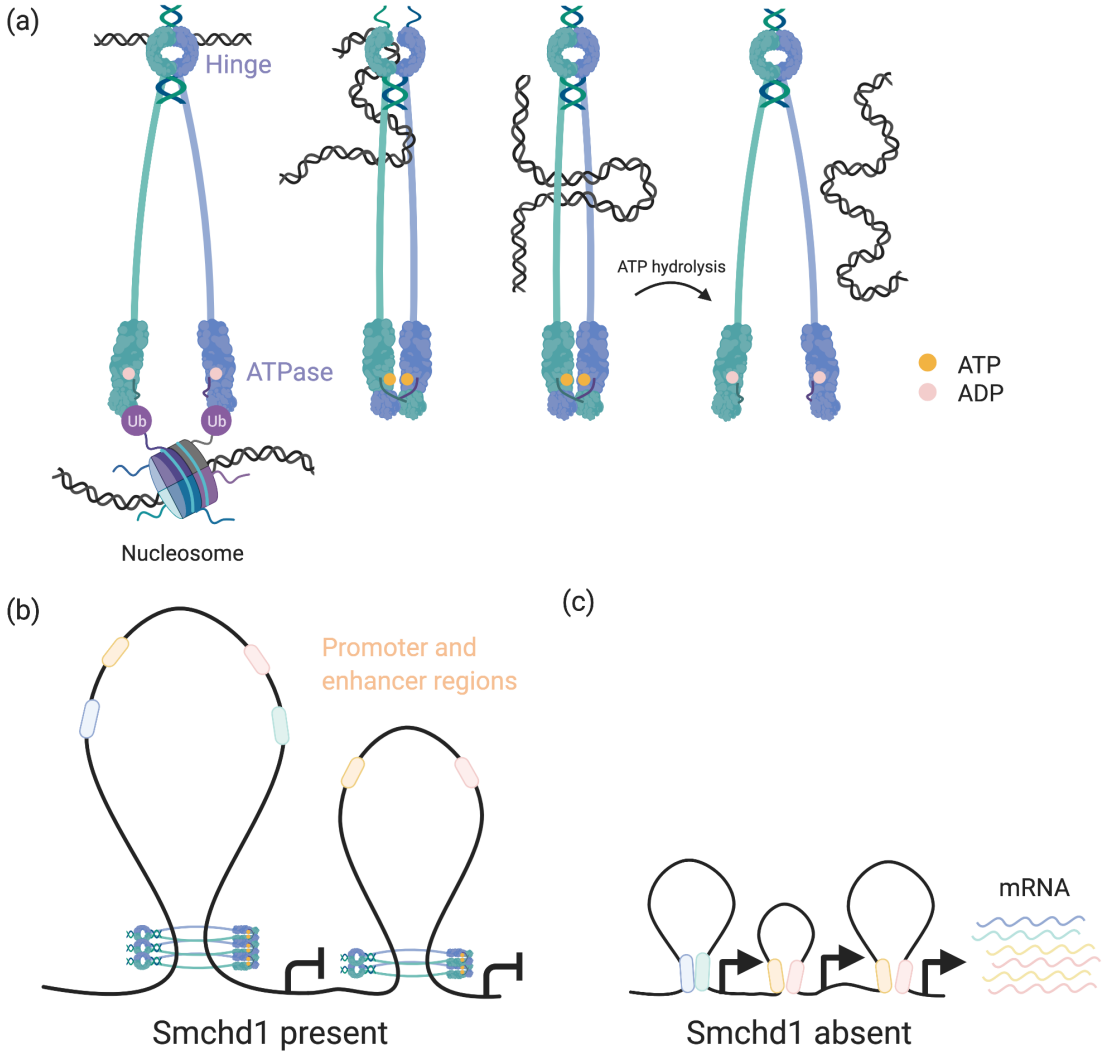


FIGURE 6.1: **A model of Smchd1's molecular mechanism and recruitment to DNA.** (a) A hypothetical model proposing that Smchd1 initially interacts with DNA via its C-terminal hinge domain, and is then specifically recruited to H2AK119ub-marked chromatin via its ATPase region, potentially via the UBL domain. Smchd1's interaction with DNA may trigger a conformational change that results in the opening of the hinge domain dimer to reveal buried residues, such as R1867 and R1848, shown to be critical in Smchd1's localisation to chromatin (Chapter 4). This conformational change could further result in the opening of the central intermediate region of Smchd1, allowing DNA to reel through and lead to the formation of chromatin loops. Upon ATP-binding, the ATPase region of Smchd1 undergoes dimerisation, leading to a UBL domain-swapping event. Following ATP hydrolysis, the Smchd1 protomers disengage at the ATPase region, resulting in the opening of the dimer and the release of DNA. (b) Long-range chromatin loop formation induced by Smchd1 prevents interactions between distal promoter and enhancer regions, leading to a transcriptionally repressive environment. (c) When Smchd1 is absent, short-range chromatin interactions take place which facilitate promoter-enhancer interactions, thereby promoting a transcriptionally permissive environment.

The presence of a UBL domain within SMCHD1's N-terminus has sparked a great interest regarding its functional role. Using AUC studies, I revealed that the UBL domain is required for the dimerization of the wild-type SMCHD1 ATPase region, which is likely driven by the domain-swapping event of the UBL regions [133]. It is therefore possible the UBL domains may likewise undergo domain-swapping with protomers of adjacent SMCHD1 dimers, resulting in a higher-order assembly (Figure 6.2). Such a phenomenon may account for the absence of nuclear foci formation upon the transfection of a UBL-less full-length Smchd1 construct into SMCHD1-knockdown cells, failing to oligomerise and to arrange into regions of high local protein concentrations. Alternatively, Smchd1's inability to localise to nuclear foci in the absence of the UBL domain may potentially indicate a disruption in the formation of phase-separated droplets. Liquid-liquid phase separation may represent a functional property of Smchd1 that could manifest as the discrete nuclear foci observed upon Smchd1-specific immunofluorescence studies, and is a phenomenon that we are furthermore interested in exploring in regard to Smchd1 function.

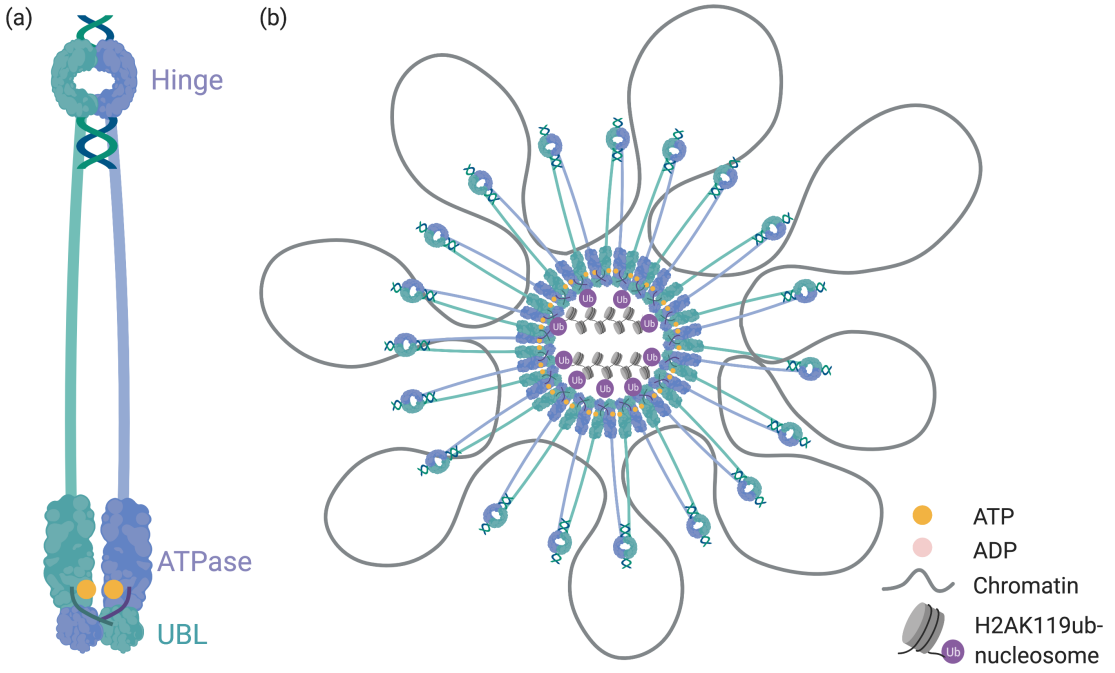


FIGURE 6.2: **A model of Smchd1 oligomerisation.** (a) A cartoon representation of the full-length Smchd1 protein and its component domains. (b) Smchd1 is proposed to be recruited to H2AK119ub-marked chromatin sites, where it may undergo a higher-order arrangement driven by UBL domain-swapping events between protomers of other Smchd1 dimers, resulting in a high localised Smchd1 presence. The associated chromatin loop formation that is proposed to be induced by Smchd1 may thereby create a transcriptionally repressive hub, where distal promoter and enhancer regions are unable to interact.

The discovery that BAMS-associated variants in *SMCHD1* lead to a gain-of-function phenotype alongside an enhanced ATPase activity *in vitro* raises the possibility that SMCHD1's function could be augmented [46, 146], an idea that may be exploited for the treatment of FSHD2 where patients suffer from SMCHD1 haploinsufficiency. Small-molecule activators would target the functional SMCHD1 allele that is not affected by heterozygous mutations, whereby boosting its function may represent a way to compensate for the dysfunctional allele. We are additionally interested in developing SMCHD1 inhibitors as a potential therapeutic treatment for Prader-Willi Syndrome (PWS). SMCHD1 normally silences several genes within the PWS cluster [44, 45, 52], whose reactivation would provide one active allele in these patients, and may be achieved via the removal of SMCHD1 function. Importantly, the targetable PWS cluster is predominantly expressed in the hypothalamus, enabling a localised treatment option. The recently reported crystallisation of the ATPase region of SMCHD1 may additionally enable its co-crystallisation with small molecule compounds emerging from the high-throughput screens we are conducting. The resulting information would be incredibly informative, facilitating compound optimisation to improve target specificity and selectivity, particularly regarding other GHKL-type ATPases that share a similar ATP-binding pocket to SMCHD1 and represent undesirable off-targets.

The work that I have presented here has contributed to our current understanding of how Smchd1 may interact with chromatin to facilitate gene expression regulation. Ongoing efforts are directed at visualising the high-resolution structure of full-length Smchd1 and the role of ATP hydrolysis in manipulating chromatin structure to result in gene silencing events. New approaches in identifying novel Smchd1 interaction partners may additionally help provide a further insight into its function and recruitment to chromatin, and highlight its potential mode of association with nucleosomes. Such findings may in turn facilitate the structural stabilisation of the full-length Smchd1 protein, reducing its observed conformational flexibility to allow high-resolution structural studies, such as cryo-EM. Providing a comprehensive understanding of Smchd1's distinct functional properties furthermore highlights the requirement of different experimental approaches that are necessary to study the enigmatic protein that Smchd1 is proving to be. Nonetheless, the resulting information will be critical to our understanding of SMCHD1's involvement in human disease and, consequently, pave the way to novel therapeutic treatments.

Appendix A

Immunoprecipitation and Mass Spectrometry identification of Smchd1-interacting proteins

ProteinID	LeadingRazor Protein	PValue_WRST GFP/WT	Gene names	Protein names	Leading proteins	GFP Imputed	WT Imputed	GFP NotImputed	WT NotImputed	Significant WRST GFP/WT
37	Q6P5D8	3.80E-137	Smchd1	Structural maintenance of chromosomes flexibl	Q6P5D8	31.648737	25.40034	31.6440989	0	++
150	E9Q616	2.14E-123	Ahnak		E9Q616	30.547555	28.573096	30.5326331	28.416301	++
67	Q9QXS1-2	9.86E-106	Plec	Plectin	Q9QXS1-2	32.766646	31.042126	32.7603967	30.994858	++
262	Q8BTM8	1.47E-46	Flna	Filamin-A	Q8BTM8	31.004864	29.903828	30.9971742	29.863226	+
126	P70227	5.51E-46	Itpr3	Inositol 1,4,5-trisphosphate receptor type 3	P70227	27.76288	25.657484	27.7248179	25.109982	++
105	Q5SWZ5	6.81E-37	Mprip		Q5SWZ5	28.628651	26.834028	28.6093963	26.654191	++
157	Q5SXR6	6.47E-36	Cltc;mKIAA00	Clathrin heavy chain;Clathrin heavy chain 1	Q5SXR6	27.880838	26.352931	27.8434796	26.087596	++
118	Q9EP71	3.67E-32	Rai14	Ankycorbin	Q9EP71	28.273037	26.353009	28.2574805	26.180372	++
53	A2AVJ7	6.04E-32	Rrbp1	Ribosome-binding protein 1	A2AVJ7	29.505441	26.803226	29.49677	26.625666	++
634	Q9DBR7	2.58E-31	Ppp1r12a	Protein phosphatase 1 regulatory subunit 12A	Q9DBR7	28.530148	26.444631	28.5142509	26.262859	++
65	A0A0R4J0J8	2.96E-31	Specc1l	Cytospin-A	A0A0R4J0J8	27.931397	26.054063	27.9130204	25.823832	++
154	Q6P5H2	9.20E-31	Nes	Nestin	Q6P5H2	28.967713	26.358548	28.9501056	26.066147	++
279	Q62261	5.02E-30	Sptbn1	Spectrin beta chain, non-erythrocytic 1	Q62261	28.988904	27.677909	28.9625942	27.513532	+
219	P16546-2	9.32E-29	Sptan1	Spectrin alpha chain, non-erythrocytic 1	P16546-2	29.223634	28.044703	29.1887279	27.882401	+
270	Q8VHX6	5.67E-28	Flnc	Filamin-C	Q8VHX6	27.613943	26.127086	27.5741317	25.825612	+
59	Q7TQD7	5.79E-27	Myo1b	Unconventional myosin-Ib	Q7TQD7	28.675046	26.30075	28.6621576	26.110101	++
956	Q9JHU4	1.60E-26	Dync1h1	Cytoplasmic dynein 1 heavy chain 1	Q9JHU4	25.343799	23.72471	25.1891666	21.569673	+
217	E9Q3Z5	2.52E-26	Svil	Supervillin	E9Q3Z5	27.351597	26.015138	27.3121532	25.738555	++
151	B2RRE0	3.15E-26	Akap12	A-kinase anchor protein 12	B2RRE0	26.005447	24.024592	25.9469697	23.116454	+
317	Q545R0	3.98E-26	Ctnna1;Ctnna	Catenin alpha-1;Catenin alpha-3	Q545R0	26.76998	24.711952	26.7410171	24.259307	++
238	S4R1P5	3.16E-25	Dst	Dystonin	S4R1P5	25.95399	24.36565	25.812388	22.990663	+
135	Q7TPR4	5.83E-25	Actn1	Alpha-actinin-1	Q7TPR4	29.675451	28.243667	29.6672659	28.191067	+
54	Q80X90	1.94E-24	Flnb	Filamin-B	Q80X90	30.026794	29.24771	30.008619	29.193034	
661	Q5FWJ3	2.42E-24	Vim	Vimentin	Q5FWJ3	33.830269	32.631899	33.8295035	32.627783	++
131	B9EK95	4.86E-24	Myof	Myoferlin	B9EK95	26.407185	24.248075	26.3403223	23.190656	+
276	A0A0J9YUR2	2.70E-21	Specc1;Ccdc9	Cytospin-B;Coiled-coil domain-containing protei	A0A0J9YUR2	26.270832	23.65483	26.2427048	22.730353	++
106	Q99104	1.97E-20	Myo5a	Unconventional myosin-Va	Q99104	29.017819	28.006735	28.995672	27.916217	
107	A0A0R4J1H6	6.93E-20	Golga3	Golgin subfamily A member 3	A0A0R4J1H6	26.282512	23.905028	26.2146436	22.442663	++
659	Q3UBA6	2.48E-19	Hspa8	Heat shock cognate 71 kDa protein	Q3UBA6	28.876345	27.207101	28.8653344	27.136651	+
564	P38647	1.25E-18	Hspa9	Stress-70 protein, mitochondrial	P38647	27.720442	25.940674	27.7045893	25.784351	++
400	A0A0R4J1E3	3.63E-18			A0A0R4J1E3	28.974349	27.23741	28.9632145	27.146714	++
28	F8VPU2	3.80E-18	Farp1	FERM, RhoGEF and pleckstrin domain-containin	F8VPU2	26.590791	24.378847	26.5649084	23.986872	++
13	E9Q8Z5	9.58E-18	Ctnnd1	Catenin delta-1	E9Q8Z5	25.901508	23.886023	25.8567288	23.185036	++
459	Q3U9G2	1.73E-17	Hspa5	78 kDa glucose-regulated protein	Q3U9G2	29.468647	27.615294	29.4624967	27.55884	+
1036	P51655	1.77E-17	Gpc4	Glypican-4;Secreted glypican-4	P51655	27.601599	24.784235	27.581451	24.391949	++
243	A6MDD3	3.40E-17	Cd109	CD109 antigen	A6MDD3	26.093112	23.410487	26.061309	22.250912	++
182	Q8K1N2	4.81E-16	Phldb2	Pleckstrin homology-like domain family B mem	Q8K1N2	26.659538	25.301301	26.6275399	25.039518	+
472	E9PXX7	2.03E-15	Txndc5	Thioredoxin domain-containing protein 5	E9PXX7	26.585299	23.364107	26.570201	22.706757	++
261	F2Z4A3	6.14E-15	Fat1		F2Z4A3	26.34703	24.966698	26.2502148	24.419444	+
862	A0A0A6YY47	6.78E-15	Ncam1	Neural cell adhesion molecule 1	A0A0A6YY47	25.139168	22.411885	25.0953016	20.785791	++
204	P62204	1.18E-14	Calm1;Calml3	Calmodulin;Calmodulin-like protein 3;Troponin C	P62204	29.511213	27.674698	29.509544	27.641297	++
195	Q8BMS1	1.45E-14	Hadha	Trifunctional enzyme subunit alpha, mitochondr	Q8BMS1	26.399547	23.784514	26.3667797	23.059981	++
38	K3W4L0	2.40E-14	Myo18a	Unconventional myosin-XVIIIa	K3W4L0	26.5494	25.385055	26.4828105	25.040029	+
167	P62141	2.58E-14	Ppp1cb	Serine/threonine-protein phosphatase PP1-beta	P62141	27.384918	25.597204	27.3707385	25.46332	++
503	Q3UDC8	7.30E-14	Eef2	Elongation factor 2	Q3UDC8	27.374896	25.876111	27.3452275	25.699628	+
328	Q01853	8.64E-14	Vcp	Transitional endoplasmic reticulum ATPase	Q01853	26.909366	25.235029	26.8761002	24.911363	+
471	Q3UDR2	9.56E-14	P4hb	Protein disulfide-isomerase	Q3UDR2	27.905883	25.518273	27.8955886	25.338409	++
295	Q3U2W2	2.31E-13	Mybbp1a	Myb-binding protein 1A	Q3U2W2	25.315912	23.46445	25.2717866	22.707596	+
88	E9Q634	3.15E-13	Myo1e;Myo1	Unconventional myosin-Ie;Unconventional myo	E9Q634	27.648505	26.641564	27.6266628	26.523345	+
138	Q6PEE6	3.81E-13	Ap2a2	AP-2 complex subunit alpha-2	Q6PEE6	24.656348	22.693663	24.6016685	21.147441	+
756	Q3TQL4	5.63E-13	Immt	MICOS complex subunit Mic60	Q3TQL4	25.898684	23.837741	25.8665604	23.364065	++
115	Q3UZT7	9.01E-13	Ctnnb1	Catenin beta-1	Q3UZT7	25.263135	22.365828	25.2269719	20.55717	++
514	B2RQQ7	1.15E-12	Cdc42bpb	Non-specific serine/threonine protein kinase;Se	B2RQQ7	24.4312	22.522752	24.2879844	20.034412	+
127	Q53WR6	1.28E-12	Glg1	Golgi apparatus protein 1	Q53WR6	26.480743	22.822903	26.4486723	21.238802	++
398	Q3TX57	1.40E-12	Col1a2	Collagen alpha-2(I) chain	Q3TX57	29.115451	26.845963	29.1011932	26.724139	+
518	Q5SWR1	2.07E-12	Ap2b1	AP complex subunit beta;AP-2 complex subunit	Q5SWR1	25.262637	23.49566	25.1756456	22.909596	+
325	Q564F4	3.28E-12	Cct4	T-complex protein 1 subunit delta	Q564F4	25.646014	23.847085	25.5882594	23.179603	+
706	Q922B9	4.68E-12	Ssfa2	Sperm-specific antigen 2 homolog	Q922B9	24.84545	22.602612	24.7630181	20.772032	++
409	A2AKH4	1.13E-11	Snap23;Snap2	Synaptosomal-associated protein;Synaptosoma	A2AKH4	25.735152	22.641934	25.7022587	21.324576	++
228	E9Q6R7	1.20E-11	Utrn;utrophin		E9Q6R7	25.890469	24.796087	25.7892681	24.236685	
258	Q04750	1.56E-11	Top1;TOP	DNA topoisomerase 1	Q04750	25.162914	22.603888	25.1341787	21.387942	++
406	P56480	2.08E-11	Atp5b	ATP synthase subunit beta, mitochondrial	P56480	26.984098	25.060496	26.9619337	24.839781	+
844	Q811N6	2.16E-11	env	MLV-related proviral Env polyprotein;Surface p	Q811N6	24.258715	21.403082	24.2163853	19.028855	+
929	Q8BMK4	2.50E-11	Ckap4	Cytoskeleton-associated protein 4	Q8BMK4	26.367982	24.9545	26.3497062	24.778389	+
979	Q58E64	3.52E-11	Eef1a1;Eef1a	Elongation factor 1-alpha;Elongation factor 1-al	Q58E64	28.195121	26.898004	28.1780336	26.809256	+
244	A7YY80	4.20E-11	Epb4.1l3		A7YY80	25.264728	23.477043	25.2053473	22.774883	+
152	Q3UZI3	4.37E-11	Snd1	Staphylococcal nuclease domain-containing pro	Q3UZI3	25.179755	23.764161	25.1036418	23.170209	+
558	P08752	5.14E-11	Gnai2;Gnai1	Guanine nucleotide-binding protein G(i) subunit	P08752	27.626302	25.718515	27.6123823	25.573843	+
717	Q80T06	6.11E-11	Eef1d	Elongation factor 1-delta	Q80T06	26.472213	23.445321	26.4543479	22.927578	++
592	Q80Y52	1.07E-10	Hsp90aa1	Heat shock protein HSP 90-alpha	Q80Y52	25.909502	23.998456	25.858044	23.524924	++
917	Q3TFA9	1.10E-10	Tmod3	Tropomodulin-3	Q3TFA9	28.689239	27.18949	28.6826859	27.147251	+
9	Q80TM2	1.15E-10	Tln1	Talin-1	Q80TM2	25.220083	24.29354	25.076349	23.635582	
101	Q811D0-2	1.25E-10	Dlg1	Disks large homolog 1	Q811D0-2	24.013475	21.849051	23.9132469	20.114042	+
136	Q3ULT2	1.28E-10	Actn4	Alpha-actinin-4	Q3ULT2	27.991075	26.920503	27.971673	26.82911	+
480	Q8BN32	1.33E-10	Pabpc1;Pabpc	Polyadenylate-binding protein;Polyadenylate-b	Q8BN32	27.026594	25.131152	26.9939519	24.864557	+
393	B2RSU6	1.83E-10	Cgnl1	Cingulin-like protein 1	B2RSU6	23.653521	21.793112	23.5112618	17.724766	+
69	Q3TFD0	1.97E-10	Shmt2	Serine hydroxymethyltransferase	Q3TFD0	24.603614	22.447614	24.5320594	21.187051	+
405	Q3UL22	2.06E-10	Cct8;Cctq	T-complex protein 1 subunit theta	Q3UL22	25.842092	23.634276	25.7973395	22.852347	++
269	Q99K58	2.89E-10	Fbln2	Fibulin-2	Q99K58	25.119257	23.364478	25.0571573	22.714237	+
869	Q9DC36	3.06E-10	Flot2	Flotillin-2	Q9DC36	25.341992	23.400385	25.2970794	22.929832	+
668	Q9WTI7-3	3.13E-10	Myo1c	Unconventional myosin-Ic	Q9WTI7-3	29.860381	29.056772	29.8493296	29.0263	
17	E9QAH1	4.16E-10	Golgb1		E9QAH1	23.779689	22.390828	23.5539181	0	
408	A0A087WQQ5	4.35E-10	Ktn1;mKIAA0	Kinectin	A0A087WQQ5	23.982943	22.530256	23.8994317	21.727738	+
306	Q99L75	4.78E-10	Hspa4	Heat shock 70 kDa protein 4	Q99L75	24.675699	23.059981	24.610707	22.407092	+
247	Q3TN84	4.85E-10	P4ha1	Prolyl 4-hydroxylase subunit alpha-1	Q3TN84	27.08365	24.475923	27.0703972	24.23663	++
538	B1AWB9	4.90E-10	Col5a1	Collagen alpha-1(V) chain	B1AWB9	26.735816	24.798366	26.6817696	24.36675	+

802	Q8CGC7	4.99E-10	Eprs	Bifunctional glutamate/proline--tRNA ligase;Glu	Q8CGC7	24.653152	22.96303	24.5559606	21.870697	+
950	B9EKE9	5.15E-10	Ddx3x;D1Pas1	ATP-dependent RNA helicase DDX3X;Putative A	B9EKE9	25.790329	24.304389	25.7527941	23.993097	+
166	B9EHJ3	9.87E-10	Tjp1	Tight junction protein ZO-1	B9EHJ3	25.576924	24.296959	25.5245832	23.968869	+
590	Q5RKP4	1.05E-09	Rpn1	Dolichyl-diphosphooligosaccharide--protein glyc	Q5RKP4	24.895109	22.976527	24.8456755	22.390449	+
1853	Q8BTI8-3	1.14E-09	Srrm2	Serine/arginine repetitive matrix protein 2	Q8BTI8-3	22.523042	20.500983	22.4580617	18.424381	+
1244	Q58EU6	1.32E-09	Rpl5	60S ribosomal protein L5	Q58EU6	24.887637	22.675992	24.8262264	21.749433	++
320	E9PX70	1.51E-09	Col12a1	Collagen alpha-1(XII) chain	E9PX70	23.945705	25.267851	23.2370715	25.095295	
614	Q540I4	1.51E-09	Flot1	Flotillin-1	Q540I4	24.90321	22.615852	24.8656312	21.681147	+
45	E9QAI5	1.71E-09	Cad	CAD protein;Glutamine-dependent carbamoyl-p	E9QAI5	23.165019	21.558051	22.9627997	18.817204	+
992	G3XA10	1.74E-09	Gm28062;Hnr	Heterogeneous nuclear ribonucleoprotein U	G3XA10	24.952485	23.259697	24.9340439	22.92514	+
178	Q5U438	2.20E-09	Npm1;Gm561	Nucleophosmin	Q5U438	26.663327	25.036208	26.6433291	24.901532	++
685	Q544Y7	2.35E-09	Cfl1	Cofilin-1	Q544Y7	27.90809	25.921966	27.9000429	25.83	++
845	Q8CBM0	2.41E-09	Zyx	Zyxin	Q8CBM0	24.468816	22.448342	24.4249247	21.283388	+
110	P63038	2.99E-09	Hspd1	60 kDa heat shock protein, mitochondrial	P63038	26.579964	25.016935	26.5528019	24.797755	+
542	P63094	3.23E-09	Gnas;Gnal	Guanine nucleotide-binding protein G(s) subunit	P63094	26.531716	23.724323	26.5205423	23.407537	++
16	Q4FZK2	3.28E-09	Eef1g	Elongation factor 1-gamma	Q4FZK2	25.762644	23.020608	25.7360731	22.346506	++
934	Q9CY58	3.70E-09	Serbp1	Plasminogen activator inhibitor 1 RNA-binding p	Q9CY58	25.985928	23.739281	25.9622979	23.385338	++
570	P29268	3.82E-09	Ctgf;fisp-12	Connective tissue growth factor	P29268	27.441194	25.118266	27.4314364	24.991559	+
975	Q8K173	3.94E-09	Col3a1	Collagen alpha-1(III) chain	Q8K173	26.053121	23.113935	25.9893991	21.911227	+
296	Q3ULF7	4.42E-09	Actr3;Actr3b	Actin-related protein 3;Actin-related protein 3B	Q3ULF7	27.289283	25.792742	27.2775378	25.697686	+
414	P11881-8	4.55E-09	Itpr1	Inositol 1,4,5-trisphosphate receptor type 1	P11881-8	23.70326	21.827086	23.5747724	18.310602	
1157	Q3UGJ5	4.55E-09	Rasa3	Ras GTPase-activating protein 3	Q3UGJ5	23.389221	21.459071	23.2837328	17.517761	+
1139	Q3UHH5	7.83E-09	Gnaq	Guanine nucleotide-binding protein G(q) subunit	Q3UHH5	24.411425	21.057716	24.3724499	0	+
1016	Q80UE5	8.14E-09	Epb4.1l2		Q80UE5	24.970264	23.399074	24.8929654	22.785499	+
111	Q3TFE8	8.42E-09	Kpnb1	Importin subunit beta-1	Q3TFE8	24.972552	23.599494	24.8611944	22.930052	
1165	B7ZP22	9.80E-09	Hnrnpa2b1	Heterogeneous nuclear ribonucleoproteins A2/B	B7ZP22	26.510985	23.903279	26.4912016	23.681067	+
63	O70400	1.09E-08	Pdlim1	PDZ and LIM domain protein 1	O70400	25.257445	22.544381	25.2025939	21.22802	+
639	Q8VCQ8	1.25E-08	Cald1		Q8VCQ8	30.530142	29.774396	30.5243014	29.759313	
649	Q6PAC1	1.30E-08	Gsn	Gelsolin	Q6PAC1	28.95799	28.334113	28.9502384	28.305983	
1460	D3YYT0	1.33E-08	Cdh2;Cdh4	Cadherin-2;Cadherin-4	D3YYT0	24.309232	21.83339	24.2613131	19.282504	+
10	Q8C845	1.45E-08	Efhd2	EF-hand domain-containing protein D2	Q8C845	26.2593	24.468009	26.246901	24.346429	++
425	Q5PPQ7	1.53E-08	Coro1c;Coro6	Coronin;Coronin-1C;Coronin-6	Q5PPQ7	28.637249	27.481038	28.6308462	27.445842	+
66	E9PWE8	1.60E-08	Dpysl3	Dihydropyrimidinase-related protein 3	E9PWE8	24.351502	22.182267	24.3007132	21.094078	+
233	E9QLA5	1.65E-08	Inf2	Inverted formin-2	E9QLA5	25.018935	23.876437	24.9248347	23.34885	+
404	Q3TE06	1.84E-08	Wdr1	WD repeat-containing protein 1	Q3TE06	26.089038	24.90605	26.0615028	24.710427	+
1166	Q3U7F3	2.30E-08	Hnrnpa1;Hnrn	Heterogeneous nuclear ribonucleoprotein A1;He	Q3U7F3	26.580821	22.923727	26.5695998	22.29675	++
369	Q4FK11	2.66E-08	Nono	Non-POU domain-containing octamer-binding p	Q4FK11	25.203262	22.886758	25.1284739	21.539871	+
920	Q3TWG9	2.65E-08	Serpinh1	Serpin H1	Q3TWG9	28.071528	26.695589	28.0594082	26.62692	+
206	P63321	2.70E-08	Rala	Ras-related protein Ral-A	P63321	23.358503	20.375246	23.2688437	0	++
114	Q7TPT7	3.18E-08	Vars	Valine--tRNA ligase	Q7TPT7	24.159525	22.516292	24.0481048	21.561599	+
123	Q3TQ70	3.28E-08	Gnb1;Gnb3	Guanine nucleotide-binding protein G(I)/G(S)/G	Q3TQ70	27.347329	25.366671	27.3341644	25.247514	+
719	O54724	3.31E-08	Ptrf	Polymerase I and transcript release factor	O54724	26.863355	24.990947	26.8485587	24.770805	+
346	P26041	3.71E-08	Msn	Moesin	P26041	26.068055	24.540936	26.0318233	24.268725	+
859	Q9WV55	3.83E-08	Vapa	Vesicle-associated membrane protein-associate	Q9WV55	24.344843	22.098581	24.3294465	21.71246	+
1211	Q80WJ7	5.08E-08	Mtdh	Protein LYRIC	Q80WJ7	23.308514	21.709634	23.2978687	21.313727	+
962	A0A0R4J0Z1	5.16E-08	Pdia4	Protein disulfide-isomerase A4	A0A0R4J0Z1	25.434166	23.741499	25.3701774	23.230621	+
85	Q8CBP1	5.37E-08	Yes1	Non-specific protein-tyrosine kinase;Tyrosine-p	Q8CBP1	25.400478	22.038696	25.3729328	20.135806	++
1386	Q545V8	5.85E-08	Csnk2a2	Casein kinase II subunit alpha	Q545V8	23.164233	20.970786	23.096618	17.949232	++
80	Q02257	6.09E-08	Jup	Junction plakoglobin	Q02257	26.372049	25.284582	26.3514984	25.16651	+
265	Q91XV3	6.18E-08	Basp1	Brain acid soluble protein 1	Q91XV3	27.830628	25.68407	27.8193369	25.580556	+
277	Q8VIJ6	6.52E-08	Sfpq	Splicing factor, proline- and glutamine-rich	Q8VIJ6	26.664663	24.211179	26.6382454	23.73472	+
102	E9PZ16	6.90E-08	Hspg2	Basement membrane-specific heparan sulfate p	E9PZ16	25.9639	24.690421	25.8660954	24.088052	
729	Q8C5Q5	6.97E-08	Cct7	T-complex protein 1 subunit eta	Q8C5Q5	25.10974	23.469849	25.0569556	22.958588	+
517	Q52KG9	7.29E-08	Cct6a;Cct6b	T-complex protein 1 subunit zeta;T-complex pro	Q52KG9	24.762271	22.302391	24.7330153	21.576623	++
332	P11983	7.51E-08	Tcp1;Tcp-1	T-complex protein 1 subunit alpha	P11983	25.54932	23.724433	25.5101473	23.127798	+
912	P21981	7.69E-08	Tgm2	Protein-glutamine gamma-glutamyltransferase	P21981	24.985256	21.308305	24.9453939	0	+
1187	Q61177	8.02E-08	Csnk2a1	Casein kinase II subunit alpha	Q61177	23.509045	20.928169	23.448882	19.023089	+
461	Q8CIJ3	8.66E-08	Eif3b	Eukaryotic translation initiation factor 3 subunit	Q8CIJ3	23.744015	22.194679	23.6518545	21.186403	+
738	Q4FZL1	1.01E-07	Eif4a1	Eukaryotic initiation factor 4A-I	Q4FZL1	25.568396	23.800832	25.5177733	23.370385	+
455	Q3UP03	1.01E-07	Ablim1	Actin-binding LIM protein 1	Q3UP03	24.849594	23.300762	24.774488	22.650332	
698	H7BX44	1.04E-07	Cdc42bpa	Non-specific serine/threonine protein kinase;Se	H7BX44	23.552384	21.314402	23.4470783	17.698922	+
319	G5E896	1.09E-07	Edc4	Enhancer of mRNA-decapping protein 4	G5E896	22.853395	21.305898	22.7309825	19.686978	
716	A0A0R4J0I9	1.13E-07	Lrp1	Prolow-density lipoprotein receptor-related prot	A0A0R4J0I9	23.974461	22.48895	23.8283151	20.882136	+
1077	Q63918	1.13E-07	Sdpr	Serum deprivation-response protein	Q63918	25.496229	22.636102	25.4679585	21.568054	++
36	Q9ERG0	1.16E-07	Lima1	LIM domain and actin-binding protein 1	Q9ERG0	28.9038	27.633812	28.8931148	27.581973	
549	P49817	1.24E-07	Cav1	Caveolin-1;Caveolin	P49817	27.445526	25.405408	27.4379568	25.309416	++
1159	Q8BTS0	1.46E-07	Ddx5;Hlr1	Probable ATP-dependent RNA helicase DDX5	Q8BTS0	25.775267	24.456504	25.7536119	24.341133	+
761	Q03265	1.47E-07	Atp5a1	ATP synthase subunit alpha, mitochondrial;ATP	Q03265	25.950396	24.406789	25.9295692	24.202109	+
562	Q545K5	1.58E-07	Rbm3	RNA-binding protein 3	Q545K5	24.86677	20.756415	24.8321578	17.203377	++
665	Q60598	1.76E-07	Cttn	Src substrate cortactin	Q60598	26.624213	25.07405	26.6077587	24.941822	+
520	Q9CPN8	1.78E-07	Igf2bp3	Insulin-like growth factor 2 mRNA-binding prote	Q9CPN8	24.37748	22.372469	24.3464343	21.673722	++
831	Q5SW83	1.79E-07	Actr2	Actin-related protein 2	Q5SW83	26.0257	24.82919	25.9873955	24.643987	+
229	Q6R891	1.88E-07	Ppp1r9b;Ppp1	Neurabin-2	Q6R891	24.314474	21.643594	24.2808823	19.92359	+
991	P63101	2.25E-07	Ywhaz	14-3-3 protein zeta/delta	P63101	26.823509	23.984052	26.8155703	23.741839	++
552	Q99JY0	2.57E-07	Hadhb	Trifunctional enzyme subunit beta, mitochondri	Q99JY0	24.719237	22.550569	24.6718824	22.032877	+
799	Q9JM76	2.65E-07	Arpc3	Actin-related protein 2/3 complex subunit 3	Q9JM76	25.318803	24.1356	25.2860119	23.94738	+
171	Q71LX8	2.70E-07	Hsp90ab1	Heat shock protein HSP 90-beta	Q71LX8	27.714815	26.212049	27.6920383	26.080431	+
84	Q64727	2.73E-07	Vcl	Vinculin	Q64727	25.362164	24.135195	25.2986045	23.819134	+
1588	Q9R0Y5	2.83E-07	Ak1	Adenylate kinase isoenzyme 1	Q9R0Y5	24.405412	21.839321	24.3601376	20.485858	++
236	Q542G9	2.98E-07	Anxa2	Annexin;Annexin A2	Q542G9	27.696891	26.613028	27.6836342	26.553766	+
1384	Q3TF41	3.05E-07	Nap1l1	Nucleosome assembly protein 1-like 1	Q3TF41	25.595206	23.118614	25.5719563	22.741125	++
880	Q925B0	3.28E-07	Pawr	PRKC apoptosis WT1 regulator protein	Q925B0	25.647522	23.535689	25.6284093	23.195427	++
1008	Q91V38	3.50E-07	Hsp90b1	Endoplasmin	Q91V38	27.45652	26.015201	27.425701	25.854573	
1708	Q91X95	3.60E-07	Gna11;Gna14	Guanine nucleotide-binding protein subunit alph	Q91X95	23.902331	21.272296	23.8747955	20.119595	+
692	Q542X7	4.95E-07	Cct2	T-complex protein 1 subunit beta	Q542X7	25.320862	23.241734	25.2592411	22.448163	+
1500	Q561M4	5.04E-07	Tom1	Target of Myb protein 1	Q561M4	23.937478	22.705182	23.8122716	21.896255	+
417	Q8BG67	5.32E-07	Efr3a	Protein EFR3 homolog A	Q8BG67	22.597189	20.484544	22.4960676	0	+
688	Q9CVB6	5.60E-07	Arpc2	Actin-related protein 2/3 complex subunit 2	Q9CVB6	26.394241	24.895666	26.376137	24.803935	

935	Q569X4	5.58E-07	Psmc4	26S protease regulatory subunit 6B	Q569X4	21.819487	20.30424	21.4840484	0	
818	Q7TMS2	5.72E-07	Sipa1	Signal-induced proliferation-associated protein 1	Q7TMS2	22.829224	20.615085	22.7285443	0	+
513	Q3U379	5.97E-07	Gpc1	Glypican-1;Secreted glypican-1	Q3U379	24.968772	22.251831	24.9034391	20.596948	+
130	Q3U9V4	6.09E-07	Gnb2;Gnb4	Guanine nucleotide-binding protein G(I)/G(S)/G	Q3U9V4	25.911031	22.998222	25.9030683	22.70103	++
595	Q564E8	6.61E-07	Rpl4	60S ribosomal protein L4	Q564E8	25.900412	24.422941	25.8762849	24.231554	+
488	Q8CD23	8.01E-07	Ncl	Nucleolin	Q8CD23	27.540444	25.865451	27.5233816	25.7748	
163	Q8CB16	8.39E-07	Palm	Paralemmin-1	Q8CB16	25.100978	22.954515	25.0566482	22.334323	+
184	Q5DQJ3	9.18E-07	Capza2	F-actin-capping protein subunit alpha-2	Q5DQJ3	26.563778	25.076831	26.543462	24.96795	
366	Q542D9	9.74E-07	Tfrc	Transferrin receptor protein 1	Q542D9	23.251812	21.768436	23.0776842	19.492965	+
1597	O70251	9.87E-07	Eef1b;Eef1b2	Elongation factor 1-beta	O70251	25.219783	23.094926	25.1983746	22.810345	+
629	P48678	9.98E-07	Lmna	Prelamin-A/C;Lamin-A/C	P48678	25.66141	24.074116	25.6091547	23.767469	
1451	Q5SX74	1.00E-06	P4ha2	Prolyl 4-hydroxylase subunit alpha-2	Q5SX74	24.895684	22.254026	24.8601236	21.246339	+
140	Q564F3	1.07E-06	Rps3a1;Rps3a	40S ribosomal protein S3a	Q564F3	26.420161	24.424915	26.3951552	24.220148	+
490	P27773	1.18E-06	Pdia3	Protein disulfide-isomerase A3	P27773	27.199561	25.763295	27.1768581	25.654748	
1346	Q9CPY7-2	1.33E-06	Lap3	Cytosol aminopeptidase	Q9CPY7-2	22.501225	20.245225	22.3251872	0	++
78	E9PVC6	1.40E-06	Eif4g1;eIF4GI	Eukaryotic translation initiation factor 4 gamma	E9PVC6	24.157536	22.571268	24.004789	21.413804	
125	Q61553	1.50E-06	Fscn1	Fascin	Q61553	25.958996	23.709913	25.9306645	23.348397	+
1155	Q7TSZ3	1.54E-06	Lars	Leucine--tRNA ligase, cytoplasmic	Q7TSZ3	23.3172	21.299384	23.1667944	17.701662	+
1783	Q91W53	1.69E-06	Golga7;GOLG	Golgin subfamily A member 7;Golgin subfamily	Q91W53	22.396738	20.032459	22.2799241	0	+
837	Q3UWP8	1.85E-06	Calr	Calreticulin	Q3UWP8	27.166958	24.630747	27.148361	24.388965	++
333	Q3TW40	1.87E-06	Hnrnpm	Heterogeneous nuclear ribonucleoprotein M	Q3TW40	24.699761	23.443485	24.6510837	23.107721	+
1324	Q3UG68	1.94E-06	Ddost	Dolichyl-diphosphooligosaccharide--protein gly	Q3UG68	23.276623	20.767406	23.2173456	18.518918	+
1133	A0A0U1RNK7	2.23E-06	Dock7;mKIAA	Dedicator of cytokinesis protein 7;Dedicator of	A0A0U1RNK7	24.721221	23.622165	24.6014085	22.923603	
235	P26645	2.27E-06	Marcks	Myristoylated alanine-rich C-kinase substrate	P26645	25.581868	23.502972	25.550202	23.174526	+
7	Q76MZ3	2.53E-06	Ppp2r1a	Serine/threonine-protein phosphatase 2A 65 kD	Q76MZ3	24.623027	22.801738	24.5455852	22.039668	+
119	Q61245	2.60E-06	Col11a1	Collagen alpha-1(XI) chain	Q61245	25.174474	22.353358	25.1125973	20.942481	+
159	P51863	2.84E-06	Atp6v0d1	V-type proton ATPase subunit d 1	P51863	23.901021	21.78204	23.8385172	20.790439	+
478	Q3THH1	3.51E-06	Pdia6	Protein disulfide-isomerase A6	Q3THH1	26.240598	24.229952	26.2284163	24.042571	++
1203	Q91VJ2	3.50E-06	Prkcdbp	Protein kinase C delta-binding protein	Q91VJ2	22.735376	20.342098	22.6157739	0	+
98	Q9D0I9	3.52E-06	Rars	Arginine--tRNA ligase, cytoplasmic	Q9D0I9	23.580548	21.603926	23.4784142	19.554989	+
283	Q9R0E2	3.66E-06	Plod1	Procollagen-lysine,2-oxoglutarate 5-dioxygenas	Q9R0E2	23.734945	20.688262	23.6880756	0	+
1702	A2RS22	3.83E-06	Coro1b	Coronin;Coronin-1B	A2RS22	25.938675	24.59301	25.9210443	24.415074	++
58	Q9D8B3	4.03E-06	Chmp4b	Charged multivesicular body protein 4b	Q9D8B3	25.087621	23.131013	25.049161	22.733455	+
330	P21956-2	4.05E-06	Mfge8;MFG-E	Lactadherin	P21956-2	26.340058	25.464171	26.3185533	25.355176	+
635	Q3UJS0	4.20E-06	Rpl8	60S ribosomal protein L8	Q3UJS0	24.688202	23.030809	24.6789428	22.833659	++
586	Q80W68	4.43E-06	Kirrel	Kin of IRRE-like protein 1	Q80W68	23.123721	21.522421	23.0126908	20.220346	+
153	P11087	4.72E-06	Col1a1	Collagen alpha-1(I) chain	P11087	29.617242	27.463812	29.6033662	27.367083	
557	A0JLV3	4.76E-06	Hist1h2bj;Hist	Histone H2B;Histone H2B type 1-P;Histone H2B	A0JLV3	25.14742	21.291822	25.13532	20.277615	++
954	Q9DCY1	5.18E-06	Ppib	Peptidyl-prolyl cis-trans isomerase;Peptidyl-pro	Q9DCY1	25.069611	23.291229	25.0534378	23.049313	++
1298	Q99N15	5.57E-06	Hsd17b10	3-hydroxyacyl-CoA dehydrogenase type-2	Q99N15	22.547413	19.915736	22.510121	0	++
1306	Q3U518	5.59E-06	Prkcsh	Glucosidase 2 subunit beta	Q3U518	23.521052	21.642024	23.4109024	20.425045	+
1210	Q3U9Y8	5.70E-06	Psmd2;Gm54	26S proteasome non-ATPase regulatory subuni	Q3U9Y8	23.746937	22.04521	23.6673788	20.996524	+
1285	Q8C292	5.77E-06	Kars	Lysine--tRNA ligase	Q8C292	23.802257	22.129822	23.7200448	20.892589	+
429	Q8BKC5	6.49E-06	Ipo5	Importin-5	Q8BKC5	24.301148	22.724405	24.0973762	21.073847	
160	P35700	6.71E-06	Prdx1	Peroxiredoxin-1	P35700	25.214088	23.100304	25.1831386	22.731154	+
528	Q3U7R1	6.88E-06	Esyt1	Extended synaptotagmin-1	Q3U7R1	22.891736	21.831286	22.7539481	20.512492	
246	Q5SUC3	7.62E-06	Canx	Calnexin	Q5SUC3	25.766964	24.457452	25.7179145	24.235727	
577	Q9Z0U1	8.61E-06	Tjp2	Tight junction protein ZO-2	Q9Z0U1	22.792051	20.878573	22.6382523	17.601386	+
969	Q3TUF3	8.93E-06	Calu	Calumenin	Q3TUF3	25.411621	23.575627	25.3539735	23.070328	+
86	Q9JKR6	9.36E-06	Hyou1	Hypoxia up-regulated protein 1	Q9JKR6	23.985057	22.325916	23.8298391	20.959132	
389	B7ZND7	9.40E-06	Ppp1r18	Phostensin	B7ZND7	25.146751	23.186226	25.1081992	22.785648	+
873	Q91YR1	9.36E-06	Twf1	Twinfilin-1	Q91YR1	25.179527	23.462351	25.147926	23.21895	+
985	E9Q3E2	1.11E-05	Synpo	Synaptopodin	E9Q3E2	23.698483	21.706723	23.6024279	19.334544	+
12	Q3THQ5	1.12E-05	Stip1	Stress-induced-phosphoprotein 1	Q3THQ5	24.800563	22.015248	24.7582742	20.949437	+
271	Q3V3R1	1.16E-05	Mthfd1l	Monofunctional C1-tetrahydrofolate synthase, r	Q3V3R1	22.611644	20.904772	22.389819	0	
681	Q80Y09	1.17E-05	Pdcd6ip	Programmed cell death 6-interacting protein	Q80Y09	24.243025	23.081181	24.1101509	22.154078	
1090	Q3TE95	1.16E-05	Rcn2	Reticulocalbin-2	Q3TE95	23.142379	22.060497	23.0352756	21.410917	+
191	Q9JJ28	1.21E-05	Flii	Protein flightless-1 homolog	Q9JJ28	25.828461	24.841553	25.763617	24.572594	
759	Q5XJF6	1.31E-05	Rpl10a	Ribosomal protein;60S ribosomal protein L10a	Q5XJF6	25.512407	24.095222	25.4977944	23.96118	++
787	Q8BTF0	1.39E-05	Copa	Coatomer subunit alpha;Xenin;Proxenin	Q8BTF0	23.618576	22.528684	23.5089937	21.748168	
1805	G3X9L6	1.48E-05	Gm10250;Atp	ATP synthase subunit d, mitochondrial	G3X9L6	23.781619	21.951053	23.7209082	21.441433	+
164	Q8VDD5	1.56E-05	Myh9	Myosin-9	Q8VDD5	33.432638	33.457481	33.4285234	33.452222	
388	A0A0R4J259	1.56E-05	Syncrip;Hnrnp	Heterogeneous nuclear ribonucleoprotein Q	A0A0R4J259	24.730809	23.201503	24.6712476	22.729758	+
924	Q3V1F2	1.56E-05	Sdc1	Syndecan;Syndecan-1	Q3V1F2	23.871525	20.754911	23.8429239	19.318492	+
493	E9Q175	1.57E-05	Myo6	Unconventional myosin-VI	E9Q175	25.819963	25.259951	25.722408	25.049007	
1020	G3X9J0	1.59E-05	Sipa1l3		G3X9J0	22.21984	22.311897	22.0316873	21.949234	++
1533	Q91YK6	1.79E-05	Rpl23a;Gm60	60S ribosomal protein L23a	Q91YK6	25.559741	22.326992	25.5484498	21.803452	++
124	Q3TRJ1	1.84E-05	Vps35	Vacuolar protein sorting-associated protein 35	Q3TRJ1	24.072678	22.651135	23.863557	21.219536	
1978	P60766	1.83E-05	Cdc42	Cell division control protein 42 homolog	P60766	24.38827	22.806309	24.3737675	22.617335	+
298	A0A0R4J169	2.20E-05	Lrrfip2	Leucine-rich repeat flightless-interacting protein	A0A0R4J169	25.588399	24.51884	25.5622988	24.372145	+
1047	Q811L7	2.49E-05	Hnrnph1	Heterogeneous nuclear ribonucleoprotein H;Het	Q811L7	24.185889	22.581417	24.1464971	22.306742	+
1448	Q64521	2.64E-05	Gpd2	Glycerol-3-phosphate dehydrogenase, mitochor	Q64521	22.006188	20.227743	21.7608167	0	+
567	Q3U4U6	2.85E-05	Cct3;Cctg	T-complex protein 1 subunit gamma	Q3U4U6	25.108856	24.866291	25.0206588	24.713217	
521	Q9JKF1	2.87E-05	Iqgap1	Ras GTPase-activating-like protein IQGAP1	Q9JKF1	26.536557	25.982123	26.4696205	25.79858	
1365	Q3UNN2	3.01E-05	Cd44	CD44 antigen	Q3UNN2	25.475455	24.030431	25.4522869	23.895655	+
1740	P99027	3.08E-05	Rplp2	60S acidic ribosomal protein P2	P99027	26.560274	25.271665	26.5548133	25.213364	+
777	Q9R0E1	3.12E-05	Plod3	Procollagen-lysine,2-oxoglutarate 5-dioxygenas	Q9R0E1	24.661748	23.000145	24.5483171	22.323833	
234	Q8R4U7	3.20E-05	Luzp1;mFLJ00	Leucine zipper protein 1	Q8R4U7	24.461092	23.378088	24.3238545	22.878171	
237	Q58EU3	3.26E-05	Rps2;Gm1802	40S ribosomal protein S2	Q58EU3	26.128463	24.546078	26.1031856	24.406077	
619	Q91V55	3.26E-05	Rps5	40S ribosomal protein S5;40S ribosomal protein	Q91V55	25.555966	23.99987	25.5338927	23.863424	+
1804	P42208	3.24E-05	Sep-02	Septin-2	P42208	23.201971	21.372512	23.1072733	18.431192	
1201	Q544H0	3.50E-05	Eif3g	Eukaryotic translation initiation factor 3 subunit	Q544H0	23.206136	21.756316	23.1496016	20.88793	+
1038	Q91VR5	3.64E-05	Ddx1	ATP-dependent RNA helicase DDX1	Q91VR5	24.812376	24.060953	24.7302215	23.806197	
40	P67778	3.74E-05	Phb;1700071I	Prohibitin	P67778	26.163895	22.569241	26.1519622	21.989072	++
56	Q58E35	3.74E-05	Rplp1;Gm100	60S acidic ribosomal protein P1	Q58E35	25.073962	23.752782	25.0544141	23.616214	+
210	Q8BRD3	3.77E-05	Nucb1	Nucleobindin-1	Q8BRD3	23.04316	20.849096	22.9576027	0	+
477	P50516-2	3.80E-05	Atp6v1a	V-type proton ATPase catalytic subunit A	P50516-2	23.855268	22.280465	23.7620668	21.455418	
433	E9QPE7	4.10E-05	Myh11	Myosin-11	E9QPE7	24.342916	24.49353	24.1548346	24.42717	--

593	Q3U741	4.15E-05	Ddx17	Probable ATP-dependent RNA helicase DDX17	Q3U741	25.145909	23.341041	25.1128887	22.932512	+
401	Q5SX50	4.24E-05	Pfn1	Profilin;Profilin-1	Q5SX50	25.372097	24.019429	25.3151874	23.78892	+
1549	Q8K2J0	4.25E-05	Plcd3	1-phosphatidylinositol 4,5-bisphosphate phosph	Q8K2J0	22.509732	20.65842	22.3840593	0	
553	P18872-2	4.33E-05	Gnao1	Guanine nucleotide-binding protein G(o) subunit	P18872-2	23.738759	21.183677	23.6693352	18.095021	+
766	A6PW84	4.56E-05	P3h1;Lepre1	Prolyl 3-hydroxylase 1	A6PW84	23.863088	21.533919	23.7704792	20.259269	+
1558	Q8C2U7	4.62E-05	Aimp1	Aminoacyl tRNA synthase complex-interacting r	Q8C2U7	22.375175	19.656923	22.2937164	0	+
300	Q5SYD0	4.72E-05	Myo1d;Myo1	Unconventional myosin-Id;Unconventional myo	Q5SYD0	26.314338	25.321944	26.2664761	25.118042	
487	Q80U72	4.77E-05	Scrib;Lrrc1	Protein scribble homolog;Leucine-rich repeat-co	Q80U72	22.594098	21.535843	22.3236338	19.751812	
1181	Q4VAC9-2	4.91E-05	Plekhg3	Pleckstrin homology domain-containing family G	Q4VAC9-2	23.147655	21.87146	22.8504047	20.5033	
794	P08032	5.21E-05	Spta1	Spectrin alpha chain, erythrocytic 1	P08032	21.635342	20.10057	21.3522001	0	
613	Q8R509	5.31E-05	Ptbp1;Ptbp3	Polypyrimidine tract-binding protein 1;Polypyrin	Q8R509	24.364975	22.312679	24.3239665	21.891599	+
1063	A0A0U1RPX7	5.32E-05	Myo7a;Myo7	Unconventional myosin-VIIa;Unconventional m	A0A0U1RPX7	23.942144	24.673485	23.7848904	24.509357	
439	Q3U0V1	5.60E-05	Khsrp	Far upstream element-binding protein 2	Q3U0V1	23.793635	21.853579	23.6983799	20.971129	
701	Q3U094	5.61E-05	Arpc1b	Actin-related protein 2/3 complex subunit 1B	Q3U094	26.232917	25.321839	26.2196604	25.270678	
275	A1A4T2	6.02E-05	Ganab	Neutral alpha-glucosidase AB	A1A4T2	23.742105	22.287882	23.6326422	21.067881	
198	Q61768	6.15E-05	Kif5b;mKIAA0	Kinesin-1 heavy chain;Kinesin-like protein;Kines	Q61768	22.500817	20.895435	22.3671864	18.464548	
902	Q9WVA4	6.55E-05	Tagln2	Transgelin-2	Q9WVA4	24.776428	23.540659	24.739279	23.277853	+
1194	P62830	6.92E-05	Rpl23	60S ribosomal protein L23	P62830	25.0562	24.134237	25.0303052	24.05987	
1089	Q9QZD9	7.17E-05	Eif3i	Eukaryotic translation initiation factor 3 subunit	Q9QZD9	23.290739	21.433319	23.1823159	19.19416	
846	P23242	7.25E-05	Gja1	Gap junction alpha-1 protein;Gap junction prote	P23242	25.038321	23.553612	25.001955	23.299166	+
93	E9Q5B5	7.38E-05	Hk2	Hexokinase;Hexokinase-2	E9Q5B5	21.586198	20.178392	21.3446481	0	
344	E9QA63	7.76E-05	Macf1	Microtubule-actin cross-linking factor 1	E9QA63	22.912475	21.892183	22.6908572	20.380614	++
641	Q3U6A2	7.83E-05	Rbm14;p16;G	RNA-binding protein 14	Q3U6A2	24.968544	23.236947	24.930849	22.922193	+
1131	Q3UC68	7.81E-05	Sdcbp	Syntenin-1	Q3UC68	23.211654	21.47967	23.0701025	20.260596	
578	P23116	7.97E-05	Eif3a	Eukaryotic translation initiation factor 3 subunit	P23116	23.956914	22.571754	23.7709364	21.489337	
662	A0A0N4SVT3	8.09E-05	Gng12	Guanine nucleotide-binding protein G(I)/G(S)/G	A0A0N4SVT3	26.712901	24.344645	26.7047435	24.219874	++
892	B2RTM0	8.10E-05	Hist2h4;Hist1H	Histone H4	B2RTM0	25.907027	22.333668	25.8980699	21.866194	++
169	Q497E9	8.30E-05	Rps8	40S ribosomal protein S8	Q497E9	25.921219	24.301273	25.9086328	24.186996	+
251	Q6ZWZ6	8.33E-05	Rps12;Gm492	40S ribosomal protein S12	Q6ZWZ6	25.587012	23.842259	25.5598393	23.64249	+
385	Q3UB90	8.46E-05	Rpl3;Rp13	60S ribosomal protein L3	Q3UB90	25.397124	24.077941	25.3452605	23.866884	
523	P17426-2	8.69E-05	Ap2a1	AP-2 complex subunit alpha-1	P17426-2	21.421086	19.654005	21.1775861	0	
507	E9Q557	8.73E-05	Dsp	Desmoplakin	E9Q557	25.71827	24.913301	25.6456096	24.659112	
260	Q99PT1	8.81E-05	Arhgdia	Rho GDP-dissociation inhibitor 1	Q99PT1	24.138201	22.491266	24.0597685	21.939613	+
982	Q99KI0	8.88E-05	Aco2	Aconitate hydratase, mitochondrial	Q99KI0	24.312416	22.47859	24.23509	21.332889	+
565	Q9Z329	9.52E-05	Itpr2	Inositol 1,4,5-trisphosphate receptor type 2	Q9Z329	22.768674	21.249652	22.5916756	17.84671	
363	Q3U962	9.61E-05	Col5a2	Collagen alpha-2(V) chain	Q3U962	25.796397	24.005101	25.7182377	23.441009	
674	Q05186	9.79E-05	Rcn1	Reticulocalbin-1	Q05186	24.402323	22.655566	24.3641811	22.257106	+
498	Q8BK18	0.000101405	Dars	Aspartate--tRNA ligase, cytoplasmic	Q8BK18	22.943284	21.729312	22.8154468	20.956534	
581	B2RSC8	0.000101771	Nedd4	E3 ubiquitin-protein ligase NEDD4	B2RSC8	24.884824	24.008944	24.8305084	23.784316	
585	Q9CWI9	0.000102805	Rps23	40S ribosomal protein S23	Q9CWI9	24.019256	21.604888	24.013317	21.401383	++
722	Q9Z2X1	0.000102878	Hnrnpf	Heterogeneous nuclear ribonucleoprotein F;Het	Q9Z2X1	24.785387	23.075029	24.7321269	22.654228	
18	Q9JIF7	0.000104264	Copb1	Coatomer subunit beta	Q9JIF7	21.976194	20.208042	21.6985599	0	
782	A2A5N2	0.000104652	Ywhab	14-3-3 protein beta/alpha;14-3-3 protein beta/a	A2A5N2	24.488028	22.477508	24.448576	21.937576	+
1283	Q69Z38	0.000110361	Peak1	Pseudopodium-enriched atypical kinase 1	Q69Z38	22.406958	20.728058	22.2655513	18.502232	
972	Q9DC51	0.000114102	Gnai3	Guanine nucleotide-binding protein G(k) subunit	Q9DC51	25.358873	23.452255	25.3213292	23.16979	+
1573	Q8QZY1	0.00011585	Eif3l	Eukaryotic translation initiation factor 3 subunit	Q8QZY1	23.590301	21.94161	23.506498	21.036319	
817	Q52LS0	0.000121857	Rap1b	Ras-related protein Rap-1b	Q52LS0	24.930533	22.413057	24.892736	21.863344	+
687	Q3UHW9	0.000125184	Cfl2	Cofilin-2	Q3UHW9	25.130742	23.565661	25.1090597	23.379846	+
1037	P17751	0.000125184	Tpi1	Triosephosphate isomerase	P17751	24.920817	23.211955	24.8993657	22.993435	++
894	Q6PB66	0.000125646	Lrpprc	Leucine-rich PPR motif-containing protein, mitoc	Q6PB66	22.63572	21.251486	22.4679133	18.700367	
117	S4R189	0.000129181	Arhgef40	Rho guanine nucleotide exchange factor 40	S4R189	22.893756	20.58028	22.7839842	0	
851	Q9DCC5	0.000129311	Cbx3;Cbx5	Chromobox protein homolog 3;Chromobox prote	Q9DCC5	23.243996	20.173723	23.1746205	0	+
1069	E9PYB0	0.000129311	Ahnak2		E9PYB0	22.638003	20.633482	22.5852247	18.906535	+
1991	O08583-2	0.000129311	Alyref;Alyref2	THO complex subunit 4;Aly/REF export factor 2	O08583-2	22.900199	20.636186	22.8195257	19.045331	+
1081	Q5M9L7	0.000136086	Rps17	40S ribosomal protein S17	Q5M9L7	24.743192	23.238954	24.7375379	23.117242	+
1220	Q9CQ65	0.000136086	Mtap	S-methyl-5-thioadenosine phosphorylase	Q9CQ65	22.561509	21.222434	22.4896536	20.584203	
209	Q543N7	0.000148468	Pacsin3	Protein kinase C and casein kinase II substrate p	Q543N7	22.67135	20.93717	22.5071828	17.106093	
371	P61164	0.000148468	Actr1a	Alpha-centractin	P61164	23.037911	21.410661	22.94411	20.535916	
30	Q60951	0.000152905	Ybx1	Nuclease-sensitive element-binding protein 1	Q60951	26.524245	24.860894	26.4966183	24.683584	
554	Q3TUI1	0.000155022	Anxa6	Annexin	Q3TUI1	25.355084	24.151736	25.2631911	23.840824	
1167	O08807	0.000155115	Prdx4	Peroxiredoxin-4	O08807	24.032461	21.931729	24.0189123	21.586264	+
1725	O70194	0.000155286	Eif3d	Eukaryotic translation initiation factor 3 subunit	O70194	23.60875	21.36913	23.5433259	19.910197	
919	Q8BU30	0.000161806	Iars	Isoleucine--tRNA ligase, cytoplasmic	Q8BU30	23.249874	22.167113	23.0687179	21.415914	
31	A6H659	0.000176538	Pag1	Phosphoprotein associated with glycosphingolip	A6H659	21.986013	20.094461	21.8167206	0	
630	Q9CYL5	0.000178269	Glipr2	Golgi-associated plant pathogenesis-related pro	Q9CYL5	27.639544	26.069188	27.6337287	26.024945	+
741	Q6ZWQ9	0.000179381	Myl12a		Q6ZWQ9	30.460929	29.94719	30.4587878	29.940945	
292	Q543M7	0.000190034	Kpna3	Importin subunit alpha;Importin subunit alpha-4	Q543M7	22.494014	19.995756	22.3451595	0	+
335	Q922Z3	0.000190034	Trap1	Heat shock protein 75 kDa, mitochondrial	Q922Z3	21.675373	19.908293	21.543314	0	
257	B2RXS4	0.000199994	Plxnb2;mKIAA	Plexin-B2	B2RXS4	23.088801	21.632883	22.8477205	17.729819	
1577	Q3TV94	0.000201172	Ssr1	Translocon-associated protein subunit alpha	Q3TV94	22.333366	20.404391	22.2927458	19.827358	+
1330	Q08288	0.000205061	Lyar	Cell growth-regulating nucleolar protein	Q08288	22.573949	20.904736	22.5013094	20.128791	
1208	B2RXM2	0.000209996	Gm6793;Hnrn	Heterogeneous nuclear ribonucleoprotein A3	B2RXM2	25.725952	23.684403	25.7075782	23.504192	+
116	Q5FW97	0.000217245	EG433182;En	Alpha-enolase;Enolase;Gamma-enolase	Q5FW97	25.989496	24.974866	25.9283144	24.778009	
1161	Q5FWJ5	0.000219252	Hnrnpk	Heterogeneous nuclear ribonucleoprotein K	Q5FWJ5	25.193972	23.838873	25.15181	23.530127	
785	P62814	0.000223986	Atp6v1b2;Atp	V-type proton ATPase subunit B, brain isoform	P62814	23.372089	21.571489	23.2928607	20.483211	+
563	Q8BH80	0.000227661	Vapb	Vesicle-associated membrane protein-associate	Q8BH80	22.665243	21.335569	22.5565062	20.570197	
1150	Q3TJG6	0.000228358	Ptges3	Prostaglandin E synthase 3	Q3TJG6	23.829856	21.35082	23.7975452	20.257205	+
1388	F6ZV59	0.000228358	Hnrnpd	Heterogeneous nuclear ribonucleoprotein D0	F6ZV59	24.436265	22.154353	24.4145652	21.735433	+
1414	Q8BGH2	0.000226971	Samm50	Sorting and assembly machinery component 50	Q8BGH2	23.19989	21.519599	23.1024107	20.400933	
133	P16045	0.000232638	Lgals1	Galectin-1	P16045	26.867542	25.104982	26.8544925	25.019906	+
156	P63242	0.000234584	Eif5a;Eif5a2	Eukaryotic translation initiation factor 5A-1;Euk	P63242	24.379198	22.2192	24.3027838	21.325576	
256	Q542F1	0.00024621	Clic1	Chloride intracellular channel protein;Chloride in	Q542F1	24.023082	22.541057	23.9757664	22.117358	
15	Q9CRB9	0.000250456	Chchd3	MICOS complex subunit Mic19	Q9CRB9	24.277399	22.348237	24.2528385	21.878091	
1041	Q6ZWS7	0.000250456	Camk2g;Cam	Calcium/calmodulin-dependent protein kinase t	Q6ZWS7	23.547223	21.807623	23.4786995	21.285966	
176	Q3UC02	0.000254844	Rps11	40S ribosomal protein S11	Q3UC02	25.304767	23.799564	25.2897122	23.648728	+
814	Q9CQ73	0.000257974	Pkp2		Q9CQ73	22.028419	20.36176	21.868281	17.428294	
1236	Q5SW88	0.000257974	Rab1;Rab1A	Ras-related protein Rab-1A	Q5SW88	24.296374	22.464952	24.2814158	22.208265	+
1394	P51859	0.000257974	Hdgf	Hepatoma-derived growth factor	P51859	22.779231	20.096701	22.6316981	0	

1276	Q60932-2	0.000261127	Vdac1	Voltage-dependent anion-selective channel pro	Q60932-2	25.696421	23.023706	25.6695186	22.598988	+
137	A1BN54	0.000274392	Actn1		A1BN54	22.980108	20.13161	22.9327685	19.068148	+
259	Q8CCG5	0.000274392	Ralb	Ras-related protein Ral-B	Q8CCG5	23.221896	19.888297	23.1726409	18.143024	+
1744	Q3UXI9	0.000274945	Ilf2	Interleukin enhancer-binding factor 2	Q3UXI9	21.95066	20.108899	21.7673738	17.78323	
462	P43274	0.000277644	Hist1h1e	Histone H1.4	P43274	26.220425	24.832224	26.210975	24.764691	++
1200	Q8BP47	0.000285842	Nars	Asparagine--tRNA ligase, cytoplasmic	Q8BP47	23.161915	21.159504	23.0231199	18.471592	+
709	P80316	0.000289627	Cct5	T-complex protein 1 subunit epsilon	P80316	24.756281	23.397204	24.6683888	22.876812	
930	Q5EBQ6	0.000290419	Rpl9;Rpl9-ps1	60S ribosomal protein L9	Q5EBQ6	25.506093	24.000027	25.484453	23.800921	
1556	S4R257	0.000290615	Gapdh;Gm383	Glyceraldehyde-3-phosphate dehydrogenase;G	S4R257	26.747074	25.402821	26.7366748	25.341254	+
289	Q9D3C4	0.000298002	Arpc4	Actin-related protein 2/3 complex subunit 4	Q9D3C4	25.368112	23.9435	25.3614197	23.863313	+
382	Q9EPU0-2	0.00030068	Upf1	Regulator of nonsense transcripts 1	Q9EPU0-2	24.287751	23.149101	24.1922335	22.571632	
682	Q3TJ52	0.0003108	Rad23b;Rad23	UV excision repair protein RAD23 homolog B;UV	Q3TJ52	20.600715	18.85634	20.4289219	0	+
705	Q3V1L5	0.0003108	Tpr	Nucleoprotein TPR	Q3V1L5	21.156257	18.94772	20.9968658	0	+
1262	A2BDX2	0.0003108	Dpm1	Dolichol-phosphate mannosyltransferase subuni	A2BDX2	21.086479	22.489776	21.01219	22.489776	
1481	P70698	0.000319723	Ctps1;Ctps	CTP synthase 1	P70698	23.355607	21.085454	23.2127472	19.61088	+
534	P68372	0.000322571	Tubb4b	Tubulin beta-4B chain	P68372	27.38167	26.735673	27.363851	26.682304	
413	Q8BQ03	0.000328396	Mcm5	DNA helicase;DNA replication licensing factor M	Q8BQ03	23.178835	21.457188	23.0993037	18.80007	
438	E0CXB1	0.000328998	Psma6	Proteasome subunit alpha type-6	E0CXB1	22.254226	19.750403	22.194573	0	+
1999	Q8R2K3	0.000329848	Ssbp1	Single-stranded DNA-binding protein;Single-stra	Q8R2K3	22.90022	20.842434	22.8703391	19.552706	
399	Q6P1A9	0.00033332	Rpl7a;Gm576	60S ribosomal protein L7a	Q6P1A9	25.164525	24.063508	25.1358153	23.965649	
749	A8IP69	0.00033332	Ywhag	14-3-3 protein gamma;14-3-3 protein gamma, f	A8IP69	25.611191	23.8152	25.5855979	23.567692	+
202	Q52JJ6	0.000337977	Enpep	Glutamyl aminopeptidase	Q52JJ6	21.742531	20.186356	21.4542968	0	
1137	Q9R0X4	0.000337977	Acot9;Acot10	Acyl-coenzyme A thioesterase 9, mitochondrial;	Q9R0X4	22.407371	20.145497	22.3519107	0	
23	Q6ZPF4	0.000342585	Fmnl3	Formin-like protein 3	Q6ZPF4	22.799835	20.979697	22.6507854	17.739814	
1279	I7HLV2	0.000342887	Rpl10;Rpl10l	60S ribosomal protein L10;60S ribosomal protei	I7HLV2	24.305275	23.159278	24.2682163	22.955348	+
231	Q60930	0.000354266	Vdac2	Voltage-dependent anion-selective channel pro	Q60930	25.313419	23.231378	25.2832592	22.954179	
273	Q3TWN8	0.000356897	Aldh18a1	Delta-1-pyrroline-5-carboxylate synthase;Gluta	Q3TWN8	25.927008	26.00481	25.8801052	25.935752	
1052	A0A0R4J0S7	0.000359831	Uaca	Uveal autoantigen with coiled-coil domains and	A0A0R4J0S7	22.346668	20.902871	22.199739	19.502704	
983	Q60865	0.000366156	Caprin1;rng10	Caprin-1	Q60865	26.187182	24.86287	26.1604887	24.715982	+
1427	Q9ESU7	0.000372206	Slc1a5	Amino acid transporter;Neutral amino acid trans	Q9ESU7	22.553413	20.479724	22.4172812	18.211038	+
1802	Q58DZ3	0.000372206	Rpl30	60S ribosomal protein L30	Q58DZ3	24.029201	22.13081	23.9973047	21.664307	+
2062	Q545X5	0.00037128	Plaur	Urokinase plasminogen activator surface recept	Q545X5	23.324385	20.068268	23.3031388	18.396603	+
454	Q6P5E4	0.000373563	Uggt1	UDP-glucose:glycoprotein glucosyltransferase 1	Q6P5E4	24.051188	22.860285	23.8584413	21.894176	
1106	P62071	0.000388447	Rras2	Ras-related protein R-Ras2	P62071	24.268316	22.955599	24.2261264	22.657546	+
112	Q3TXS7	0.000392437	Psmd1;Gm21	26S proteasome non-ATPase regulatory subuni	Q3TXS7	22.461183	20.826664	22.3756217	19.87102	
905	P11276	0.000392437	Fn1	Fibronectin;Anastellin	P11276	23.751061	20.336582	23.7204855	18.953016	
1290	Q8BU31	0.00039326	Rap2c;Rap2a	Ras-related protein Rap-2c;Ras-related protein F	Q8BU31	22.871546	20.478449	22.7754604	0	+
375	Q3UIJ2	0.000394949	Eif2s3x	Eukaryotic translation initiation factor 2 subunit	Q3UIJ2	24.347425	23.140412	24.2753902	22.798067	
378	B9EKB8	0.000400879	Igf2r	Cation-independent mannose-6-phosphate rece	B9EKB8	22.617547	21.059205	22.4093855	0	
610	G3UYD0	0.000409904	Gtf2i;Gtf2ird2	General transcription factor II-I;General transcri	G3UYD0	23.332314	21.700062	23.1861736	20.028234	
664	Q9D2G2	0.000415242	Dlst	Dihydrolipoyllysine-residue succinyltransferase	Q9D2G2	22.34127	21.26059	22.1857721	20.622117	
822	P62737	0.000453441	Acta2	Actin, aortic smooth muscle	P62737	31.356326	30.925616	31.3542465	30.921181	
865	Q91YS2	0.000468756	Rangap1	Ran GTPase-activating protein 1	Q91YS2	22.699873	20.975365	22.5275716	19.93404	
881	Q545F8	0.000482739	Rps4x;Rps4l;G	40S ribosomal protein S4;40S ribosomal protein	Q545F8	25.951354	24.657123	25.9339704	24.551028	
361	Q64337	0.000490252	Sqstm1	Sequestosome-1	Q64337	25.13078	23.808712	25.0550873	23.457802	
1391	Q0VBA8	0.000495534	Plau	Urokinase-type plasminogen activator;Urokinas	Q0VBA8	21.509542	19.899293	21.2995139	18.435943	
1626	Q5BLJ7	0.000511256	Rps13	40S ribosomal protein S13	Q5BLJ7	25.090346	23.377646	25.0789875	23.255491	+
712	Q4VBF8	0.000552572	Sipa1l1	Signal-induced proliferation-associated 1-like pr	Q4VBF8	22.078115	27.354122	21.7143735	27.341263	++
252	E9PYL9	0.000557234	Gm10036;Rpl	60S ribosomal protein L11	E9PYL9	24.771152	23.383436	24.7542354	23.281191	
746	Q3TF84	0.00055979	Lrrc59	Leucine-rich repeat-containing protein 59	Q3TF84	25.267239	24.218558	25.2494052	24.130732	
1078	P09055	0.000585129	Itgb1	Integrin beta-1	P09055	24.519169	23.66364	24.4207262	23.35152	
735	Q99P72	0.000590619	Rtn4	Reticulon-4	Q99P72	24.582371	23.339976	24.5006163	22.916525	
1141	Q5KU03	0.000590412	Gsk3b	Glycogen synthase kinase-3 beta	Q5KU03	20.7957	19.58941	20.4896734	0	+
280	Q5YLW3	0.000594395	Rps3	40S ribosomal protein S3	Q5YLW3	26.850157	25.808213	26.8348269	25.748545	
92	Q3ULW0	0.000620515	Ran;1700009I	GTP-binding nuclear protein Ran;GTP-binding nu	Q3ULW0	24.428283	23.787354	24.4001373	23.679404	+
990	Q5SVG5	0.000621601	Ap1b1	AP complex subunit beta;AP-1 complex subunit	Q5SVG5	20.109021	18.647256	20.0284015	0	
1475	B1AQR8	0.000621601	Lgals9	Galectin;Galectin-9	B1AQR8	21.591397	18.536774	21.5913972	0	+
1582	Q6NSP9	0.000621601	Hmga2	High mobility group protein HMGI-C	Q6NSP9	21.293232	19.731214	21.2932325	18.984736	
676	Q3UD67	0.000632572	Aars	Alanine--tRNA ligase, cytoplasmic	Q3UD67	23.19452	21.096917	23.0822795	19.097619	+
505	Q3UKC1	0.000657503	Tax1bp1	Tax1-binding protein 1 homolog	Q3UKC1	23.856182	22.96775	23.6927523	22.459296	
616	G5E8C3	0.000656027	Gprc5a	Retinoic acid-induced protein 3	G5E8C3	22.889957	19.513523	22.8584067	17.598292	+
1640	Q9Z1G3	0.000656027	Atp6v1c1	V-type proton ATPase subunit C 1	Q9Z1G3	21.448844	19.62277	21.3201134	0	
760	Q3UCH0	0.00066908	Rpl6;Gm5428	60S ribosomal protein L6	Q3UCH0	25.550916	23.851209	25.5271216	23.695738	
1684	Q5M9M4	0.000667562	Rps15a	40S ribosomal protein S15a	Q5M9M4	23.938316	22.489045	23.9089079	22.238361	
411	P09411	0.000677566	Pgk1;Pgk2	Phosphoglycerate kinase 1;Phosphoglycerate k	P09411	25.032064	23.432651	24.974451	22.993432	
190	P27546-3	0.000744169	Map4	Microtubule-associated protein 4;Microtubule-a	P27546-3	22.728746	21.758279	22.4903775	20.139435	
1393	F7DBB3	0.000742533	Ahnak2		F7DBB3	26.475387	25.150726	26.4480962	25.02319	
1620	Q8R050-2	0.000747408	Gspt1;Gspt2	Eukaryotic peptide chain release factor GTP-bin	Q8R050-2	22.398754	20.276344	22.2187642	0	
1489	E9QB02	0.000757447	Mars	Methionine--tRNA ligase, cytoplasmic	E9QB02	21.986171	20.341091	21.7603485	0	
311	B2RQQ5	0.000767178	Map1b;Map1	Microtubule-associated protein 1B;MAP1B heav	B2RQQ5	22.995677	22.000453	22.747531	20.988533	
334	Q5M8R8	0.000785476	Rplp0	60S acidic ribosomal protein P0	Q5M8R8	26.08118	24.645972	26.066017	24.547475	
1319	Q5PR09	0.000835766	Rpl32	60S ribosomal protein L32	Q5PR09	23.774675	21.187775	23.7025195	20.425468	+
304	Q99JB2	0.000848669	Stoml2	Stomatin-like protein 2, mitochondrial	Q99JB2	22.440413	21.257264	22.2722064	20.301426	
672	Q9EPK2-3	0.000857939	Rp2	Protein XRP2	Q9EPK2-3	22.67742	19.881435	22.6295326	17.840719	
1205	Q61833	0.000870014	Rpn2	Dolichyl-diphosphooligosaccharide--protein glyc	Q61833	23.612568	22.409172	23.5337226	21.873737	
559	Q5DTI2	0.000872785	Atp2a2;Atp2a	Sarcoplasmic/endoplasmic reticulum calcium AT	Q5DTI2	24.287821	22.912827	24.1918169	22.457696	
450	P60335	0.000908423	Pcbp1	Poly(rC)-binding protein 1	P60335	24.313264	22.867847	24.2449589	22.597181	
35	B1AWE0	0.000999669	Clta	Clathrin light chain A	B1AWE0	24.508467	22.191225	24.4791365	21.62219	
313	P30416	0.000999669	Fkbp4	Peptidyl-prolyl cis-trans isomerase FKBP4;Peptic	P30416	22.966457	20.429278	22.8573616	0	
4	O35226	0.001059778	Psmd4	26S proteasome non-ATPase regulatory subuni	O35226	21.404024	20.068939	21.0954125	0	
566	Q3TV20	0.001060435	Asns	Asparagine synthetase [glutamine-hydrolyzing]	Q3TV20	22.468855	21.41683	22.1250286	20.341016	
215	G3X922	0.001087256	Dnajc13		G3X922	22.52906	21.628181	22.2593911	19.909629	
349	Q0VGQ1	0.001087801	Hsd17b12	Very-long-chain 3-oxoacyl-CoA reductase	Q0VGQ1	21.159704	18.95206	21.0077622	0	
1176	P19157	0.001087801	Gstp1;Gstp2	Glutathione S-transferase P 1;Glutathione S-trar	P19157	21.495189	19.345063	21.33707	17.770803	
1209	Q8BGU5-2	0.001087801	Ccny	Cyclin-Y	Q8BGU5-2	22.2551	18.454256	22.1601864	0	
1412	Q8K298	0.001087801	Anln	Actin-binding protein anillin	Q8K298	19.936814	18.925249	19.4135979	0	
1712	P62137	0.001083935	Ppp1ca	Serine/threonine-protein phosphatase PP1-alpH	P62137	24.678173	22.788492	24.6550507	22.549117	+
651	Q3V235	0.00111259	Phb2	Prohibitin-2	Q3V235	22.865198	21.02283	22.8167449	20.245094	+

1300	P40336	0.001114581	Vps26a	Vacuolar protein sorting-associated protein 26A	P40336	21.712597	19.474083	21.5878899	0	
355	P08249	0.001130562	Mdh2	Malate dehydrogenase, mitochondrial	P08249	25.839394	24.398327	25.8090658	24.243106	
255	Q8R5L1	0.001160073	C1qbp	Complement component 1 Q subcomponent-bi	Q8R5L1	24.11363	21.960641	24.0723509	21.693211	
1080	Q3TMZ1	0.001163487	Pycr2	Pyrroline-5-carboxylate reductase;Pyrroline-5-c	Q3TMZ1	22.625029	21.063695	22.4843617	19.81931	
49	A0A0G2JEU1	0.001202385	Aldh2	Aldehyde dehydrogenase, mitochondrial	A0A0G2JEU1	23.024281	21.831935	22.888778	21.129635	
1255	Q61171	0.001238157	Prdx2	Peroxiredoxin-2	Q61171	23.688638	20.802721	23.6560727	19.745693	+
1362	Q9CQM8	0.001238157	Rpl21	60S ribosomal protein L21	Q9CQM8	23.262242	21.082912	23.211161	20.165954	+
1571	Q3U1S6	0.001238157	Csde1	Cold shock domain-containing protein E1	Q3U1S6	21.933301	20.227703	21.855821	18.73971	
1893	Q3TWV4	0.001238157	Ap2m1	AP-2 complex subunit mu	Q3TWV4	22.734428	20.625673	22.6908965	19.498185	
767	Q6P5F9	0.001271988	Xpo1	Exportin-1	Q6P5F9	22.195096	20.945101	21.9392311	18.531596	
899	Q4FZH2	0.001271988	Rpl26	60S ribosomal protein L26	Q4FZH2	25.188039	23.199383	25.1717841	23.019605	+
666	P54823	0.001287268	Ddx6	Probable ATP-dependent RNA helicase DDX6	P54823	23.904283	22.27903	23.7770729	21.449814	
541	Q8BFY9-2	0.001350077	Tnpo1	Transportin-1	Q8BFY9-2	22.462905	21.301458	22.2838497	19.351986	
203	Q80ZZ0	0.001362146	Ehd1;Ehd3	EH domain-containing protein 1;EH domain-con	Q80ZZ0	23.277584	22.596174	23.0559883	21.910735	
24	Q3V1M8	0.001379995	Hdlbp	Vigilin	Q3V1M8	23.196208	22.402716	22.9271394	21.643655	
627	P39688	0.00143261	Fyn	Tyrosine-protein kinase Fyn;Non-specific protei	P39688	22.014712	20.002283	21.8679622	18.854046	
680	Q545K4	0.001434955	Eif6	Eukaryotic translation initiation factor 6	Q545K4	22.217839	20.3206	22.0954015	0	
1456	D6REF7	0.00143261	Sgpl1	Sphingosine-1-phosphate lyase 1	D6REF7	22.7008	21.03216	22.6375424	20.758753	
1589	P97450	0.00143261	Atp5j	ATP synthase-coupling factor 6, mitochondrial	P97450	21.686591	19.55187	21.405445	18.405167	
134	Q9DCL9	0.001442136	Paics	Multifunctional protein ADE2;Phosphoribosylam	Q9DCL9	22.707271	21.000139	22.6105055	20.017892	
1149	Q5M9P3	0.001593159	Rps19	40S ribosomal protein S19	Q5M9P3	26.196218	24.41708	26.1921654	24.385351	
434	Q3UFJ3	0.001648178	Pdha1	Pyruvate dehydrogenase E1 component subuni	Q3UFJ3	21.488927	19.141797	21.3690661	0	
801	E9QMQ3	0.001663241	Triobp	TRIO and F-actin-binding protein	E9QMQ3	22.818857	21.071106	22.7262838	19.569907	
921	Q543N5	0.001676761	Clic4	Chloride intracellular channel protein;Chloride in	Q543N5	21.727513	19.73015	21.5424037	0	
945	P97855	0.001692601	G3bp1	Ras GTPase-activating protein-binding protein 1	P97855	25.795079	24.36159	25.7643414	24.191305	
509	Q7TMI0	0.001778827	Psmd11	26S proteasome non-ATPase regulatory subuni	Q7TMI0	22.672758	20.697711	22.6170498	18.693112	+
1273	Q9QYJ0	0.001784334	Dnaja2	DnaJ homolog subfamily A member 2	Q9QYJ0	23.231447	21.194007	23.1820921	20.20138	
481	Q91YZ8	0.00184518	Pabpc4;Gm10	Polyadenylate-binding protein	Q91YZ8	23.561008	21.846859	23.4841092	21.0825	
694	P63213	0.001864802	Gng2	Guanine nucleotide-binding protein G(I)/G(S)/G	P63213	21.735306	19.84053	21.6940268	19.071373	
1752	Q7JD03	0.001864802	COX2;mt-Co2	Cytochrome c oxidase subunit 2	Q7JD03	22.026965	18.90714	21.9270991	0	+
1594	Q3TF25	0.00187728	Atp5o	ATP synthase subunit O, mitochondrial	Q3TF25	22.846441	20.869929	22.6733121	18.999941	
1003	Q5BLK1	0.001905847	Rps6	40S ribosomal protein S6	Q5BLK1	25.497086	23.887763	25.4717844	23.767441	
1518	Q4KL76	0.00194382	Hspe1;Cpn10-	10 kDa heat shock protein, mitochondrial	Q4KL76	24.693451	22.277242	24.6829367	22.025966	+
670	Q6ZQ38	0.001973111	Cand1	Cullin-associated NEDD8-dissociated protein 1	Q6ZQ38	22.884548	21.564763	22.6334205	19.852698	
821	Q8BFW7-4	0.001991323	Lpp	Lipoma-preferred partner homolog	Q8BFW7-4	23.617652	22.295203	23.5500559	21.767849	
46	Q5SS40	0.002120919	Ywhae	14-3-3 protein epsilon	Q5SS40	26.584441	25.257399	26.561501	25.130954	
395	O35646	0.002165549	Capn6	Calpain-6	O35646	23.151128	20.5033	23.0856693	0	+
168	O08749	0.002247676	Dld	Dihydrolipoyl dehydrogenase, mitochondrial;Dil	O08749	21.925564	20.553076	21.7605538	19.855187	
221	A2RSV8	0.002247676	Cox4i1	Cytochrome c oxidase subunit 4 isoform 1, mitoc	A2RSV8	23.422528	21.49166	23.3734265	21.048134	+
525	Q8C7Q6	0.002316434	Cdh11	Cadherin-11	Q8C7Q6	21.028681	19.587572	20.7615244	17.88782	
1632	Q925I1-2	0.002354183	Atad3;Atad3a	ATPase family AAA domain-containing protein 3	Q925I1-2	22.155016	20.923483	21.9722109	19.6841	
852	Q0P6B2	0.002393533	Fubp1	Far upstream element-binding protein 1	Q0P6B2	21.585823	20.127445	21.3194646	0	
890	P84084	0.002393533	Arf5	ADP-ribosylation factor 5	P84084	23.615088	22.65404	23.5917405	22.467017	
1332	Q922I0	0.002393533	U2af2	Splicing factor U2AF 65 kDa subunit	Q922I0	21.913961	20.593629	21.6459075	19.037885	
73	Q8CHH7	0.00243142	Gcn1l1		Q8CHH7	20.471491	19.524456	20.1697176	0	+
1004	Q9Z247	0.002478968	Fkbp9	Peptidyl-prolyl cis-trans isomerase FKBP9	Q9Z247	22.958182	21.377222	22.8881291	20.029751	
752	Q3UT02	0.002535725	Larp4	La-related protein 4	Q3UT02	22.313581	20.711991	22.1427231	18.106518	
1452	Q9DBZ6	0.002535725	Abcd3	ATP-binding cassette sub-family D member 3	Q9DBZ6	22.50837	21.100143	22.3930574	20.07124	
1494	Q8BUM1	0.002559421	Tardbp	TAR DNA-binding protein 43	Q8BUM1	22.82568	21.581803	22.7762708	21.176389	
1553	Q9CQS8	0.002592801	Sec61b;Gm10	Protein transport protein Sec61 subunit beta	Q9CQS8	24.329396	22.788624	24.3228011	22.702175	
268	Q5M9M0	0.002642631	Rpl13a	60S ribosomal protein L13a	Q5M9M0	24.385014	23.052185	24.3628452	22.917001	+
64	Q8CFQ9	0.00278297	Fus	RNA-binding protein FUS	Q8CFQ9	24.345717	21.159312	24.304926	19.435596	
20	P28667	0.002793468	Marcksl1	MARCKS-related protein	P28667	23.078614	20.938889	23.0328603	19.41468	
174	Q6ZWM8	0.002939446	Ppp1cc	Serine/threonine-protein phosphatase;Serine/t	Q6ZWM8	22.206232	20.537924	22.1331644	19.946937	
384	Q542S9	0.002914511	Sntb2	Beta-2-syntrophin	Q542S9	22.368408	20.800052	22.2230526	18.724268	
428	Q3UPK6	0.002914033	Psma5;Gm83	Proteasome subunit alpha type;Proteasome sul	Q3UPK6	23.027567	20.144384	22.9552253	19.279816	
432	Q4FZE6	0.002936794	Rps7;Gm9493	40S ribosomal protein S7	Q4FZE6	25.680143	24.04357	25.6701243	23.976407	
501	Q3UVJ2	0.002856973	Cap1	Adenylyl cyclase-associated protein;Adenylyl cy	Q3UVJ2	23.331887	22.018218	23.2345292	21.460957	
643	Q9WUD1	0.002914033	Stub1	STIP1 homology and U box-containing protein 1	Q9WUD1	21.260318	19.696897	20.971652	0	
772	Q8BPG5	0.002936794	Rac1;Rac3;Ra	Ras-related C3 botulinum toxin substrate 1;Ras-	Q8BPG5	25.211022	22.775803	25.1824823	22.56863	
796	Q52L87	0.002952603	Tuba1c	Tubulin alpha-1C chain	Q52L87	21.15433	19.061794	20.9903155	0	++
960	Q3TJA9	0.002952603	Dnaja3	DnaJ homolog subfamily A member 3, mitochor	Q3TJA9	21.308826	18.851597	21.2766571	0	
1191	P54116	0.002939446	Stom	Erythrocyte band 7 integral membrane protein	P54116	20.516192	19.673898	20.419598	0	
1245	Q3UF30	0.002952603	S100a10	Protein S100-A10	Q3UF30	23.895371	18.85141	23.8953706	0	+
1354	Q7TMG8	0.002914033	Gbas	Protein NipSnap homolog 2	Q7TMG8	21.162503	19.083617	20.9580091	0	
1417	Q9ESU8	0.002952603	Slc1a4	Amino acid transporter;Neutral amino acid trans	Q9ESU8	20.825477	19.011817	20.6659687	0	
1515	Q91VM2	0.002939446	Snrpd3	Small nuclear ribonucleoprotein Sm D3	Q91VM2	22.64785	20.176539	22.6188226	19.10367	
1550	Q8CDD9	0.002952603	Lrif1	Ligand-dependent nuclear receptor-interacting f	Q8CDD9	22.725235	19.800113	22.69271	0	
1628	Q9DC23	0.002952603	Dnajc10	DnaJ homolog subfamily C member 10	Q9DC23	20.768549	19.730364	20.6883481	19.008648	
1678	Q4VA29	0.002936794	2700060E02R	UPF0568 protein C14orf166 homolog	Q4VA29	23.736913	22.355005	23.7066604	22.078345	
1737	Q3UNJ3	0.002952603	Fkbp10	Peptidyl-prolyl cis-trans isomerase FKBP10	Q3UNJ3	21.43728	18.820488	21.3567096	0	
1824	Q3TWW8	0.002914033	Srsf6;Srsf4	Serine/arginine-rich splicing factor 6;Serine/arg	Q3TWW8	22.639652	20.820232	22.5555202	20.199394	
1947	Q9DBH5	0.002952603	Lman2	Vesicular integral-membrane protein VIP36	Q9DBH5	21.414386	18.979556	21.2197221	0	
2016	Q6ZWV7	0.002914033	Rpl35;Gm102	60S ribosomal protein L35	Q6ZWV7	24.31209	19.968805	24.3006316	18.909711	++
508	Q8C1X9	0.002969701	Anxa3	Annexin;Annexin A3	Q8C1X9	24.23237	22.784798	24.1287868	22.441248	
364	Q9JKB3-2	0.00302205	Ybx3;Igf2bp3	Y-box-binding protein 3	Q9JKB3-2	25.193014	23.343098	25.1440138	22.935488	
284	Q3U4W8	0.003079735	Usp5	Ubiquitin carboxyl-terminal hydrolase;Ubiquitin	Q3U4W8	21.426941	19.892751	21.2866702	0	
1113	Q3TDD8	0.003087881	Eif4b	Eukaryotic translation initiation factor 4B	Q3TDD8	22.548087	20.172673	22.4445183	18.450498	
1613	O08734	0.003087881	Bak1	Bcl-2 homologous antagonist/killer	O08734	21.546682	19.931597	21.4267251	0	
1175	P47963	0.003157811	Rpl13	60S ribosomal protein L13	P47963	25.492083	24.118191	25.4806829	24.051621	
1382	Q3TTY6	0.003330262	Lin7c;Lin7a;Li	Protein lin-7 homolog C;Protein lin-7 homolog B;	Q3TTY6	22.816568	20.708588	22.7721243	18.640373	
348	Q9CWJ9	0.003354482	Atic	Bifunctional purine biosynthesis protein PURH;P	Q9CWJ9	22.968825	21.688816	22.8492828	20.832515	
367	Q561N5	0.00336011	Rps18;Gm102	40S ribosomal protein S18	Q561N5	26.273443	25.816558	26.268165	25.780468	
376	Q6RI64	0.003425505	Psmb1	Proteasome subunit beta type;Proteasome sub	Q6RI64	22.683617	20.709541	22.5514999	19.830151	
1073	Q8VEH3	0.003533771	Arl8a	ADP-ribosylation factor-like protein 8A	Q8VEH3	23.22738	21.247923	23.1311948	20.550788	
1490	Q9DBS1	0.003533771	Tmem43	Transmembrane protein 43	Q9DBS1	22.250616	20.593382	22.0917393	18.740045	
604	P05201	0.003636625	Got1	Aspartate aminotransferase, cytoplasmic	P05201	20.858573	19.001068	20.5122018	0	
744	A0A0G2JGT5	0.003636625	Eif4e	Eukaryotic translation initiation factor 4E	A0A0G2JGT5	22.728263	19.879688	22.6751675	18.842866	

1570	Q3TD51	0.003636625	Picalm;Snap9	Phosphatidylinositol-binding clathrin assembly p	Q3TD51	21.696374	19.361029	21.6220619	0	
1800	O55108	0.003636625	Bsg	Basigin	O55108	21.807517	19.331897	21.6377091	0	
1356	Q5SF07	0.003765819	Igf2bp2	Insulin-like growth factor 2 mRNA-binding protei	Q5SF07	22.684195	21.410816	22.5911221	20.393259	
1945	P63168	0.003824005	Dynll1	Dynein light chain 1, cytoplasmic	P63168	23.968415	23.046953	23.9203161	22.849476	
840	Q8R1B4	0.003855052	Eif3c;G730046	Eukaryotic translation initiation factor 3 subunit	Q8R1B4	23.555509	22.405353	23.3815555	21.610046	
1546	Q8C180	0.003849924	Frs2;Frs3	Fibroblast growth factor receptor substrate 2;Fil	Q8C180	21.164069	18.858396	21.0522646	0	
860	Q5NTY0	0.003885006	Dnaja1	DnaJ homolog subfamily A member 1	Q5NTY0	22.605542	21.17549	22.4948195	20.335479	
1499	Q3U6V5	0.003885006	Etf1	Eukaryotic peptide chain release factor subunit	Q3U6V5	21.745265	20.370846	21.5125648	17.646809	
872	H7BX95	0.003905478	Srsf1	Serine/arginine-rich splicing factor 1	H7BX95	23.673553	21.134861	23.6072474	19.426706	
43	Q8C4D5	0.003923612	Rftn1	Raftlin	Q8C4D5	21.549375	20.088737	21.2442095	18.040016	
1112	Q8CHA4	0.003923612	Sorcs2	VPS10 domain-containing receptor SorCS2	Q8CHA4	21.620416	24.830487	21.416734	24.776644	+
1002	A0A087WPL5	0.004231441	Dhx9	ATP-dependent RNA helicase A	A0A087WPL5	20.627987	19.691902	20.4077664	0	
324	Q99JI4	0.004291511	Psmd6	26S proteasome non-ATPase regulatory subuni	Q99JI4	22.935786	20.851938	22.8524325	19.124914	
829	Q3U935	0.004318364	Cnbp	Cellular nucleic acid-binding protein	Q3U935	23.089258	21.150494	23.033054	20.265655	
809	A0A0G2JDV8	0.004430361	Cnn3	Calponin-3	A0A0G2JDV8	21.932571	20.070232	21.8542084	0	
342	Q3UBB0	0.004513053	Gns	N-acetylglucosamine-6-sulfatase	Q3UBB0	21.921566	19.741167	21.767771	18.678635	
415	B1AU25	0.004513053	Aifm1;Pdcd8	Apoptosis-inducing factor 1, mitochondrial	B1AU25	22.073984	19.644538	21.9780809	0	
618	Q4FJQ0	0.0045185	Rab7;Rab7a	Ras-related protein Rab-7a	Q4FJQ0	25.006912	23.160963	24.9515478	22.861212	
780	P68510	0.004513053	Ywhah	14-3-3 protein eta	P68510	22.624742	20.556878	22.5881549	19.597469	
1807	E9PYX7	0.004513053	Mllt4	Afadin	E9PYX7	20.676735	19.368759	20.3082424	0	
1964	A0A0A6YX73	0.004513053	Prkar2a	cAMP-dependent protein kinase type II-alpha r	A0A0A6YX73	21.495537	19.800545	21.3465213	17.936109	
1967	E9Q6K3	0.004513053	Pibf1		E9Q6K3	23.007541	19.875394	22.9678579	18.36676	++
108	B9EHN0	0.004555541	Uba1	Ubiquitin-like modifier-activating enzyme 1	B9EHN0	24.142359	22.951571	24.0181037	22.26402	
287	Q8BR92	0.004662005	Palm2	Paralemmin-2	Q8BR92	21.852439	19.194728	21.8299679	0	
545	Q3TDA7	0.004662005	Pacsin2	Protein kinase C and casein kinase substrate in n	Q3TDA7	21.402566	19.402986	21.2147116	18.289435	
574	A0A0G2JEP0	0.004623781	Fxr1	Fragile X mental retardation syndrome-related p	A0A0G2JEP0	24.527145	23.513824	24.4621854	23.235955	
1320	E9Q5G3	0.004662005	Kif23	Kinesin-like protein KIF23;Kinesin-like protein	E9Q5G3	20.626125	18.80199	20.5727353	0	
1562	Q9CQD1	0.004662005	Rab5a	Ras-related protein Rab-5A	Q9CQD1	20.620456	18.895631	20.5292306	0	
2121	Q3THA6	0.004662005	Srsf7	Serine/arginine-rich splicing factor 7	Q3THA6	21.598738	20.006117	21.5692738	19.594283	
82	Q5RL55	0.004882902	Abcf1	ATP-binding cassette sub-family F member 1	Q5RL55	22.366093	21.123656	22.1400792	19.752555	
1117	Q3U1P2	0.004948138	Snx9	Sorting nexin;Sorting nexin-9	Q3U1P2	23.183943	21.611552	23.0546607	20.641489	
1284	A1L0U3	0.005113346	Hist1h3e;Hist2	Histone H3;Histone H3.2;Histone H3.3;Histone	A1L0U3	22.479819	19.875024	22.4264494	0	
200	Q9DB05	0.005166141	Napa;Napb	Alpha-soluble NSF attachment protein;Beta-solu	Q9DB05	25.219212	21.134277	25.1877488	20.073789	
600	Q3UE40	0.005351034	Gna13	Guanine nucleotide-binding protein subunit alpH	Q3UE40	20.790561	18.764544	20.667623	0	
927	Q8BK51	0.005351034	Pmpca	Mitochondrial-processing peptidase subunit alp	Q8BK51	21.169398	18.934861	21.0800676	0	
281	Q3U630	0.005530342	Tars	Threonine--tRNA ligase, cytoplasmic	Q3U630	23.203724	20.671143	23.109001	0	
877	Q9CXP8	0.00555959	Gng10;Gm205	Guanine nucleotide-binding protein G(I)/G(S)/G	Q9CXP8	22.879604	20.054412	22.8433468	18.724162	
1344	Q9CY93	0.005558238	Rpl31	60S ribosomal protein L31	Q9CY93	24.270241	21.980951	24.236691	21.485206	
1510	F8WIV2	0.00555959	Serpinb6a;Ser	Serpin B6	F8WIV2	24.217889	22.633798	24.1962278	22.448624	+
1545	Q8BVV6	0.00555959	Eif5	Eukaryotic translation initiation factor 5	Q8BVV6	21.587176	19.83593	21.408437	17.817805	
997	Q3V2D0	0.005598932	Lonp1	Lon protease homolog, mitochondrial	Q3V2D0	21.935089	19.97015	21.7805424	0	
1264	D3YU22	0.005818297	Limch1	LIM and calponin homology domains-containing	D3YU22	19.918664	22.159363	0	22.015991	
1506	Q8BJX0	0.005818297	Psmb2	Proteasome subunit beta type;Proteasome sub	Q8BJX0	22.042903	20.849935	21.8688288	20.101264	
1487	Q5M8Q0	0.00591703	Rpl15;Gm100	Ribosomal protein L15;60S ribosomal protein L1	Q5M8Q0	24.99398	23.730104	24.9666959	23.626179	
537	Q9QXT0	0.00608788	Cnpy2	Protein canopy homolog 2	Q9QXT0	21.913164	19.816922	21.7911057	18.462393	
1136	Q91VM5	0.00608788	Rbmxl1;Rbmx	RNA binding motif protein, X-linked-like-1;RNA-b	Q91VM5	22.320186	19.818126	22.3063667	0	
1331	Q6ZWN5	0.006098167	Rps9	40S ribosomal protein S9	Q6ZWN5	24.491452	23.166302	24.4301708	22.837634	
918	Q811J2	0.006267488	LOC72520		Q811J2	22.81038	20.408752	22.7506396	18.710121	
41	O55029	0.006311811	Copb2	Coatomer subunit beta	O55029	22.62332	21.875698	22.346523	20.925615	
466	Q99LP6	0.006606155	Grpel1	GrpE protein homolog 1, mitochondrial;GrpE pro	Q99LP6	22.907669	20.512476	22.8517979	19.241834	
1053	O08547	0.006606155	Sec22b	Vesicle-trafficking protein SEC22b	O08547	21.296618	20.496115	21.0580455	19.498855	
1325	Q497I3	0.006606155	Fabp5	Fatty acid-binding protein, epidermal	Q497I3	23.13777	19.945947	23.1013243	0	
1429	Q5M8R9	0.006606155	Fdps	Farnesyl pyrophosphate synthase	Q5M8R9	23.069995	20.400029	23.0119856	18.425551	
357	P58871	0.006628359	Tnks1bp1	182 kDa tankyrase-1-binding protein	P58871	23.950034	22.920089	23.7651646	21.924524	
1050	B2RT97	0.006715009	Psmd13	26S proteasome non-ATPase regulatory subuni	B2RT97	22.258181	20.611044	22.1457392	19.158922	
314	P62627	0.006811737	Dynlrb1	Dynein light chain roadblock-type 1	P62627	21.703019	20.088352	21.6270133	19.168211	
397	Q3UZH2	0.006811737	Mcm3	DNA helicase;DNA replication licensing factor M	Q3UZH2	21.178432	19.650252	20.8814293	0	
645	Q3TDM8	0.006811737	Scamp3;Tu52	Secretory carrier-associated membrane protein	Q3TDM8	20.100311	19.380961	19.606464	0	
1387	A0A0A6YX18	0.006818061	Atp6v1h	V-type proton ATPase subunit H	A0A0A6YX18	22.529942	21.328115	22.3940835	18.570853	
2	Q3TLR3	0.006993007	Trim32	E3 ubiquitin-protein ligase TRIM32	Q3TLR3	21.000869	18.84584	20.9448592	0	
70	Q9DB15	0.006993007	Mrpl12	39S ribosomal protein L12, mitochondrial	Q9DB15	21.293126	18.624977	21.1094808	0	
633	B2RQE8	0.006993007	Arhgap42	Rho GTPase-activating protein 42	B2RQE8	20.284309	18.614008	20.004791	0	
640	E9Q226	0.006993007	Ccnyl1		E9Q226	20.706449	18.481945	20.5360785	0	
774	Q3UJN2	0.006993007	Ruvbl1	RuvB-like 1	Q3UJN2	20.451136	18.748281	20.394389	0	
808	E9QMV2	0.006993007	Abracl	Costars family protein ABRACL	E9QMV2	21.183832	19.474251	20.9807649	18.662929	
1058	Q5NCB5	0.006993007	Irgm1	Immunity-related GTPase family M protein 1	Q5NCB5	20.789538	19.319026	20.7472438	17.612322	
1127	Q5M9P1	0.006993007	Rpl36a;Gm65	60S ribosomal protein L36a	Q5M9P1	23.593449	22.311556	23.5852538	22.19696	
1322	Q8BTA2	0.006993007	Pafah1b2	Platelet-activating factor acetylhydrolase IB sub	Q8BTA2	20.161761	18.34247	19.9261922	0	
1531	G3UZA4	0.006993007	Csnk2b;Csnk2	Casein kinase II subunit beta	G3UZA4	21.712447	19.566737	21.641292	18.286308	
1862	Q3TQY2	0.006993007	Dctn4	Dynactin subunit 4	Q3TQY2	20.786362	19.447606	20.657619	0	
2007	Q3UKC8	0.006993007	Gng5	Guanine nucleotide-binding protein subunit gam	Q3UKC8	22.360289	19.089533	22.3463195	0	
677	P58022-2	0.007038765	Loxl2	Lysyl oxidase homolog 2	P58022-2	22.367158	20.353836	22.2626452	0	
1709	Q9CQX5	0.007247285	Cldnd1	Claudin domain-containing protein 1	Q9CQX5	22.019291	19.753101	21.8968462	0	
938	Q543B9	0.007485168	Prep	Prolyl endopeptidase	Q543B9	22.009886	19.86127	21.8231093	0	
476	Q3UIH5	0.007995382	Psmc2	26S protease regulatory subunit 7	Q3UIH5	22.583209	21.256152	22.4629198	20.727381	
60	Q05BN2	0.008025015	Pa2g4	Proliferation-associated protein 2G4	Q05BN2	24.277724	22.539143	24.185087	22.042175	
800	Q9CYD3	0.008056393	Crtap	Cartilage-associated protein	Q9CYD3	23.436992	20.888736	23.3269038	18.915752	
1444	Q9D6G1	0.008230024	Hnrnpab	Heterogeneous nuclear ribonucleoprotein A/B	Q9D6G1	24.660642	22.44877	24.6449534	22.171066	
804	O54962	0.008293161	Banf1	Barrier-to-autointegration factor;Barrier-to-auto	O54962	22.952307	21.278294	22.9469973	20.971585	
1642	Q9JJ44	0.008293161	Dut		Q9JJ44	21.576896	19.671999	21.4723126	0	
723	Q8BX10-2	0.00846349	Pgam5	Serine/threonine-protein phosphatase PGAM5,	Q8BX10-2	23.518042	22.454599	23.5027157	22.348149	
928	P10852	0.00846349	Slc3a2	4F2 cell-surface antigen heavy chain	P10852	23.072251	20.623205	23.0097817	19.428877	
1405	E9Q1W0	0.00846349	Camk2d	Calcium/calmodulin-dependent protein kinase t	E9Q1W0	23.235682	22.084075	23.1784501	21.876761	
1769	Q3TIC8	0.00846349	Uqcrc1	Cytochrome b-c1 complex subunit 1, mitochond	Q3TIC8	22.56498	20.885507	22.4719522	20.192545	
463	Q3TKM9	0.008544372	Arpc5	Actin-related protein 2/3 complex subunit 5	Q3TKM9	25.004505	24.005119	24.9868618	23.895052	
1260	Q9D0M3-2	0.008601105	Cyc1	Cytochrome c1, heme protein, mitochondrial	Q9D0M3-2	21.002176	19.407007	20.8146958	0	
1221	Q3UA17	0.008651039	Mtch2	Mitochondrial carrier homolog 2	Q3UA17	23.293386	21.882106	23.1885041	21.438309	
81	Q3UHT6	0.008714295	Fasn	Fatty acid synthase;[Acyl-carrier-protein] S-ace	Q3UHT6	23.912681	23.009235	23.6867867	22.200191	

464	Q1WWK3	0.008711881	Hist1h1b	Histone H1.5	Q1WWK3	25.122384	23.730524	25.0943556	23.655789	
266	Q6IRU2	0.008893041	Tpm4	Tropomyosin alpha-4 chain	Q6IRU2	28.336394	28.003113	28.3251626	27.982729	
449	B2M1R7	0.009356868	Pcbp2;Pcbp3	Poly(rC)-binding protein 2;Poly(rC)-binding prote	B2M1R7	24.822625	21.856292	24.7803278	21.114718	
1955	Q5BLJ9	0.009483607	Rpl27	60S ribosomal protein L27	Q5BLJ9	24.570242	23.44568	24.5337731	23.349773	
1463	E9PUD2	0.009550533	Dnm1l	Dynamin-1-like protein	E9PUD2	21.28698	19.580531	21.0137498	0	
1915	P14115	0.009550533	Rpl27a	60S ribosomal protein L27a	P14115	24.639501	23.04648	24.6272903	22.862475	
381	F8WIX8	0.010044539	Hist1h2al;Hist	Histone H2A;Histone H2A type 1-H;Histone H2A	F8WIX8	24.920228	22.318484	24.9081159	22.14476	
441	Q8BK46	0.010055744	Psmd3	26S proteasome non-ATPase regulatory subuni	Q8BK46	22.811066	21.193829	22.6591877	19.731628	
1227	Q6ZWZ4	0.010044539	Rpl36;Gm897	60S ribosomal protein L36	Q6ZWZ4	24.144377	22.563543	24.1347715	22.449736	
2107	Q5BLK2	0.010044539	Rps20	40S ribosomal protein S20	Q5BLK2	22.919522	21.190685	22.900545	20.835925	
532	Q99KE1	0.01041181	Me2	NAD-dependent malic enzyme, mitochondrial;M	Q99KE1	21.830128	18.798915	21.7501308	0	
1302	P12787	0.01041181	Cox5a	Cytochrome c oxidase subunit 5A, mitochondria	P12787	22.597332	18.543969	22.5575024	0	
1530	Q8CFE6	0.01041181	Slc38a2	Sodium-coupled neutral amino acid transporter	Q8CFE6	21.910978	18.678399	21.8439504	0	
1537	P99026	0.01041181	Psmb4	Proteasome subunit beta type-4	P99026	21.539658	20.13609	21.516542	19.761454	
1868	E9Q0X0	0.01041181	Psma4	Proteasome subunit alpha type;Proteasome sul	E9Q0X0	21.572525	18.604026	21.5225306	0	
2092	Q8R3X4	0.01041181	Timm44	Mitochondrial import inner membrane transloca	Q8R3X4	20.68579	18.555182	20.5331585	0	
1691	Q9CT23	0.010692097	Eif3e	Eukaryotic translation initiation factor 3 subunit	Q9CT23	23.308177	22.020915	23.2221517	21.253759	
170	Q14AQ1	0.01075553	Psmc6	26S protease regulatory subunit 10B	Q14AQ1	21.975812	20.415558	21.8305881	18.980839	
967	Q3THE2	0.01075553	Myl12b	Myosin regulatory light chain 12B	Q3THE2	24.109668	24.847708	24.0963233	24.815505	
704	Q3KQQ1	0.010951733	Nsfl1c	NSFL1 cofactor p47	Q3KQQ1	22.079704	20.67631	21.8894338	19.705668	
132	Q5RKP0	0.011206767	Vat1	Synaptic vesicle membrane protein VAT-1 hom	Q5RKP0	22.675973	20.589181	22.4986699	18.660771	
547	Q3THU8	0.011537737	Slc25a3	Phosphate carrier protein, mitochondrial	Q3THU8	24.100921	22.447871	24.0783329	22.075073	
212	Q5DTS3	0.012020504	Sep-07	Septin-7	Q5DTS3	22.504767	21.625153	22.4202984	21.090208	
1070	Q3TFN7	0.012004235	Rsu1	Ras suppressor protein 1	Q3TFN7	21.022672	19.443159	20.8629093	0	
527	Q8K4Q8	0.012093977	Colec12	Collectin-12	Q8K4Q8	21.56874	19.746144	21.4044794	0	
866	Q91YN9	0.012093977	Bag2	BAG family molecular chaperone regulator 2	Q91YN9	22.008963	19.448295	21.9018724	0	
765	A0A0A0MQA5	0.012120618	Tuba4a	Tubulin alpha-4A chain	A0A0A0MQA5	22.235666	20.069308	22.154543	0	
207	O70569	0.012842328	rps14;Rps14	40S ribosomal protein S14	O70569	25.372571	24.058944	25.350976	23.989342	
360	Q3U8R9	0.013194383	Txnl1	Thioredoxin-like protein 1	Q3U8R9	22.688564	20.984872	22.5069871	20.078583	
638	Q3U868	0.013194383	Parp1	Poly [ADP-ribose] polymerase 1	Q3U868	21.727265	20.237286	21.4777489	0	
953	Q3UC73	0.013194383	Ppic	Peptidyl-prolyl cis-trans isomerase;Peptidyl-pro	Q3UC73	24.133544	21.331907	24.0916459	20.551936	
502	Q3TS64	0.013561537	Rsl1d1	Ribosomal L1 domain-containing protein 1	Q3TS64	22.132494	20.334308	22.0526938	18.908864	
1478	Q3TJ39	0.014492507	Rab5c	Ras-related protein Rab-5C	Q3TJ39	22.208815	20.900536	22.1376404	20.489295	
1496	Q8BRF7-3	0.014492507	Scfd1	Sec1 family domain-containing protein 1	Q8BRF7-3	20.997164	19.626948	20.6973342	18.347186	
232	E0CY24	0.014763015	Pard3	Partitioning defective 3 homolog	E0CY24	20.420295	18.756453	20.0199889	0	
310	Q9ES97-3	0.014763015	Rtn3	Reticulon-3	Q9ES97-3	21.452901	20.761725	21.3195943	20.237086	
1343	Q3TFD9	0.014763015	Vim		Q3TFD9	24.960148	21.509315	24.9601482	21.387861	
1492	A0A0R4J034	0.014763015	Pdxdc1	Pyridoxal-dependent decarboxylase domain-cor	A0A0R4J034	20.094869	18.69358	19.5062761	0	
1687	Q8BKE0	0.014763015	Psma2	Proteasome subunit alpha type;Proteasome sul	Q8BKE0	21.318284	18.878224	21.1821577	0	
1972	A0A0A6YW80	0.014763015	Naa15;Naa16	N-alpha-acetyltransferase 15, NatA auxiliary sul	A0A0A6YW80	20.074442	18.69706	19.8261473	0	
2088	Q99PV0	0.014763015	Prpf8	Pre-mRNA-processing-splicing factor 8	Q99PV0	20.391495	18.947957	20.1198912	0	
326	Q3UJP8	0.015184754	Gart	Trifunctional purine biosynthetic protein adenos	Q3UJP8	21.632747	20.193844	21.3288237	18.005625	
390	Q91V28	0.015184754	Pgd	6-phosphogluconate dehydrogenase, decarboxy	Q91V28	21.967892	20.47935	21.7618236	18.809205	
536	A4FUS1	0.01627718	Rps16	40S ribosomal protein S16	A4FUS1	25.267366	24.529885	25.2492077	24.472332	
702	F7CBP1	0.016648393	Eif4g2	Eukaryotic translation initiation factor 4 gamma	F7CBP1	21.538393	20.252724	21.3600654	0	
1838	A2RTT4	0.016690044	Ube2n	Ubiquitin-conjugating enzyme E2 N	A2RTT4	22.463218	21.141025	22.3649202	20.563396	
2009	Q3U2J2	0.016742394	Slc2a1	Solute carrier family 2, facilitated glucose transp	Q3U2J2	23.234085	20.803058	23.1889438	19.185559	
340	P20108	0.016970246	Prdx3	Thioredoxin-dependent peroxide reductase, mit	P20108	22.140255	20.373718	22.0762526	19.136814	
626	B7ZP28	0.016970246	Dennd2a	DENN domain-containing protein 2A	B7ZP28	21.595806	20.631375	21.547488	20.058709	
1257	Q99LB4	0.016970246	Capg	Macrophage-capping protein	Q99LB4	21.91026	20.282235	21.8149158	18.638074	
1471	P57759	0.016970246	Erp29	Endoplasmic reticulum resident protein 29	P57759	22.137978	20.476792	22.0130213	18.134074	
1665	Q642K1	0.016970246	Rpl18	60S ribosomal protein L18	Q642K1	25.218754	24.215059	25.2077549	24.189674	
29	Q5EBQ2	0.017271193	Pebp1	Phosphatidylethanolamine-binding protein 1;Hi	Q5EBQ2	22.552494	20.189744	22.485638	18.737541	
32	P52480	0.017265476	Pkm	Pyruvate kinase PKM	P52480	26.53411	25.525053	26.500033	25.395197	
605	Q542I9	0.017271193	Psmc1	26S protease regulatory subunit 4	Q542I9	21.59605	19.997822	21.471452	18.864403	
700	Q5HZY7	0.017271193	Atp6v1g1	V-type proton ATPase subunit G 1	Q5HZY7	23.020636	20.647313	22.9806783	19.943086	
1337	A0A087WQS2	0.017271193	Bzw1	Basic leucine zipper and W2 domain-containing	A0A087WQS2	21.354498	20.046649	21.2812663	18.02869	
1241	Q9EQU5-2	0.017858503	Set;BC085271	Protein SET	Q9EQU5-2	24.895577	22.49537	24.8734056	22.153656	
87	Q8BMF4	0.018044095	Dlat	Dihydrolipoyllysine-residue acetyltransferase co	Q8BMF4	20.034796	18.860214	19.821516	0	
1415	Q3UVI9	0.018044095	Abcf2	ATP-binding cassette sub-family F member 2	Q3UVI9	23.730796	19.603793	23.7085693	0	
1905	Q9QXS1-3	0.018093579	Plec	Plectin	Q9QXS1-3	26.292712	23.455294	26.2821822	23.227026	
358	Q9QZE5	0.018634782	Copg1;Copg2	Coatomer subunit gamma-1;Coatomer subunit	Q9QZE5	22.791543	21.660142	22.5718267	19.743113	
535	B2RSN3	0.018931075	Tubb2b	Tubulin beta-2B chain	B2RSN3	23.074628	21.175148	23.0219465	20.69923	
882	Q8BTQ1	0.018931075	Elavl1	ELAV-like protein;ELAV-like protein 1	Q8BTQ1	23.053105	21.288505	22.9803406	20.826821	
1315	E0CYM1	0.019286711	Filip1l	Filamin A-interacting protein 1-like	E0CYM1	22.620939	21.570379	22.4259386	21.109901	
1017	A0A0A6YVU8	0.020668221	Gm9774;Adrm	Proteasomal ubiquitin receptor ADRM1	A0A0A6YVU8	21.154702	19.878133	21.0757752	19.399063	
1511	Q91WP1	0.020668221	Pvr		Q91WP1	20.678058	18.715326	20.448537	0	
1614	A3KG36	0.020668221	G6pdx;G6pd2	Glucose-6-phosphate 1-dehydrogenase;Glucose	A3KG36	21.0789	18.875128	20.7961246	0	
1689	Q8C2Q8	0.020489203	Atp5c1	ATP synthase subunit gamma;ATP synthase su	Q8C2Q8	22.441348	20.98479	22.3770802	20.512252	
1718	Q60960	0.020668221	Kpna1;Kpna6	Importin subunit alpha-5;Importin subunit alpha	Q60960	20.808672	19.683618	20.7151568	19.132913	
1782	G3UW48	0.020668221	Coro2b	Coronin;Coronin-2B	G3UW48	20.495181	18.85457	20.192277	0	
1792	Q3UAZ7	0.020668221	Hmgb2	High mobility group protein B2	Q3UAZ7	20.68264	18.860977	20.4315927	0	
1857	Q922A2	0.020489203	Anxa7	Annexin;Annexin A7	Q922A2	21.788827	20.327592	21.5539772	19.67699	
1875	A0A0A6YX02	0.020668221	Lamtor1	Ragulator complex protein LAMTOR1	A0A0A6YX02	22.037166	19.001031	21.9956267	0	
1926	Q8R550-5	0.020668221	Sh3kbp1	SH3 domain-containing kinase-binding protein 1	Q8R550-5	20.328847	19.082945	20.01799	0	
95	Q3UHL7	0.021076794	Lamb1	Laminin subunit beta-1	Q3UHL7	22.050932	20.176298	21.9474266	0	
1103	B7ZN02	0.021076794	9430020K01R	Junctional protein associated with coronary arte	B7ZN02	21.677844	20.175885	21.5690018	0	
1304	Q03145	0.021079919	Epha2;Epha6;	Ephrin type-A receptor 2;Receptor protein-tyros	Q03145	21.615012	20.496998	21.4476214	18.710202	
482	Q545L9	0.022718728	Pcmt1	Protein-L-isoaspartate O-methyltransferase;Pro	Q545L9	23.138684	21.61906	23.0353397	20.982111	
1031	Q3TWZ9	0.023431039	Cltb	Clathrin light chain B	Q3TWZ9	22.69665	21.024885	22.6138702	20.245789	
1459	Q3U0T9	0.023431039	Rab35	Ras-related protein Rab-35	Q3U0T9	23.814256	21.287675	23.790119	20.568883	
1713	Q80TX7	0.023845235	mKIAA0567;D	Dynamin-like 120 kDa protein, mitochondrial;Dy	Q80TX7	20.691648	18.744731	20.3613745	0	
853	P47856-2	0.023896551	Gfpt1;Gfpt2	Glutamine--fructose-6-phosphate aminotransfe	P47856-2	22.006469	20.758753	21.7935197	18.943393	
122	Q4FJV4	0.02402801	Anxa1	Annexin;Annexin A1	Q4FJV4	26.339651	25.327627	26.311248	25.242635	
684	Q3U292	0.024184256	Hist1h1d	Histone H1.3	Q3U292	24.722815	23.013053	24.7189342	22.923449	
824	Q3TL79	0.024184256	Ahsa1	Activator of 90 kDa heat shock protein ATPase h	Q3TL79	21.66222	20.042809	21.6034445	18.120341	
1311	Q542I7	0.024184256	Lamtor3	Ragulator complex protein LAMTOR3	Q542I7	21.846462	20.421137	21.7430749	19.631141	
1634	Q3UL43	0.024184256	Nup155	Nuclear pore complex protein Nup155	Q3UL43	20.949872	19.651171	20.6106296	0	

1655	Q545Q2	0.024184256	Surf4	Surfeit locus protein 4	Q545Q2	22.083537	20.617619	20.069433	19.993325	
1028	Q8VHY0	0.024280143	Cspg4	Chondroitin sulfate proteoglycan 4	Q8VHY0	23.413747	22.483609	23.1636764	21.740838	
857	Q5BLK0	0.024748117	Rpl12;Gm165	60S ribosomal protein L12	Q5BLK0	24.858274	23.872406	24.829969	23.755746	
473	Q9D898	0.024916414	Arpc5l	Actin-related protein 2/3 complex subunit 5-like	Q9D898	22.945693	21.071533	22.8973267	20.397349	
301	Q60597	0.025692206	Ogdh	2-oxoglutarate dehydrogenase, mitochondrial	Q60597	21.78574	20.690443	21.3365239	19.455493	
33	Q3U2Z5	0.02599825	Eps8	Epidermal growth factor receptor kinase substra	Q3U2Z5	21.721011	20.396911	21.5360174	19.178255	
699	Q542K4	0.02599825	Cat	Catalase	Q542K4	21.685012	20.364733	21.4033647	18.96644	
1907	Q14AX9	0.026196796	Mrc2	C-type mannose receptor 2	Q14AX9	21.197546	19.695329	21.0226829	0	
1198	Q14BH8	0.027147143	Cacna2d1;Cac	Voltage-dependent calcium channel subunit alp	Q14BH8	22.281445	21.291163	22.0758513	20.626677	
230	Q3UKA1	0.028420698	Dnajc3	DnaJ homolog subfamily C member 3	Q3UKA1	21.599799	19.414922	21.4229815	0	
368	Q6PDI5-2	0.028571429	Ecm29;AI314	Proteasome-associated protein ECM29 homolo	Q6PDI5-2	18.474398	17.682997	18.0978568	0	
646	Q9DCZ0	0.028571429	Atp5d	ATP synthase subunit delta, mitochondrial	Q9DCZ0	22.698738	17.707919	22.6987382	0	
748	Q9CQG4	0.028571429	Reep5	Receptor expression-enhancing protein;Recepto	Q9CQG4	21.379453	18.023701	21.2935577	0	
788	Q921M7	0.028571429	Fam49b	Protein FAM49B	Q921M7	19.397999	17.52139	18.8119181	0	
797	B2RQS1	0.028571429	Strn3	Striatin-3	B2RQS1	17.183294	19.237135	0	18.792603	
836	Q3TX21	0.028571429	Cyr61	Protein CYR61	Q3TX21	18.236979	18.809376	17.6058024	0	
847	Q9CZX7	0.028127428	Tmem55a	Type 2 phosphatidylinositol 4,5-bisphosphate 4-	Q9CZX7	20.994383	18.773666	20.7745652	0	
868	Q5U448	0.028571429	Srsf5	Serine/arginine-rich splicing factor 5	Q5U448	17.173304	19.777253	0	19.555832	
1012	Q3U893	0.028127428	Cpne1	Copine-1	Q3U893	20.288284	18.926922	20.0695899	0	
1101	Q8R5C5	0.028571429	Actr1b	Beta-centractin	Q8R5C5	19.381324	17.783373	18.9493172	0	
1108	A0A0A0U6W1	0.028571429			A0A0A0U6W1	26.20116	24.654843	26.198639	24.64398	
1110	Q549D0	0.028127428	Cd63	Tetraspanin;CD63 antigen	Q549D0	23.408681	21.64884	23.4006433	21.488364	
1178	Q3U029	0.028571429	Cpox	Oxygen-dependent coproporphyrinogen-III oxid	Q3U029	18.942543	17.40615	18.0801876	0	
1185	Q3TG58	0.028127428	Eif4h;mKIAA0	Eukaryotic translation initiation factor 4H	Q3TG58	21.845164	19.82112	21.7731586	19.048378	
1261	Q5SVY2	0.028224594	Ppia;Gm5160	Peptidyl-prolyl cis-trans isomerase;Peptidyl-pro	Q5SVY2	26.474575	25.144896	26.4513741	25.079506	
1303	Q7TNC9	0.028127428	Inpp5a		Q7TNC9	21.620312	19.106071	21.6016186	0	
1335	Q543R4	0.028571429	Cpe	Carboxypeptidase E	Q543R4	18.186417	17.225362	17.7260891	0	
1439	Q3U6E4	0.027896718	Ptma;Gm1250	Prothymosin alpha;Prothymosin alpha, N-termi	Q3U6E4	24.640883	22.224231	24.6071815	21.753084	
1502	Q9WVA2	0.028571429	Timm8a1;Tim	Mitochondrial import inner membrane transloca	Q9WVA2	20.628672	17.242958	20.4438861	0	
1569	P08228	0.028571429	Sod1	Superoxide dismutase [Cu-Zn]	P08228	20.029259	18.037776	19.8729635	0	
1610	Q3TC38	0.028127428	Atp2b1	Plasma membrane calcium-transporting ATPase	Q3TC38	19.860131	19.243087	19.1968957	18.086868	
1646	Q922M3	0.028127428	Kctd10	BTB/POZ domain-containing adapter for CUL3-r	Q922M3	21.970303	20.253708	21.926883	19.704496	
1652	Q3THL1	0.028127428	Tmed2	Transmembrane emp24 domain-containing pro	Q3THL1	21.903064	19.630893	21.8016549	18.7051	
1661	F7D1I5	0.028571429	Gypc	Glycophorin-C	F7D1I5	20.202246	17.917281	20.0286713	0	
1703	A0A0G2JGX3	0.028571429	Atp5f1	ATP synthase F(0) complex subunit B1, mitocho	A0A0G2JGX3	20.911559	17.530311	20.9115589	0	
1742	Q3TXR9	0.028571429	Hexb	Beta-hexosaminidase;Beta-hexosaminidase sut	Q3TXR9	19.862382	17.612738	19.5754755	0	
1779	Q9CQM5	0.028571429	Txndc17	Thioredoxin domain-containing protein 17	Q9CQM5	19.69803	17.927977	19.2803537	0	
1789	Q80XY1	0.028571429	Ppap2a	Lipid phosphate phosphohydrolase 1	Q80XY1	20.438875	17.558432	20.2894839	0	
1818	Q9R0Q6	0.028571429	Arpc1a	Actin-related protein 2/3 complex subunit 1A	Q9R0Q6	20.841947	19.257904	20.8419473	18.858886	
1819	D3Z5B1	0.028127428	Zbed5;Chchd2	Coiled-coil-helix-coiled-coil-helix domain-contain	D3Z5B1	20.737234	18.933186	20.5637219	0	
1914	Q3TG45	0.028571429	Psmd8	26S proteasome non-ATPase regulatory subuni	Q3TG45	18.699126	17.814891	17.8783439	0	
1957	Q3URM4	0.028571429	Erp44	Endoplasmic reticulum resident protein 44	Q3URM4	19.951535	17.696923	19.7011619	0	
2045	Q5SZA3	0.028571429	Hist1h1c	Histone H1.2	Q5SZA3	21.528169	17.321894	21.5042263	0	
2075	Q6DFW4	0.028127428	Nop58	Nucleolar protein 58	Q6DFW4	20.312296	19.109501	19.7661087	0	
2090	F8WIG4	0.028127428	Zpr1	Zinc finger protein ZPR1	F8WIG4	20.281148	18.992835	19.6917957	0	
2150	Q4V9X0	0.028571429	Ppp3ca;Ppp3c	Serine/threonine-protein phosphatase;Serine/t	Q4V9X0	23.087333	21.053485	23.0873329	20.935068	
1305	Q9QZM0	0.028796977	Ubqln2	Ubiquilin-2	Q9QZM0	22.09994	20.050248	21.9233402	0	
293	Q3UZ39	0.029011273	Lrrfip1	Leucine-rich repeat flightless-interacting protein	Q3UZ39	24.826053	24.316416	24.7584289	24.146271	
1151	Q5RKN9	0.030497737	Capza1	F-actin-capping protein subunit alpha-1	Q5RKN9	25.246462	24.63084	25.2189378	24.535941	
499	P52479	0.031842057	Usp10	Ubiquitin carboxyl-terminal hydrolase 10	P52479	21.578671	19.843968	21.3918239	0	
1647	P05132-2	0.031842057	Prkaca	cAMP-dependent protein kinase catalytic subur	P05132-2	22.176727	20.855046	22.0358935	19.788953	
582	Q6NXX6	0.031936275	Atp6v0a1	V-type proton ATPase subunit a;V-type proton A	Q6NXX6	22.667603	21.365594	22.560748	20.067814	
249	F8VPR1	0.032957784	Lmod1	Leiomodin-1	F8VPR1	24.401366	25.738391	24.2279764	25.682898	
336	Q9D1R9	0.033241425	Rpl34;Gm217	60S ribosomal protein L34	Q9D1R9	23.979656	22.852978	23.9687291	22.790803	
1900	Q542M5	0.033241425	Lgals8	Galectin;Galectin-8	Q542M5	21.157292	19.774825	21.0801286	17.739374	
663	P05202	0.033817538	Got2	Aspartate aminotransferase, mitochondrial	P05202	23.215156	21.84012	23.1101362	21.266379	
771	A0A1B0GX25	0.034740091	Hdac6	Histone deacetylase 6	A0A1B0GX25	23.106402	22.707244	22.9001134	22.270868	
883	Q6GT24	0.035129968	Prdx6;Prdx6b	Peroxiredoxin-6	Q6GT24	23.780635	22.045522	23.6639664	21.135022	
1676	Q14AF6	0.035150037	Snrpd2;Gm54	Small nuclear ribonucleoprotein Sm D2	Q14AF6	21.725642	20.687632	21.4772196	19.821956	
697	Q3TW74	0.036039611	Mthfd1	C-1-tetrahydrofolate synthase, cytoplasmic;Me	Q3TW74	22.143439	21.208523	21.9855614	20.435973	
242	Q61699	0.036734541	Hsph1	Heat shock protein 105 kDa	Q61699	23.224205	21.991361	23.071166	20.853142	
297	A5GZX3	0.037917638	Glo1	Lactoylglutathione lyase	A5GZX3	25.857484	19.069472	25.8522232	0	
671	B1AUX2	0.037917638	Hcfc1;Hcfc2	Host cell factor 1;HCF N-terminal chain 1;HCF N-	B1AUX2	20.013041	18.611061	19.905768	0	
790	Q45VK5	0.037917638	Ilf3	Interleukin enhancer-binding factor 3	Q45VK5	20.373697	18.912534	20.0024896	0	
879	Q80UU9	0.037917638	Pgrmc2	Membrane-associated progesterone receptor c	Q80UU9	21.178223	18.99293	21.1064158	0	
1291	Q99JZ4	0.037917638	Sar1a	GTP-binding protein SAR1a	Q99JZ4	21.876671	19.966506	21.8577735	19.352152	
1423	O35841	0.037917638	Api5	Apoptosis inhibitor 5	O35841	20.226796	18.541506	20.0263574	0	
1469	A0A0G2JDZ4	0.037917638	Myl10		A0A0G2JDZ4	25.39242	24.794098	25.3903424	24.780778	
1616	F7CDT0	0.037917638	Ube2m	NEDD8-conjugating enzyme Ubc12	F7CDT0	21.281454	18.999511	21.1698593	0	
1812	Q3UW32	0.037917638	Anp32b	Acidic leucine-rich nuclear phosphoprotein 32 fa	Q3UW32	22.027282	19.793365	21.9556668	19.14089	
1842	Q9CWU3	0.037917638	Ppp2r2a;Ppp2	Serine/threonine-protein phosphatase 2A 55 kD	Q9CWU3	21.246481	19.166736	21.1198208	17.253815	
331	Q3U781	0.038721139	Srsf3	Serine/arginine-rich splicing factor 3	Q3U781	23.782971	21.491207	23.7468102	21.209212	
871	Q8C671	0.038721139	Srsf2	Serine/arginine-rich splicing factor 2	Q8C671	21.50796	20.15598	21.3715317	19.253538	
1115	Q5J7N1	0.038736675	Kras	GTPase KRas;GTPase KRas, N-terminally proces	Q5J7N1	23.25457	21.179632	23.2233253	20.632729	
1128	D3Z7P3-2	0.038721139	Gls	Glutaminase kidney isoform, mitochondrial	D3Z7P3-2	21.50422	20.224709	21.4070402	19.448852	
1266	Q3TRK3	0.038721139	Dbn1	Drebrin	Q3TRK3	22.950863	22.123343	22.895728	21.876039	
1426	Q80XU2	0.038721139	Src	Non-specific protein-tyrosine kinase;Neuronal p	Q80XU2	21.523585	19.585468	21.4698158	17.44452	
1068	Q6ZQ74	0.040175085	Daam1;Daam	Disheveled-associated activator of morphogene	Q6ZQ74	21.695062	20.358807	21.545874	17.800962	
1454	Q80UI9	0.040421346	Wfs1	Wolframin	Q80UI9	21.209651	18.661164	21.1450257	0	
2029	Q99KP6-3	0.040421346	Prpf19	Pre-mRNA-processing factor 19	Q99KP6-3	21.392945	18.859526	21.3461842	0	
512	Q3U7Z6	0.040941496	Pgam1;Pgam	Phosphoglycerate mutase 1;Phosphoglycerate	Q3U7Z6	24.174237	22.731233	24.1110023	22.344529	
739	Z4YJT3	0.042619461	Larp1	La-related protein 1	Z4YJT3	21.786744	20.203156	21.6014869	0	
887	S4R2J9	0.042984633	Prrc2c	Protein PRRC2C	S4R2J9	21.603519	21.265603	21.3506411	20.665577	
1032	Q9ERK4	0.042984633	Cse1l	Exportin-2	Q9ERK4	21.336506	20.551176	20.8563397	18.898991	
1022	Q5SUW3	0.043657636	Grb10	Growth factor receptor-bound protein 10	Q5SUW3	21.21675	22.058034	20.3779757	21.748688	
1608	P57784	0.043923921	Snrpa1	U2 small nuclear ribonucleoprotein A	P57784	24.619873	26.721359	24.5746634	26.701587	
1377	Q6ZWZ7	0.044901995	Rpl17;Gm613	60S ribosomal protein L17	Q6ZWZ7	23.390011	19.525869	23.3414389	0	
1763	Q9DAV6	0.044901995	Serpinb9b		Q9DAV6	20.408497	19.408652	20.0739895	0	

377	Q924W7	0.045447243	St5	Suppression of tumorigenicity 5 protein	Q924W7	22.887031	22.296106	22.6259316	21.64223	
440	A0A0H2UH17	0.045889995	Ubap2l;lig-2a	Ubiquitin-associated protein 2-like	A0A0H2UH17	22.901107	21.816283	22.6972833	20.791326	
690	Q3UK08	0.046814993	Tsg101	Tumor susceptibility gene 101 protein	Q3UK08	21.688302	20.461704	21.5842759	19.57593	
947	Q99JX4	0.046814993	Eif3m	Eukaryotic translation initiation factor 3 subunit	Q99JX4	22.467657	20.944558	22.4327276	20.124642	
145	A0A067XG53	0.048908253	Cask	Peripheral plasma membrane protein CASK	A0A067XG53	22.416876	21.020658	22.3267213	18.689842	
939	Q923G3	0.049140067	Capzb	F-actin-capping protein subunit beta	Q923G3	26.979131	26.093665	26.9581298	26.041695	
856	Q80UN0	0.04988345	Usp14	Ubiquitin carboxyl-terminal hydrolase;Ubiquitin	Q80UN0	21.050353	18.69231	20.8743156	0	
1007	Q497N1	0.04988345	Rps26	40S ribosomal protein S26	Q497N1	22.867845	21.563551	22.8450837	21.395274	
1104	Q8C5L1	0.04988345	Mcm6	DNA helicase;DNA replication licensing factor M	Q8C5L1	20.295943	19.297248	19.9107387	18.027672	
1410	A7VJ98	0.04988345	Gmfb;Gmfg	Glia maturation factor beta;Glia maturation fact	A7VJ98	21.068284	18.867161	20.9522942	0	
1497	Q8BMP9	0.04988345	Stt3a	Dolichyl-diphosphooligosaccharide--protein glyc	Q8BMP9	21.785764	20.009742	21.7314134	19.461642	
1566	Q4V9T8	0.04988345	Eif1;Eif1b	Eukaryotic translation initiation factor 1;Eukaryo	Q4V9T8	22.223128	19.603746	22.1321024	18.654677	
1721	Q3UCT9	0.04988345	Itm2b	Integral membrane protein 2B;BRI2, membrane	Q3UCT9	21.151027	19.034375	21.0807173	0	
1996	Q571A2	0.04988345	Cul2	Cullin-2	Q571A2	19.386744	18.603776	18.9446708	0	
2099	Q0VBA4	0.04988345	Rpl22l1	60S ribosomal protein L22-like 1	Q0VBA4	21.397248	19.192893	21.2959	0	
539	Q61753	0.051014406	Phgdh	D-3-phosphoglycerate dehydrogenase	Q61753	24.321378	23.622888	24.2385456	23.441388	
708	Q3TPZ5	0.051341555	Dctn2	Dynactin subunit 2	Q3TPZ5	22.410908	20.948483	22.3259415	20.041279	
1555	Q50HX3	0.051341555	Rab14	Ras-related protein Rab-14	Q50HX3	22.112697	20.292284	21.9528801	19.19835	
161	Q6P1B9	0.05185702	Bin1	Myc box-dependent-interacting protein 1	Q6P1B9	20.146566	18.467813	20.0300766	0	
843	P10639	0.051864611	Txn	Thioredoxin	P10639	23.732381	20.135959	23.7126628	18.642093	
886	Q6PB52	0.051864611	Urpap1	Alpha-2-macroglobulin receptor-associated prot	Q6PB52	20.54974	19.334212	20.1901264	0	
51	Q3TAH3	0.053046278	Npm3	Nucleoplasmin-3	Q3TAH3	22.334251	19.552661	22.2807874	0	
1188	Q3UPJ2	0.053297186	Impdh2	Inosine-5-monophosphate dehydrogenase;Inos	Q3UPJ2	22.450587	21.248456	22.2844995	20.150991	
8	Q3U8E7	0.057142857	Mthfd2	Bifunctional methylenetetrahydrofolate dehydr	Q3U8E7	20.545938	17.841262	20.5459385	0	
68	Q3V4B3	0.057142857	Vta1	Vacuolar protein sorting-associated protein VTA	Q3V4B3	20.462156	17.831865	20.4621565	0	
71	Q9DBR1-2	0.057142857	Xrn2	5-3 exoribonuclease 2	Q9DBR1-2	18.923214	17.657324	18.7010617	0	
104	B2RUC7	0.055888439	Strap	Serine-threonine kinase receptor-associated pro	B2RUC7	22.338732	20.827883	22.1681446	19.199599	
492	Q5ND51	0.057142857	Crk	Adapter molecule crk	Q5ND51	18.475313	17.423682	17.9094748	0	
658	P50518	0.057142857	Atp6v1e1	V-type proton ATPase subunit E 1	P50518	20.935087	17.659923	20.8157496	0	
754	Q3UQM7	0.056216317	Slc7a5	Large neutral amino acids transporter small sub	Q3UQM7	22.389257	20.179668	22.23412	17.755524	
763	P40142	0.055809697	Tkt	Transketolase	P40142	22.564311	21.885816	22.3394369	21.25136	
1035	Q3TRI2	0.057142857	Dab2	Disabled homolog 2	Q3TRI2	20.053552	18.030367	20.0535518	0	
1192	Q3U8N1	0.057142857	Sars	Serine--tRNA ligase, cytoplasmic	Q3U8N1	21.086498	18.191093	21.0271878	0	
1235	Q6PDC2	0.057142857	Tmed9	Transmembrane emp24 domain-containing pro	Q6PDC2	21.292385	17.76854	21.2923845	0	
1269	P70699	0.057142857	Gaa	Lysosomal alpha-glucosidase	P70699	18.901713	17.46716	18.3383652	0	
1318	G5E898	0.057142857	Ppl	Periplakin	G5E898	19.746836	22.014918	19.1801696	21.904527	
1342	Q8BJY1	0.057142857	Psmd5	26S proteasome non-ATPase regulatory subuni	Q8BJY1	18.850675	17.967316	18.6221903	0	
1349	B2RQH0	0.056216317	Dsg1b;Dsg1a;	Desmoglein-1-alpha;Desmoglein-1-beta;Desmo	B2RQH0	22.35963	24.644065	22.1710128	24.625184	
1435	Q3U8F7	0.057142857	Npc2	Epididymal secretory protein E1	Q3U8F7	19.891767	17.385583	19.6986619	0	
1438	Q8K2A1	0.057142857	Gulp1	PTB domain-containing engulfment adapter pro	Q8K2A1	19.161629	17.90556	18.7682934	0	
1458	Q9QZB7	0.057142857	Actr10	Actin-related protein 10	Q9QZB7	20.344545	18.443135	20.2317727	0	
1473	Q9D6K8	0.057142857	Fundc2	FUN14 domain-containing protein 2	Q9D6K8	21.085645	19.619993	20.9741187	19.535918	
1485	D3YWF6	0.057142857	Otub1	Ubiquitin thioesterase OTUB1	D3YWF6	17.545799	18.459234	0	17.676125	
1544	D3Z789	0.057142857	Snx3	Sorting nexin-3	D3Z789	19.701427	18.077421	19.1970387	0	
1674	A0A0G2JEA9	0.057142857	RBM8;Rbm8a	RNA-binding protein 8A	A0A0G2JEA9	21.447679	17.542608	21.4209974	0	
1699	Q1A602	0.057142857	Actn4		Q1A602	23.186825	21.183046	23.1868251	21.146879	
1707	B2RRH9	0.057142857	Gmps	GMP synthase [glutamine-hydrolyzing]	B2RRH9	17.922356	19.919496	0	19.472738	
2000	D3Z563	0.057142857	Dbi	Acyl-CoA-binding protein	D3Z563	20.879417	17.676778	20.8143145	0	
2023	Q505A8	0.057142857	Rpl39	60S ribosomal protein L39	Q505A8	21.521041	17.886875	21.4897071	0	
2043	Q8C5F9	0.057142857	Serinc1	Serine incorporator 1	Q8C5F9	20.444045	18.15152	20.3360144	0	
245	Q8VDZ4-2	0.059657061	Zdhhc5	Palmitoyltransferase ZDHHC5	Q8VDZ4-2	21.641937	20.15466	21.4784483	18.368125	
753	Q3UHW2	0.059657061	Hsd17b4	Peroxisomal multifunctional enzyme type 2;(3R	Q3UHW2	23.79758	19.361219	23.7738645	0	
775	E9Q1S3	0.059657061	Sec23a	Protein transport protein Sec23A	E9Q1S3	20.735628	18.955422	20.5091301	16.428126	
1336	D3YV69	0.059657061	Rab6a;Rab6b;	Ras-related protein Rab-6A;Ras-related protein	D3YV69	23.978958	22.594098	23.961153	22.446573	
1696	Q6P5I1	0.059657061	Sdf2	Stromal cell-derived factor 2	Q6P5I1	22.210094	23.411799	22.120388	23.392593	
447	A0A0H3XWX3	0.061458212	Igf2bp1	Insulin-like growth factor 2 mRNA-binding prote	A0A0H3XWX3	22.459687	20.441968	22.3786159	19.021791	
1075	Q53YX2	0.062561084	Thy1	Thy-1 membrane glycoprotein	Q53YX2	26.427872	26.01994	26.4176803	26.010455	
636	Q8VDP3-3	0.064953837	Mical1	Protein-methionine sulfoxide oxidase MICAL1	Q8VDP3-3	21.954879	20.945298	21.4470013	19.480448	
728	Q8BIF9	0.064957265	Chmp2b	Charged multivesicular body protein 2b	Q8BIF9	21.846115	19.659181	21.728539	18.750842	
755	Q8K0C4	0.064957265	Cyp51a1	Lanosterol 14-alpha demethylase	Q8K0C4	20.623511	18.903077	20.4820246	0	
1042	Q99JW7	0.064957265	Cdk1	Cyclin-dependent kinase 1	Q99JW7	20.144278	18.953936	19.7626863	0	
1074	Q3UK39	0.064957265	Mcm2	DNA helicase;DNA replication licensing factor M	Q3UK39	20.374426	18.770354	20.2002525	0	
1083	Q4FJK0	0.064957265	Decr1	2,4-dienoyl-CoA reductase, mitochondrial	Q4FJK0	19.899859	18.598111	19.4953486	0	
1228	E0CXB9	0.064957265	Ctnna2	Catenin alpha-2	E0CXB9	21.896427	19.905081	21.8528222	18.927454	
1622	Q8BFZ3	0.06461309	Actbl2	Beta-actin-like protein 2	Q8BFZ3	27.027527	26.193544	27.0220228	26.181917	
1636	Q14BR4	0.064957265	Arf4	ADP-ribosylation factor 4	Q14BR4	21.711439	23.391677	21.5288369	23.380172	
1677	Q4FK57	0.064957265	Ero1l;Ero1lb	ERO1-like protein alpha;ERO1-like protein beta	Q4FK57	20.903176	18.630299	20.6624777	0	
1726	Q3U878	0.064957265	Dazap1	DAZ-associated protein 1	Q3U878	21.639027	18.695673	21.5681783	0	
1705	P46460	0.065882764	Nsf	Vesicle-fusing ATPase	P46460	20.785287	18.868188	20.714028	0	
113	Q5SUT0	0.067084794	Ewsr1	RNA-binding protein EWS	Q5SUT0	21.984659	20.524464	21.7058475	19.512401	
561	A0A087WQE6	0.067501145	Tceb1	Transcription elongation factor B polypeptide 1	A0A087WQE6	22.3445	21.063688	22.186952	20.303539	
984	Q8C1L7	0.067501145	mCG_6739;Rp	40S ribosomal protein S21	Q8C1L7	24.002321	22.279704	23.9638601	22.010176	
213	Q4FZH7	0.068361441	Erh	Enhancer of rudimentary homolog	Q4FZH7	23.082998	21.289779	23.0396633	20.987256	
764	Q9DCD8	0.068361441	Psma3	Proteasome subunit alpha type;Proteasome sul	Q9DCD8	21.830571	20.746145	21.6986496	20.238894	
1476	A0A0R4J150	0.06890398	Osbpl8;mKIA(	Oxysterol-binding protein;Oxysterol-binding pro	A0A0R4J150	21.021273	19.725677	20.7802052	0	
19	A0A0J9YTY0	0.07095965	Sept11;Sept6	Septin-11;Septin-6	A0A0J9YTY0	22.470863	22.249699	22.2327732	21.961389	
386	Q5M9N8	0.071215863	Rpl7	60S ribosomal protein L7	Q5M9N8	25.736756	24.630526	25.6927666	24.504448	
819	Q9QZ83	0.071776753	Actg1		Q9QZ83	29.48251	27.811385	29.4821233	27.808575	
1765	Q8CBE6	0.073391608	Dag1	Dystroglycan;Alpha-dystroglycan;Beta-dystrog	Q8CBE6	21.57944	20.362607	21.4299455	18.953837	
1268	Q3TJ01	0.076465408	Rtcb	tRNA-splicing ligase RtcB homolog	Q3TJ01	24.320289	23.782516	24.2180194	23.592366	
597	Q3UJR8	0.078024345	Btf3	Transcription factor BTF3	Q3UJR8	21.538836	19.809816	21.3387947	18.416526	
1207	Q91VZ1	0.078187428	Snx2;Snx1	Sorting nexin-2;Sorting nexin-1	Q91VZ1	21.452679	20.342821	21.2893463	18.98107	
1374	Q91YU6	0.078024345	Lzts2	Leucine zipper putative tumor suppressor 2	Q91YU6	21.054737	19.938724	20.8070881	18.243835	
1443	Q3TQP7	0.078187428	Acat1	Acetyl-CoA acetyltransferase, mitochondrial	Q3TQP7	21.43226	20.163341	21.3247205	18.532289	
1413	P48036	0.07864961	Anxa5	Annexin A5	P48036	25.64825	23.924464	25.6117978	23.745927	
937	P68040	0.078926017	Gnb2l1	Guanine nucleotide-binding protein subunit beta	P68040	24.71887	23.340829	24.6528947	23.008845	
1087	Q60749	0.079567523	Khdrbs1;Khdr	KH domain-containing, RNA-binding, signal trans	Q60749	23.266654	20.870454	23.1967655	20.10028	
410	P62717	0.080976306	Rpl18a;Gm17	60S ribosomal protein L18a	P62717	24.090513	22.819548	24.0419583	22.591332	

146	P63024	0.082983683	Vamp3;Vamp	Vesicle-associated membrane protein 3;Vesicle	P63024	21.98406	20.394916	21.9118611	20.113577	
430	P12382	0.082983683	Pfkl	ATP-dependent 6-phosphofructokinase, liver ty	P12382	20.72038	19.002967	20.5306308	0	
673	A0A087WRP7	0.082983683	Ly6c1;Ly6c2	Lymphocyte antigen 6C1;Lymphocyte antigen 6	A0A087WRP7	22.105332	20.089223	21.983322	19.322381	
695	Q3TSX8	0.082983683	Tomm70a	Mitochondrial import receptor subunit TOM70	Q3TSX8	20.693575	18.740961	20.6123588	0	
855	E9Q3Y7	0.082983683	C530008M17	Uncharacterized protein KIAA1211	E9Q3Y7	19.976776	20.037053	19.0938764	19.672992	
1039	A1L3B8	0.082983683	Psmd7	26S proteasome non-ATPase regulatory subuni	A1L3B8	21.053075	19.793947	20.8678717	18.902589	
1057	F7BE32	0.082983683	Rab34	Ras-related protein Rab-34	F7BE32	20.700246	19.448217	20.6668275	18.596715	
1204	H3BKN0	0.082983683	Nsun2	tRNA (cytosine(34)-C(5))-methyltransferase	H3BKN0	20.21989	18.923931	20.1283911	0	
1894	P70460	0.082983683	Vasp	Vasodilator-stimulated phosphoprotein	P70460	20.633084	19.103678	20.5182192	17.168882	
1959	A0A0R4J0R1	0.082983683	Vamp8	Vesicle-associated membrane protein 8	A0A0R4J0R1	20.571113	18.920915	20.409439	0	
416	F8WJK8	0.086460718	St13	Hsc70-interacting protein	F8WJK8	23.272074	20.579183	23.1999772	19.445379	
427	Q52L97	0.086460718	Kpna2	Importin subunit alpha;Importin subunit alpha-1	Q52L97	21.459422	20.205153	21.2292516	18.465284	
989	Q3U9A8	0.086460718	Sh3bgrl	SH3 domain-binding glutamic acid-rich-like prot	Q3U9A8	22.381904	20.730388	22.2285086	19.624453	
1010	Q8R010	0.086460718	Aimp2	Aminoacyl tRNA synthase complex-interacting n	Q8R010	20.821878	20.537693	20.1873091	19.197779	
908	P11031	0.088733786	Sub1	Activated RNA polymerase II transcriptional coa	P11031	21.953009	19.960158	21.835158	18.691072	
1098	Q544Q7	0.088733786	Atp1b3	Sodium/potassium-transporting ATPase subuni	Q544Q7	20.8165	19.536827	20.582614	0	
1348	Q9JJ20	0.088733786	Sfn	14-3-3 protein sigma	Q9JJ20	24.803926	22.405413	24.7896946	22.272694	
1363	A0A140LHF7	0.088733786	Prmt1;Prmt8	Protein arginine N-methyltransferase 1;Protein	A0A140LHF7	21.687705	20.358985	21.6006261	19.449921	
2057	A0A0R4J1C2	0.088733786	Capns1	Calpain small subunit 1	A0A0R4J1C2	21.866662	20.605854	21.7511285	19.972069	
932	D3Z6U8	0.091919028	Fmr1	Fragile X mental retardation protein 1 homolog	D3Z6U8	25.168195	23.85088	25.1195141	23.620515	
452	A2AEX8	0.093046618	Fhl1	Four and a half LIM domains protein 1	A2AEX8	21.837062	22.636438	21.4252728	22.382139	
850	Q6NXX7	0.093481766	Abce1	ATP-binding cassette sub-family E member 1	Q6NXX7	21.509628	20.223111	21.3537177	17.167915	
1097	Q4VAE6	0.097204238	Rhoa;Rhoc;49	Transforming protein RhoA;Rho-related GTP-bir	Q4VAE6	22.928024	21.548903	22.8405732	20.838718	
423	Q9EQ66	0.100530443	ASPH;Asph	Aspartyl/asparaginyl beta-hydroxylase	Q9EQ66	27.820886	28.891152	27.8191315	28.888847	
1367	Q3TZZ7	0.100530443	Esyt2	Extended synaptotagmin-2	Q3TZZ7	20.473603	19.56741	20.0100212	0	
1601	Q3V1C9	0.100530443	Gpc6	Glypican-6;Secreted glypican-6	Q3V1C9	21.634136	19.630914	21.45823	0	
218	Q5EBP9	0.101506067	Trim28	Transcription intermediary factor 1-beta	Q5EBP9	22.409497	21.590569	22.1884586	21.032256	
556	Q3U8Y7	0.101641785	Psmc3	26S protease regulatory subunit 6A	Q3U8Y7	22.615789	20.988406	22.5398859	20.244112	
188	Q56A15	0.104895105	Cycs	Cytochrome c, somatic	Q56A15	20.801641	19.573638	20.7498021	17.524596	
491	Q3TTH9	0.104895105	Pum2	Pumilio homolog 2	Q3TTH9	20.547496	19.044688	20.3646665	0	
806	Q3TG12	0.104895105	Farsb	Phenylalanine--tRNA ligase beta subunit	Q3TG12	20.41756	19.194829	19.8794112	0	
830	Q3TMB5	0.104895105	Trip6	Thyroid receptor-interacting protein 6	Q3TMB5	20.181281	18.862919	19.9430379	0	
922	Q91WB7	0.104895105	Ubtd1;Ubtd2	Ubiquitin domain-containing protein 1;Ubiquitin	Q91WB7	21.421418	19.072914	21.2951966	0	
1242	A1L0U6	0.104895105	Nckap1	Nck-associated protein 1	A1L0U6	20.497509	18.79965	20.386464	0	
1270	F6QA74	0.104895105	Apex1	DNA-(apurinic or apyrimidinic site) lyase;DNA-(a	F6QA74	20.124295	18.710908	19.7291227	0	
1271	Q8CHP5-2	0.104895105	Wibg	Partner of Y14 and mago	Q8CHP5-2	20.455784	19.151036	20.38648	0	
1309	Q5EBI8	0.104895105	Atp5k;Atp5i	ATP synthase subunit e, mitochondrial	Q5EBI8	21.760802	20.519807	21.6909172	20.075836	
1808	Q60692	0.104895105	Psmb6	Proteasome subunit beta type-6	Q60692	21.446717	19.990468	21.4201805	19.348447	
1986	Q99KU0	0.104895105	Vmp1	Vacuole membrane protein 1	Q99KU0	20.021104	18.487233	19.5872228	0	
453	Q3TED3	0.106589635	Acly	ATP-citrate synthase	Q3TED3	23.004939	22.160885	22.7699879	21.303392	
21	Q9CPQ3	0.114285714	Tomm22	Mitochondrial import receptor subunit TOM22 h	Q9CPQ3	20.87603	19.277278	20.8760299	18.889272	
143	Q8BFZ9	0.113502328	Erlin2	Erlin-2	Q8BFZ9	22.726006	21.058668	22.6622129	20.790498	
173	Q91VS8	0.108077074	Farp2	FERM, RhoGEF and pleckstrin domain-containin	Q91VS8	21.975556	20.961777	21.7885661	19.488754	
290	Q69ZW3-2	0.114285714	Ehbp1	EH domain-binding protein 1	Q69ZW3-2	19.279691	17.650065	18.629978	0	
307	D3YX34	0.114285714	Dctn1	Dynactin subunit 1	D3YX34	19.59715	17.612908	19.379605	0	
316	Q9CR68	0.114285714	Uqcrfs1	Cytochrome b-c1 complex subunit Rieske, mitoc	Q9CR68	20.855434	18.349954	20.7011186	0	
347	Q3U541	0.110547711	G3bp2	Ras GTPase-activating protein-binding protein 2	Q3U541	24.718127	23.126938	24.6950834	22.929415	
356	Q3TWG5	0.113502328	Dync1li1	Cytoplasmic dynein 1 light intermediate chain 1	Q3TWG5	21.200558	19.486107	21.0894506	0	
359	Q05BH6	0.114285714	Tcof1	Treacle protein	Q05BH6	17.666091	18.348097	0	17.935337	
421	P70452	0.114285714	Stx4	Syntaxin-4	P70452	17.785249	18.629447	0	17.834664	
445	Q8BSH3	0.114285714	Tpm1		Q8BSH3	20.243064	17.751919	20.1729465	0	
467	Q6ZWR0	0.114285714	Rap2b	Ras-related protein Rap-2b	Q6ZWR0	21.055378	17.888567	21.0185188	0	
497	Q8K297	0.108077074	Colgalt1	Procollagen galactosyltransferase 1	Q8K297	22.855964	21.273271	22.7425135	20.912738	
526	Q3UM23	0.109965793	Rnh1	Ribonuclease inhibitor	Q3UM23	21.30814	21.643845	20.815508	21.361679	
612	A2A7S7	0.114285714	Yars	Tyrosine--tRNA ligase;Tyrosine--tRNA ligase, cy	A2A7S7	20.987824	17.354761	20.8524264	0	
624	Q3UB60	0.113502328	Ppid	Peptidyl-prolyl cis-trans isomerase D	Q3UB60	21.935649	19.52727	21.901162	0	
689	Q8BTX5	0.113502328	Eif3h	Eukaryotic translation initiation factor 3 subunit	Q8BTX5	21.50493	20.710669	21.3059972	20.151955	
691	B2RQS6	0.113502328	Dhx36	ATP-dependent RNA helicase DHX36	B2RQS6	20.981749	19.600807	20.7906633	0	
693	Q8BTJ1	0.108077074	Psat1	Phosphoserine aminotransferase	Q8BTJ1	22.902979	21.586722	22.8299091	21.084856	
734	Q3UYV9	0.114285714	Ncbp1	Nuclear cap-binding protein subunit 1	Q3UYV9	18.762547	17.866954	17.9739187	0	
757	A0A0G2JGL8	0.113502328	Dhx30;mKIAA	Putative ATP-dependent RNA helicase DHX30	A0A0G2JGL8	20.580954	19.797374	20.3853505	16.963022	
762	Q9CRY0	0.114285714	Sec62	Translocation protein SEC62	Q9CRY0	22.789609	20.452336	22.789609	20.18833	
826	P97792-2	0.114285714	Cxadr	Coxsackievirus and adenovirus receptor homolo	P97792-2	18.58857	17.415617	18.0986721	0	
889	Q9ET54-3	0.11048207	Palld	Palladin	Q9ET54-3	21.906848	22.354901	21.561757	22.123819	
895	Q8C5J1	0.114285714	Aldh3a1;Aldh	Aldehyde dehydrogenase;Aldehyde dehydroge	Q8C5J1	17.838703	18.248953	17.4701692	0	
936	Q9DC85	0.114285714	Rpl35a	60S ribosomal protein L35a	Q9DC85	22.450804	20.408614	22.4508038	20.150114	
943	Q99LD8	0.114285714	Ddah2	N(G),N(G)-dimethylarginine dimethylaminohydr	Q99LD8	18.456885	17.43878	17.6977772	0	
952	Q9ERE7	0.114285714	Mesdc2	LDLR chaperone MESD	Q9ERE7	19.425518	18.544589	18.7467505	18.231868	
987	Q7M703	0.114285714	Klc1;Klc4;Klc2	Kinesin light chain 1;Kinesin light chain 4;Kinesin	Q7M703	19.037654	17.886153	18.5970023	0	
1033	E9PWK1	0.109965793	Ephx1	Epoxide hydrolase 1	E9PWK1	21.635353	20.832618	21.5235806	20.072574	
1134	E9Q4K7	0.114285714	Kif13b;Stard9	Kinesin-like protein;Kinesin-like protein KIF1C;K	E9Q4K7	27.112435	18.957695	27.1100668	0	
1162	Q8CEI0	0.114285714	Lyn	Non-specific protein-tyrosine kinase;Tyrosine-p	Q8CEI0	20.867208	17.987988	20.8226182	0	
1184	Q640N1	0.114285714	Aebp1;Cpxm2	Adipocyte enhancer-binding protein 1;Inactive c	Q640N1	18.031057	17.312654	17.6828229	0	
1195	Q3U1W3	0.114285714	Adam9	Disintegrin and metalloproteinase domain-conta	Q3U1W3	18.995133	17.576459	18.4421102	0	
1281	Q8VD75	0.109965793	Hip1	Huntingtin-interacting protein 1	Q8VD75	21.124895	20.081183	20.7804687	0	
1288	Q3UI46	0.114285714	Nedd8	NEDD8	Q3UI46	18.040411	20.103001	0	20.103001	
1329	D3Z3J6	0.114285714	Paip1	Polyadenylate-binding protein-interacting prote	D3Z3J6	18.856752	17.686972	18.2526731	0	
1371	G5E850	0.109965793	Cyb5a	Cytochrome b5	G5E850	21.696553	20.921094	21.5443664	20.223384	
1375	Q8CJ96	0.114285714	Rassf8	Ras association domain-containing protein 8	Q8CJ96	19.036316	17.999309	18.5818144	0	
1400	P62715	0.113502328	Ppp2cb;Ppp2c	Serine/threonine-protein phosphatase 2A catal	P62715	22.457684	21.588878	22.4117776	21.399697	
1440	Q80UJ8	0.109965793	Lactb	Serine beta-lactamase-like protein LACTB, mitoc	Q80UJ8	21.975286	20.665154	21.7724707	19.972035	
1468	H3BKT5	0.114285714	Ahcyl2;Ahcyl1	Adenosylhomocysteinase;Putative adenosylho	H3BKT5	19.307022	18.08056	19.1386653	0	
1508	A0A0U1RQ05	0.10744538	Cyfip1;Cyfip2	Cytoplasmic FMR1-interacting protein 1;Cytopla	A0A0U1RQ05	21.657376	20.995924	21.2259617	19.809255	
1516	Q3TIR6	0.114285714	Vbp1	Prefoldin subunit 3	Q3TIR6	19.430235	17.61588	19.0227516	0	
1564	A0A0R4J0K1	0.114285714	Sep-15	15 kDa selenoprotein	A0A0R4J0K1	19.992705	17.839253	19.9186373	0	
1633	Q6DFU5	0.114254481	Mmp14	Matrix metalloproteinase-14	Q6DFU5	22.443454	21.769801	22.2701545	21.366597	
1662	Q9D7H9	0.114285714	Arl1	ADP-ribosylation factor-like protein 1	Q9D7H9	18.519859	17.855022	18.21999	0	
1814	F8WHT2	0.114285714	Ssh1	Protein phosphatase Slingshot homolog 1	F8WHT2	18.556038	17.672959	17.7250808	0	

1825	Q923F9	0.114285714	Ndufs4	NADH dehydrogenase [ubiquinone] iron-sulfur p	Q923F9	21.052717	17.347059	21.0216684	0	
1833	Q9D2Q8	0.114285714	S100a14	Protein S100-A14	Q9D2Q8	18.279542	19.981475	18.0576802	19.848755	
1834	A2ACH6	0.114285714	Sgcd	Delta-sarcoglycan	A2ACH6	19.71203	17.923922	19.5948318	0	
1836	Q5BKS5	0.114285714	Hook3	Protein Hook homolog 3	Q5BKS5	18.232613	18.331936	17.9872154	0	
1937	Q9CQW9	0.114285714	Ifitm3	Interferon-induced transmembrane protein 3	Q9CQW9	17.681341	20.820706	0	20.686777	
2022	Q9D7P1	0.109965793	Rps24	40S ribosomal protein S24	Q9D7P1	23.396582	22.329918	23.364803	22.142632	
2025	Q4FK79	0.114285714	Ssr3	Translocon-associated protein subunit gamma	Q4FK79	21.492186	19.59659	21.4095018	19.316942	
2076	Q3TU87	0.114285714	Wnk1	Serine/threonine-protein kinase WNK1	Q3TU87	18.466308	18.644166	17.9492165	0	
2082	Q66JR8	0.114285714	Ptms	Parathymosin	Q66JR8	20.893889	17.651305	20.8056205	0	
2086	Q91QS6	0.114285714			Q91QS6	21.439147	17.6481	21.3652025	0	
2089	E9Q4T8	0.114285714	Cul3;mKIAA06	Cullin-3	E9Q4T8	19.074306	17.744721	18.5698248	0	
2128	Q08189	0.114285714	Tgm3	Protein-glutamine gamma-glutamyltransferase	Q08189	22.582	18.0148	22.5675061	0	
2133	Q9DAS3	0.114285714	Dynlt1	Dynein light chain Tctex-type 1	Q9DAS3	19.170636	17.815691	19.1706359	0	
751	Q8VDN2	0.116021774	Atp1a1;Atp1a	Sodium/potassium-transporting ATPase subuni	Q8VDN2	23.108353	22.723747	22.971943	22.411311	
942	P97434	0.118802984	Mprip	Myosin phosphatase Rho-interacting protein	P97434	24.605867	22.91805	24.5681586	22.716109	
1692	A2A547	0.118802984	Rpl19	Ribosomal protein L19;60S ribosomal protein L1	A2A547	24.569646	23.495845	24.5565977	23.431899	
511	Q3UWS9	0.120214723	Prdx5	Peroxiredoxin-5, mitochondrial	Q3UWS9	22.620249	21.895485	22.4451771	21.567541	
175	Q6ZWX6	0.124042844	Eif2s1	Eukaryotic translation initiation factor 2 subunit	Q6ZWX6	23.97811	23.377473	23.9027974	23.10696	
26	Q3TX38	0.125281744	Vdac3	Voltage-dependent anion-selective channel pro	Q3TX38	23.99009	22.186518	23.9248029	21.906456	
1424	A0A087WNN8	0.125939785	Soat1	O-acyltransferase;Sterol O-acyltransferase 1	A0A087WNN8	20.737124	19.517483	20.4711207	0	
976	Q3USG5	0.126875167	Emilin1	EMILIN-1	Q3USG5	21.711548	20.243897	21.5459865	0	
90	Q4FK49	0.12741907	Ppa1	Inorganic pyrophosphatase	Q4FK49	22.011159	21.01418	21.7271921	20.401371	
1019	Q564E2	0.127326562	Ldha;Ldhc	L-lactate dehydrogenase;L-lactate dehydrogena	Q564E2	25.402308	25.59963	25.370495	25.537679	
529	Q4FE56	0.1276879	Usp9x;Usp9y	Ubiquitin carboxyl-terminal hydrolase;Probable	Q4FE56	20.369379	19.924924	20.0076471	0	
1913	D3YZ62	0.1276879	Myo5a		D3YZ62	22.957838	22.211863	22.9353548	22.024037	
707	A0A0A6YXK7	0.128169518	Cep170	Centrosomal protein of 170 kDa	A0A0A6YXK7	21.497838	20.450511	21.430469	18.989059	
548	Q3U1J4	0.128363683	Ddb1	DNA damage-binding protein 1	Q3U1J4	22.710512	22.618078	22.3959742	22.038524	
589	Q8R2Y8	0.13038073	Ptrh2	Peptidyl-tRNA hydrolase 2, mitochondrial	Q8R2Y8	20.473229	18.688761	20.3196457	0	
591	Q8R2Y2-2	0.13038073	Mcam	Cell surface glycoprotein MUC18	Q8R2Y2-2	21.077228	18.739765	20.9300103	0	
789	Q3TQZ8	0.13038073	Slc7a1;SLC7A	High affinity cationic amino acid transporter 1	Q3TQZ8	20.288963	19.149602	19.7914936	0	
1286	B7ZWL1	0.13038073	Cnot1	CCR4-NOT transcription complex subunit 1	B7ZWL1	20.22173	18.725942	20.1237115	0	
1310	Q9CZ04	0.13038073	Cops7a	COP9 signalosome complex subunit 7a	Q9CZ04	20.43974	19.454496	20.1737349	18.70202	
1450	Q99K78	0.13038073	Csnk1g2;Csnk	Casein kinase I isoform gamma-2;Casein kinase	Q99K78	21.067452	19.691042	20.9400278	18.550216	
1517	Q5FWK3	0.13038073	Arhgap1	Rho GTPase-activating protein 1	Q5FWK3	19.819534	19.024346	19.3745971	0	
1667	Q3UKQ5	0.13038073	M6pr	Cation-dependent mannose-6-phosphate recep	Q3UKQ5	20.195363	19.017343	19.8360354	0	
1841	A0A0J9YUQ8	0.13038073			A0A0J9YUQ8	22.722888	21.684854	22.6886681	21.451019	
1948	Q5M9N5	0.13038073	Rpl28	60S ribosomal protein L28	Q5M9N5	23.72866	22.973718	23.7240125	22.918521	
1446	B2RSH3	0.130510373	Cnn1	Calponin;Calponin-1	B2RSH3	21.466816	22.942464	20.7938063	22.836714	
451	Q78ZJ8	0.133264965	Rab11b;Rab11	Ras-related protein Rab-11B;Ras-related protein	Q78ZJ8	22.909029	22.248138	22.8065279	21.928528	
495	Q9CZU6	0.135903357	Cs;Csl	Citrate synthase, mitochondrial;Citrate synthase	Q9CZU6	23.287746	22.449412	23.1808315	22.169667	
1126	O35295	0.138082984	Purb	Transcriptional activator protein Pur-beta	O35295	22.394418	21.037816	22.2638638	20.010834	
343	H3BJP2	0.140401103	Esd	S-formylglutathione hydrolase	H3BJP2	22.292246	21.912897	22.1815977	21.547423	
740	Q3UN82	0.141705031	Pcolce	Procollagen C-endopeptidase enhancer 1	Q3UN82	22.117548	21.620797	21.9310497	21.356431	
555	Q9ESP1	0.143167776	Sdf2l1	Stromal cell-derived factor 2-like protein 1	Q9ESP1	21.497813	21.97137	21.2631177	21.945231	
1119	F6V8M6	0.143167776	Atxn2	Ataxin-2	F6V8M6	20.661097	19.534571	20.2761291	0	
1214	Q9D1M0	0.143167776	Sec13	Protein SEC13 homolog	Q9D1M0	21.723744	19.721657	21.6068121	0	
1488	Q8C6E9	0.143167776	Pura	Transcriptional activator protein Pur-alpha	Q8C6E9	21.566467	20.338094	21.4025944	19.229128	
1892	Q3UF75	0.143167776	Parva	Alpha-parvin	Q3UF75	20.295775	21.104033	19.6533325	20.975535	
2055	Q9WVM1	0.143167776	Racgap1	Rac GTPase-activating protein 1	Q9WVM1	20.42707	19.577529	20.0943861	0	
686	Q4FK36	0.143416831	Dstn	Destrin	Q4FK36	24.069121	23.523607	24.0089372	23.402091	
798	F8VQD7	0.146022578	Ptprg	Protein-tyrosine-phosphatase;Receptor-type ty	F8VQD7	21.829562	20.456523	21.6386097	0	
995	Q60817	0.147193088	Naca	Nascent polypeptide-associated complex subun	Q60817	23.781868	21.250644	23.7193483	20.332511	
1790	Q6ZQI3	0.148558242	Mlec	Malectin	Q6ZQI3	21.302984	21.041076	21.191202	20.491847	
338	Q8VH34	0.14928356	Lamp1	Lysosome-associated membrane glycoprotein 1	Q8VH34	23.568492	21.573405	23.5169154	21.046635	
402	Q3TA75	0.153095027	Fxr2	Fragile X mental retardation syndrome-related p	Q3TA75	22.533998	21.995828	22.2786371	21.596456	
1353	Q8BHE0	0.155949394	Prr11	Proline-rich protein 11	Q8BHE0	20.588385	19.116123	20.36154	0	
1169	I7HJS1	0.157160056	Emd	Emerin	I7HJS1	22.393931	20.655127	22.2944787	19.560713	
83	A0A087WNP6	0.160528361	Cdv3	Protein CDV3	A0A087WNP6	21.536659	18.992214	21.4185656	0	
272	Q0VEI6	0.159611595	Avil	Advillin	Q0VEI6	22.780125	20.150583	22.7067961	0	
460	Q8BTU4	0.160528361	Ssb	Lupus La protein homolog	Q8BTU4	21.887338	20.205653	21.8049934	19.553271	
550	D3Z158	0.160528361	Qars		D3Z158	20.722793	19.103018	20.5401243	0	
647	P70349	0.160528361	Hint1	Histidine triad nucleotide-binding protein 1	P70349	22.505788	19.5424	22.4791561	18.638892	
959	A0A0A6YWH2	0.160528361	Aif1l	Allograft inflammatory factor 1-like	A0A0A6YWH2	23.871873	18.985436	23.8598064	0	
1143	Q8VHG2-2	0.160528361	Amot	Angiomotin	Q8VHG2-2	20.688077	18.988171	20.5371629	0	
1153	P37040	0.160528361	Por	NADPH--cytochrome P450 reductase	P37040	19.673803	19.279624	19.3849789	0	
1347	Q8BXZ1	0.160528361	Tmx3	Protein disulfide-isomerase TMX3	Q8BXZ1	19.90896	18.58117	19.3969086	0	
1351	D3Z6B9	0.159611595	Aldh1l2	Mitochondrial 10-formyltetrahydrofolate dehyd	D3Z6B9	21.266267	19.817363	21.029545	0	
1477	Q4FK22	0.159973019	Ergic1	Endoplasmic reticulum-Golgi intermediate comp	Q4FK22	21.463287	20.748045	21.2450179	19.969124	
1638	A2AFQ0	0.160528361	Huwe1	E3 ubiquitin-protein ligase HUWE1	A2AFQ0	19.088043	19.217238	18.3649812	0	
1671	Q8CJ26	0.160528361	Nradd	Death domain-containing membrane protein NR	Q8CJ26	20.22297	18.886412	19.9881423	0	
1758	Q3TPD9	0.160528361	Maoa	Amine oxidase [flavin-containing] A	Q3TPD9	20.349043	18.981567	20.1788274	0	
1764	Q9D828	0.160528361	Tfg		Q9D828	20.139904	19.074169	19.8959064	0	
1849	Q9WVC3	0.159973019	Cav2	Caveolin-2;Caveolin	Q9WVC3	23.5457	20.757888	23.5243421	19.72505	
1887	Q9CZQ0	0.160528361	Nudt21	Cleavage and polyadenylation specificity factor	Q9CZQ0	20.384853	18.766298	20.2833377	0	
1975	Q3TLW5	0.159611595	Aqp1	Aquaporin-1	Q3TLW5	20.647196	23.085317	19.8982241	23.032283	
2113	Q9D880	0.160528361	Timm50	Mitochondrial import inner membrane transloca	Q9D880	20.779152	19.219374	20.7431287	0	
2153	Q9D7Y7	0.160528361	2210010C04Rik		Q9D7Y7	29.493617	32.429502	29.4935278	32.429468	
214	Q5F2E7	0.162326793	Nufip2	Nuclear fragile X mental retardation-interacting	Q5F2E7	23.156144	22.613704	22.9076008	22.101752	
999	O08553	0.165340019	Dpysl2	Dihydropyrimidinase-related protein 2	O08553	22.550086	21.453602	22.3805632	21.24067	
34	P37804	0.16793758	Tagln	Transgelin	P37804	25.668881	25.086645	25.64234	25.025232	
1045	Q9D1D4	0.171256261	Tmed10	Transmembrane emp24 domain-containing pro	Q9D1D4	21.927694	20.373078	21.7734321	17.872779	
1296	Q3TVS6	0.171256261	Ctsb	Cathepsin B;Cathepsin B light chain;Cathepsin B	Q3TVS6	22.212234	21.045517	22.0777707	20.358036	
648	A4FUJ5	0.172550789	Ccdc102a	Coiled-coil domain-containing protein 102A	A4FUJ5	22.574781	22.175306	22.4258441	21.72908	
784	E9Q035	0.173828612	Gm20425;Srp	Signal recognition particle receptor subunit beta	E9Q035	22.193451	21.198721	21.9460775	20.551047	
839	Q91X98	0.173828612	Vcam1	Vascular cell adhesion protein 1	Q91X98	23.309309	21.559442	23.2417351	20.878406	
309	Q3UKB1	0.176067041	Pls3;Pls1	Plastin-3;Plastin-1	Q3UKB1	23.78508	23.019217	23.6054907	22.591904	
55	Q3U9P7	0.178182028	Oxct1	Succinyl-CoA:3-ketoacid-coenzyme A transfera	Q3U9P7	21.838674	20.521168	21.7696613	19.701495	
72	Q3U5H8	0.178182028	Hmox1	Heme oxygenase 1	Q3U5H8	25.668247	20.740049	25.6539421	19.669347	

606	Q9CWZ5	0.178182028	Ddx21	Nucleolar RNA helicase 2	Q9CWZ5	21.388579	20.645833	21.2300263	20.165871	
1715	Q52KP0	0.178182028	Rpl38	60S ribosomal protein L38	Q52KP0	24.304002	24.328182	24.2944119	24.305174	
1650	Q8C5G6	0.18055693	Tollip	Toll-interacting protein	Q8C5G6	22.589168	21.218463	22.4834564	20.349049	
621	E9QKV6	0.182622171	Myo9b	Unconventional myosin-IXb	E9QKV6	22.563035	22.16101	22.2083334	21.624472	
660	A0A0G2JGD2	0.183506762	S100a4	Protein S100;Protein S100-A4	A0A0G2JGD2	22.439627	23.396435	22.2894572	23.340171	
1251	G3UYV7	0.183506762	Rps28;Gm102	40S ribosomal protein S28	G3UYV7	25.420792	23.93253	25.4158338	23.90034	
792	Q3UUX9	0.185272618	Gdi2;Gdi1	Rab GDP dissociation inhibitor beta;Rab GDP dis	Q3UUX9	22.530047	21.500924	22.2334478	20.923987	
185	F8VPV0	0.188404393	Pcnt	Pericentrin	F8VPV0	22.618359	22.735513	22.1687961	22.264713	
6	Q8BMC5	0.2	Adk	Adenosine kinase	Q8BMC5	19.379247	17.410128	19.0915207	0	
227	B7ZCT0	0.2	Rgs19	Regulator of G-protein signaling 19	B7ZCT0	20.409139	17.668679	20.3584983	0	
248	Q62418-3	0.194871795	Dbnl	Drebrin-like protein	Q62418-3	19.871513	18.835566	19.4621764	0	
282	A0A140LJ79	0.2	Synm	Synemin	A0A140LJ79	18.125314	21.835226	0	21.821363	
288	Q3U9U5	0.194871795	Bcap31	B-cell receptor-associated protein 31	Q3U9U5	22.345582	20.608221	22.3203178	19.910415	
318	Q9CZD3	0.191751319	Gars;F13a1	Glycine--tRNA ligase	Q9CZD3	22.795548	20.790184	22.6539266	19.289035	
353	H3BIW0	0.2	Zc3h18	Zinc finger CCCH domain-containing protein 18	H3BIW0	18.448236	18.122741	17.8265671	0	
407	Q6S393	0.2	Plec	Plectin	Q6S393	20.470482	17.404981	20.3667814	0	
435	Q542M2	0.2	Copz1	Coatomer subunit zeta-1	Q542M2	19.570024	17.519641	19.3004616	0	
469	A0A0A0MQH5	0.2	Macf1		A0A0A0MQH5	20.050387	17.842637	19.9886022	0	
470	Q69ZD1	0.194871795	Exoc4	Exocyst complex component 4	Q69ZD1	18.985551	18.73104	18.3071305	0	
475	Q9JIG8	0.194871795	Praf2	PRA1 family protein 2	Q9JIG8	20.92051	19.621022	20.6636641	0	
530	Q9DBZ5	0.2	Eif3k	Eukaryotic translation initiation factor 3 subunit	Q9DBZ5	21.837707	18.333393	21.7634642	0	
568	Q9D1M4	0.2	Eef1e1	Eukaryotic translation elongation factor 1 epsilo	Q9D1M4	19.91287	18.09404	19.8357614	0	
625	Q9DBG5	0.2	Plin3	Perilipin-3	Q9DBG5	18.622908	17.828242	17.9762002	0	
730	Q9ERR1-2	0.194871795	Ndel1;Nde1	Nuclear distribution protein nudE-like 1;Nuclear	Q9ERR1-2	20.364947	18.836305	20.1260733	0	
813	P68134	0.194871795	Acta1	Actin, alpha skeletal muscle	P68134	18.623496	19.884754	0	19.376625	
848	E9Q5F4	0.2	Actb		E9Q5F4	25.708945	24.513541	25.7089455	24.504237	
885	Q3TC93	0.2	Hs1bp3	HCLS1-binding protein 3	Q3TC93	19.629908	17.853959	19.3395776	0	
906	Q9D8Y1	0.2	Tmem126a	Transmembrane protein 126A	Q9D8Y1	17.773684	19.109778	0	18.904354	
911	P62869	0.197808115	Tceb2	Transcription elongation factor B polypeptide 2	P62869	22.522812	21.694065	22.4832157	21.473443	
913	Q9CQB4	0.2	Uqcrb	Cytochrome b-c1 complex subunit 7	Q9CQB4	21.747981	18.623383	21.7479814	17.883463	
941	D3Z3D2	0.2	Ncbp2	Nuclear cap-binding protein subunit 2	D3Z3D2	18.449044	17.760504	18.2072028	0	
973	Q6P2K2	0.2	Acot2;Acot6;A	Acyl-coenzyme A thioesterase 2, mitochondrial;	Q6P2K2	19.128699	17.680231	19.0305516	0	
1025	P14152	0.2	Mdh1	Malate dehydrogenase, cytoplasmic	P14152	21.205372	18.005719	21.1652891	0	
1082	A2RRJ4	0.194871795	Xpo5	Exportin-5	A2RRJ4	19.89032	19.250465	19.4257428	0	
1091	B1AXN9	0.2	Rps6ka3;Rps6	Ribosomal protein S6 kinase alpha-3;Ribosomal	B1AXN9	18.613107	18.034277	18.0734156	0	
1116	E9PVP1	0.2	Aim1		E9PVP1	17.82762	19.189826	0	18.581265	
1121	Q9D1J2	0.194871795	Cope	Coatomer subunit epsilon	Q9D1J2	20.820333	19.85735	20.6657148	19.137077	
1147	D0ESZ6	0.2	Serpine1	Plasminogen activator inhibitor 1	D0ESZ6	18.913473	19.536907	18.5970357	19.37698	
1163	Q3U9K9	0.2	Trappc3	Trafficking protein particle complex subunit 3	Q3U9K9	19.213044	17.518601	18.7368498	0	
1247	Q5Y5T1-2	0.194871795	Zdhhc20	Probable palmitoyltransferase ZDHHC20	Q5Y5T1-2	20.343985	18.974286	20.1770795	0	
1312	Q91VC3	0.194871795	Eif4a3;Gm895	Eukaryotic initiation factor 4A-III;Eukaryotic init	Q91VC3	20.663341	19.102741	20.4735212	0	
1328	Q3TAM9	0.197808115	Bag3	BAG family molecular chaperone regulator 3	Q3TAM9	20.170009	19.919502	19.5231541	0	
1360	Q58A84	0.2	Scarf2	Scavenger receptor class F member 2	Q58A84	18.790799	18.474857	18.1664148	0	
1432	A0A140LIN9	0.2	Dnah3	Dynein heavy chain 3, axonemal	A0A140LIN9	21.131978	17.548025	21.0528949	0	
1445	B1ARW4	0.2	Ndufs5	NADH dehydrogenase [ubiquinone] iron-sulfur p	B1ARW4	19.811768	17.605384	19.6421923	0	
1491	Q3TH46	0.2	Rdx	Radixin	Q3TH46	19.243223	17.844467	19.1138602	0	
1495	Q545G0	0.194871795	Psmb3;Gm49	Proteasome subunit beta type;Proteasome sub	Q545G0	19.952777	19.724331	19.2166195	19.207849	
1526	G3UYY2	0.2	Selenbp2;Sele	Selenium-binding protein 2;Selenium-binding pr	G3UYY2	20.22814	18.056215	19.9064238	0	
1584	E9QM38	0.2	Slc12a2	Solute carrier family 12 member 2	E9QM38	19.198722	18.165479	18.7676272	0	
1600	P08030	0.2	Aprt	Adenine phosphoribosyltransferase	P08030	19.959828	17.839315	19.8907527	0	
1619	Q3TI40	0.194871795	Mtx2	Metaxin-2	Q3TI40	20.795511	18.785841	20.6459092	0	
1624	Q9Z1N5	0.2	Ddx39b	Spliceosome RNA helicase Ddx39b	Q9Z1N5	19.860771	17.755358	19.6510024	0	
1654	Q3UTI7	0.2	Pin1	Peptidyl-prolyl cis-trans isomerase NIMA-intera	Q3UTI7	18.868855	18.042934	18.868855	0	
1664	Q9CQQ8	0.2	Lsm7	U6 snRNA-associated Sm-like protein LSm7	Q9CQQ8	18.425841	17.290217	17.9913395	0	
1698	Q58E29	0.2	Plp2	Proteolipid protein 2	Q58E29	21.965876	18.990734	21.9393615	18.211046	
1711	A0A0U1RQ95	0.2	Mat2a	S-adenosylmethionine synthase;S-adenosylme	A0A0U1RQ95	17.840133	19.338365	0	19.166825	
1733	P62965	0.2	Crabp1	Cellular retinoic acid-binding protein 1	P62965	18.326184	20.175493	0	20.065281	
1777	Q61735-2	0.2	Cd47	Leukocyte surface antigen CD47	Q61735-2	19.688409	17.844735	19.4289511	0	
1831	Q80YX0	0.2	Tnc	Tenascin	Q80YX0	18.508748	17.99349	17.7292238	17.013381	
1856	Q9CQU3	0.2	Rer1	Protein RER1	Q9CQU3	20.460801	19.355312	20.3283375	19.172017	
1883	O88587-2	0.2	Comt	Catechol O-methyltransferase	O88587-2	19.631348	18.051382	19.3360898	0	
1949	Q05BR0	0.2	Emp1	Epithelial membrane protein 1	Q05BR0	21.351959	19.918447	21.3222532	19.738996	
1983	A0A0U1RPH9	0.2	Prr14	Proline-rich protein 14	A0A0U1RPH9	24.336857	23.256575	24.3368571	23.229702	
1984	Q3TCR4	0.2	Hps3	Hermansky-Pudlak syndrome 3 protein homolo	Q3TCR4	21.190598	18.077004	21.0919299	0	
1987	Q3UBQ4	0.2	Actb		Q3UBQ4	20.700315	23.067045	20.5596661	23.041697	
2033	Q9EQQ2	0.2	Yipf5	Protein YIPF5	Q9EQQ2	20.347477	18.889911	20.3474772	18.288828	
2042	A0A0K6G3H8	0.2			A0A0K6G3H8	18.071525	28.328249	0	28.328096	
2049	Q3UX10	0.2	Tuba3	Tubulin alpha chain-like 3	Q3UX10	19.787544	17.807329	19.7275487	0	
496	Q642K0	0.202741345	Myl6	Myosin light polypeptide 6	Q642K0	29.451383	29.341131	29.4462564	29.335219	
1922	P62075	0.209871322	Timm13	Mitochondrial import inner membrane transloca	P62075	21.727381	20.690266	21.5442473	19.971232	
1598	Q4FJX4	0.215835374	Csrp1	Cysteine and glycine-rich protein 1	Q4FJX4	23.543655	22.807197	23.4260803	22.477342	
1361	P19973-2	0.217867081	Lsp1	Lymphocyte-specific protein 1	P19973-2	23.183708	22.722046	23.1180978	22.412097	
442	Q14AI7	0.218920802	Cops4	COP9 signalosome complex subunit 4	Q14AI7	20.603902	19.528277	20.3597241	0	
971	Q99PC9	0.218920802	Ppp2r5d;ppp2r5d		Q99PC9	20.53386	19.486212	20.102386	0	
1001	Q3V2X2	0.218920802	Kidins220;mKIAA1250		Q3V2X2	20.43934	19.605189	19.803336	0	
1527	Q62371	0.218920802	Ddr2	Discoidin domain-containing receptor 2	Q62371	20.638944	19.670858	20.4066435	17.545145	
1574	Q9CX86	0.218920802	Hnrnpa0	Heterogeneous nuclear ribonucleoprotein A0	Q9CX86	23.638739	22.255324	23.6182094	22.139464	
679	Q3TXH3	0.22599954	Lox	Protein-lysine 6-oxidase	Q3TXH3	23.557208	22.949177	23.441818	22.745355	
623	Q3TJZ6	0.230404426	Fam98a	Protein FAM98A	Q3TJZ6	22.924922	22.376646	22.7992314	22.121735	
3	Q9CX34	0.234498834	Sugt1	Suppressor of G2 allele of SKP1 homolog	Q9CX34	19.682017	18.995731	19.1967829	0	
494	A0A1B0GR60	0.234498834	Ftl1;Ftl2	Ferritin;Ferritin light chain 1;Ferritin light chain 2	A0A1B0GR60	21.981843	20.624008	21.9156611	20.228284	
981	D3YU17	0.234498834	Ncln	Nicalin	D3YU17	19.782688	19.043642	19.6753519	0	
1183	O09131	0.234498834	Gsto1;Gsto2	Glutathione S-transferase omega-1;Glutathione	O09131	20.956731	18.934724	20.7176919	0	
1581	D3YXT0	0.234498834	Ndufs2	NADH dehydrogenase [ubiquinone] iron-sulfur p	D3YXT0	20.348636	19.137521	20.1067538	17.506301	
1621	Q0PD35	0.234498834	Rab21	Ras-related protein Rab-21	Q0PD35	20.542494	19.538605	20.2995723	0	
1971	P05533	0.234498834	Ly6a	Lymphocyte antigen 6A-2/6E-1	P05533	22.599807	21.739794	22.5872948	21.629473	
2074	Q9CWS0	0.234498834	Ddah1	N(G),N(G)-dimethylarginine dimethylaminohydr	Q9CWS0	20.703544	19.831199	20.6586144	19.217193	
267	Q7TQH0-2	0.238963011	Atxn2l	Ataxin-2-like protein	Q7TQH0-2	23.306764	22.613858	23.1235662	22.09669	

189	Q3UDD3	0.241523788	Poldip3	Polymerase delta-interacting protein 3	Q3UDD3	21.427596	20.578613	21.1859341	18.889629	
1129	Q7TSC1	0.244744964	Prrc2a	Protein PRRC2A	Q7TSC1	21.597704	20.985154	21.3685241	20.287172	
915	P09528	0.253379899	Fth1	Ferritin heavy chain;Ferritin heavy chain, N-terr	P09528	22.884396	20.434669	22.7690522	0	
543	F6SQH7	0.258496022	Poldip2	Polymerase delta-interacting protein 2	F6SQH7	23.345013	23.611382	23.249442	23.495184	
1888	Q9ERU9	0.260132568	Ranbp2	E3 SUMO-protein ligase RanBP2	Q9ERU9	20.832107	19.851791	20.7074418	0	
50	Q543H0	0.265669584	Srm	Spermidine synthase	Q543H0	20.929886	19.777912	20.5511689	18.83982	
1504	Q8JZU2	0.265669584	Slc25a1	Tricarboxylate transport protein, mitochondrial	Q8JZU2	22.038292	22.538764	21.9217626	22.442703	
192	Q3UH59	0.266498846	Myh10	Myosin-10	Q3UH59	30.984007	31.120995	30.9732518	31.109126	
747	E9Q7U2	0.270325768	Calcoco1	Calcium-binding and coiled-coil domain-containi	E9Q7U2	21.433748	20.43943	21.1833642	18.647706	
1044	A0A0G2JDW7	0.270325768	Rps27	40S ribosomal protein S27	A0A0G2JDW7	24.345539	23.905614	24.3161835	23.854716	
1100	Q9JIG7	0.269796079	Ccdc22	Coiled-coil domain-containing protein 22	Q9JIG7	19.692159	18.593962	19.3720969	0	
89	Q3TDN2-2	0.278632479	Faf2	FAS-associated factor 2	Q3TDN2-2	20.613525	20.19733	20.4651844	19.621637	
129	Q543N3	0.278632479	Lasp1	LIM and SH3 domain protein 1	Q543N3	22.060631	20.841581	22.0095689	20.552004	
205	Q3TQ74	0.278632479	Ehd4	EH domain-containing protein 4	Q3TQ74	19.777226	19.502826	19.3332357	0	
424	Q3TQP6	0.278632479	Me1	Malic enzyme;NADP-dependent malic enzyme	Q3TQP6	20.743967	18.917989	20.5991842	0	
576	B2RTB0	0.278632479	Pdap1	28 kDa heat- and acid-stable phosphoprotein	B2RTB0	21.428548	20.164611	21.3075837	19.705424	
617	F6YH22	0.278632479	Gm5218;Gm1	60S ribosomal protein L29	F6YH22	23.313754	22.985848	23.2966782	22.954553	
669	O35127	0.278632479	Grcc10	Protein C10	O35127	20.655906	19.438736	20.5291341	18.214708	
768	Q4FJX1	0.278632479	Kpna4	Importin subunit alpha;Importin subunit alpha-3	Q4FJX1	20.563934	19.134693	20.3720442	0	
770	Q91VC9	0.278632479	Ghitm	Growth hormone-inducible transmembrane pro	Q91VC9	20.553685	19.005364	20.3483191	0	
1067	D3YTQ9	0.278632479	Rps15	40S ribosomal protein S15	D3YTQ9	20.974769	19.670334	20.8032733	19.078825	
1467	D3Z061	0.278632479	Uba6	Ubiquitin-like modifier-activating enzyme 6	D3Z061	20.189622	19.027224	19.769592	0	
1585	A0A0J9YUD8	0.278632479	Hmgb1;Hmgl	High mobility group protein B1	A0A0J9YUD8	20.685134	19.166816	20.5286468	0	
1643	Q8BGB5	0.278632479	Limd2	LIM domain-containing protein 2	Q8BGB5	21.746949	21.578063	21.7301504	21.451147	
1670	P26883	0.278632479	Fkbp1a	Peptidyl-prolyl cis-trans isomerase FKBP1A;Pep	P26883	22.074357	20.906039	21.9761203	20.710679	
2048	Q8VDT3	0.278632479	Dclk1	Serine/threonine-protein kinase DCLK1	Q8VDT3	20.387298	19.077216	20.13976	17.395485	
1762	Q3TW21	0.285371559	Ola1	Obg-like ATPase 1	Q3TW21	21.055666	19.468121	20.9484512	0	
805	A2RRJ6	0.291171026	Mib1	E3 ubiquitin-protein ligase MIB1	A2RRJ6	21.785387	21.875308	21.366943	21.52816	
1196	Q9R059	0.291342659	Fhl3	Four and a half LIM domains protein 3	Q9R059	20.415416	20.398394	19.8937108	19.645664	
811	P26443	0.301131854	Glud1	Glutamate dehydrogenase 1, mitochondrial	P26443	22.774023	22.260071	22.6110897	21.747032	
599	Q5NCU4	0.301890183	Sparc	SPARC	Q5NCU4	22.551026	22.130649	22.2436884	21.548093	
750	Q9DB77	0.304518416	Uqcrc2	Cytochrome b-c1 complex subunit 2, mitochond	Q9DB77	22.266092	21.218586	22.1836936	20.96754	
1109	E9PZF0	0.306584368	Gm20390;Nm	Nucleoside diphosphate kinase;Nucleoside diph	E9PZF0	23.267113	22.997212	23.1033232	22.713459	
391	Q543F3	0.310444458	Cnn2	Calponin;Calponin-2	Q543F3	23.504944	22.991996	23.4230274	22.796106	
225	Q5M9K7	0.31858517	Rps10	40S ribosomal protein S10	Q5M9K7	23.741916	22.089118	23.7040794	21.962628	
642	Q8BTU5	0.31858517	Psma1	Proteasome subunit alpha type;Proteasome sul	Q8BTU5	21.293487	19.616625	21.1458124	0	
914	A0A0R4J049	0.31858517	Prmt5	Protein arginine N-methyltransferase 5	A0A0R4J049	20.732091	19.771029	20.3053585	18.336501	
955	Q8CI51	0.317271656	Pdlim5	PDZ and LIM domain protein 5	Q8CI51	21.298783	21.177005	20.9384952	19.658082	
1521	Q3TN44	0.31858517	Cyb5b	Cytochrome b5 type B	Q3TN44	22.994201	21.302626	22.9463521	21.062137	
1579	B2RUJ2	0.31858517	Erbb2ip	Protein LAP2	B2RUJ2	20.71736	19.698665	20.4247902	0	
1840	Q8BG95	0.31858517	Ppp1r12b	Protein phosphatase 1 regulatory subunit 12B;P	Q8BG95	22.031712	20.506853	21.9570815	19.069594	
27	D3Z3Q3	0.331590988	Smtn	Smoothelin	D3Z3Q3	24.077117	23.761974	23.9704361	23.594539	
39	O70591	0.342857143	Pfdn2	Prefoldin subunit 2	O70591	20.036982	18.67918	19.8895815	0	
48	Q8K2B3	0.340418419	Sdha	Succinate dehydrogenase [ubiquinone] flavopro	Q8K2B3	23.824985	24.345035	23.7470999	24.287239	
61	A0A140LHJ4	0.342857143	Stxbp1	Syntaxin-binding protein 1	A0A140LHJ4	19.724129	18.075584	19.7241295	0	
121	P48962	0.337596519	Slc25a4	ADP/ATP translocase 1	P48962	26.015111	26.225577	26.0009967	26.199799	
148	Q3TG07	0.342857143	Gorasp2	Golgi reassembly-stacking protein 2	Q3TG07	17.910278	18.674124	0	18.217945	
302	E9PWQ3	0.328205128	Col6a3		E9PWQ3	19.52686	20.437675	18.8165087	20.193691	
312	A2AW05	0.342857143	Ssrp1	FACT complex subunit SSRP1	A2AW05	19.387952	18.015264	18.9745916	0	
362	Q9DCW4	0.330525182	Etfb	Electron transfer flavoprotein subunit beta	Q9DCW4	22.600047	21.786673	22.463323	21.148537	
373	F7CIP8	0.342857143	Pcyox1	Prenylcysteine oxidase	F7CIP8	18.826022	17.306349	18.5074942	0	
387	Q4VAG4	0.340784852	Rpl22	60S ribosomal protein L22	Q4VAG4	24.253326	24.083888	24.2126816	24.025464	
448	Q3UW40	0.34086786	Rpl24;Gm174	60S ribosomal protein L24	Q3UW40	24.627738	23.356976	24.5866385	23.225238	
465	E9Q221	0.342857143	Slc29a1	Equilibrative nucleoside transporter 1	E9Q221	19.707682	17.907193	19.4929088	0	
571	Q99K70	0.342857143	Rragc;Rragd	Ras-related GTP-binding protein C;Ras-related G	Q99K70	18.475372	20.274944	0	20.178798	
583	Q8C129	0.342857143	Lnpep	Leucyl-cystinyl aminopeptidase	Q8C129	19.226966	17.810722	18.8906736	0	
601	A2A6U5	0.342857143	Sep-09	Septin-9	A2A6U5	18.871448	18.275968	18.0132089	0	
602	Q3UJ34	0.328205128	Ass1;Gm5424	Argininosuccinate synthase	Q3UJ34	20.886217	27.104786	20.4484064	27.100054	
632	Q80YP5	0.328205128	Itga5	Integrin alpha-5;Integrin alpha-5 heavy chain;I	Q80YP5	20.473362	19.661484	20.3301376	18.574264	
737	Q9CQ91	0.342857143	Ndufa3	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q9CQ91	18.679685	18.231233	18.5768146	0	
781	A3KML3	0.321134531	Ywhaq	14-3-3 protein theta	A3KML3	24.959508	24.15408	24.9137881	24.041224	
832	Q8C6Z2	0.342857143	Nid2	Nidogen-2	Q8C6Z2	18.58734	18.132579	18.2306624	0	
861	Q3V3R4	0.342857143	Itga1	Integrin alpha-1	Q3V3R4	18.271962	17.651433	17.6075732	0	
904	Q3TKP3	0.342857143	Atxn10	Ataxin-10	Q3TKP3	18.791278	18.881118	18.221521	18.122187	
926	Q8R3X7	0.337736903	Pck2	Phosphoenolpyruvate carboxykinase [GTP], mit	Q8R3X7	22.266981	21.383619	22.0673081	19.85391	
964	Q8R016	0.342857143	Blmh	Bleomycin hydrolase	Q8R016	18.636719	18.220648	18.1587399	0	
1006	Q3TK27	0.342857143	Gnl3	Guanine nucleotide-binding protein-like 3	Q3TK27	18.932644	18.042379	18.8492582	0	
1014	Q61176	0.328205128	Arg1	Arginase-1	Q61176	19.101773	20.44287	0	20.186873	
1060	F6VQH5	0.342857143	Hnrnpdl	Heterogeneous nuclear ribonucleoprotein D-like	F6VQH5	20.525179	18.126253	20.4239711	0	
1132	A2RS58	0.342857143	Crkl	Crk-like protein	A2RS58	18.27302	17.864293	17.644716	0	
1138	P62835	0.342857143	Rap1a;Gm939	Ras-related protein Rap-1A	P62835	21.276047	18.909019	21.2480866	18.297143	
1193	D3YXX5	0.342857143	Ndufv1	NADH dehydrogenase [ubiquinone] flavoprotein	D3YXX5	18.994525	18.742701	18.0292333	18.351826	
1206	E9Q6Z0	0.342857143	Cul5	Cullin-5	E9Q6Z0	18.88728	20.381137	0	20.095059	
1219	Q3UHM0	0.342857143	Pafah1b1	Platelet-activating factor acetylhydrolase IB sub	Q3UHM0	19.470302	18.749131	19.3052651	18.346137	
1238	D3Z7P2	0.342857143	Tmem109	Transmembrane protein 109	D3Z7P2	21.460184	18.121037	21.3912881	0	
1250	P06837	0.328205128	Gap43	Neuromodulin	P06837	20.946587	19.528947	20.7628328	0	
1252	Q3TFP8	0.342857143	Pgrmc1	Membrane-associated progesterone receptor c	Q3TFP8	17.763472	18.581464	0	18.185276	
1254	Q3V2X9	0.342857143	Prkar2b	cAMP-dependent protein kinase type II-beta re	Q3V2X9	19.494728	18.489659	19.4947278	18.12835	
1292	Q8C0V9-2	0.342857143	Frmd6	FERM domain-containing protein 6	Q8C0V9-2	18.510426	18.259275	18.0039611	0	
1295	Q99JY8	0.342857143	Ppap2b	Lipid phosphate phosphohydrolase 3	Q99JY8	19.781641	19.098691	19.6515217	18.683467	
1308	Q99L61	0.342857143	Tes;Gm4907	Testin	Q99L61	19.584892	17.896666	19.3462969	0	
1338	Q3U4F0	0.328205128	Sfxn3	Sideroflexin-3	Q3U4F0	19.436485	19.997243	18.3009871	19.533985	
1341	O35685	0.328205128	Nudc	Nuclear migration protein nudC	O35685	20.891315	20.729491	20.5885683	20.541473	
1368	A2AKI5	0.342857143	Itgav	Integrin alpha-V;Integrin alpha-V heavy chain;I	A2AKI5	18.336511	18.070804	18.0860816	0	
1369	P62315	0.328205128	Snrpd1	Small nuclear ribonucleoprotein Sm D1	P62315	21.107794	19.235145	20.9521976	0	
1370	K3W4R2	0.342857143	Myh14	Myosin-14	K3W4R2	28.25552	27.653154	28.2541836	27.651545	
1389	Q62426	0.342857143	Cstb	Cystatin-B	Q62426	21.638586	18.239691	21.6209087	0	
1419	Q8C5E7	0.342857143	Snx5	Sorting nexin-5	Q8C5E7	19.641355	17.754062	19.3870186	0	

1455	Q3V2C6	0.342857143	Des	Desmin	Q3V2C6	21.781606	18.830781	21.6947079	0	
1514	A0A0R4J093	0.342857143	Cmpk1	UMP-CMP kinase	A0A0R4J093	18.627286	17.584907	18.5304302	0	
1536	Q9WV98	0.342857143	Timm9	Mitochondrial import inner membrane transloca	Q9WV98	19.820909	17.827608	19.5514061	0	
1541	Q9CWW6	0.342857143	Pin4	Peptidyl-prolyl cis-trans isomerase NIMA-intera	Q9CWW6	19.670602	18.147836	19.2730998	17.569706	
1567	B2RVP5	0.342857143	H2afv;H2afz	Histone H2A;Histone H2A.V;Histone H2A.Z	B2RVP5	21.139945	17.975146	21.0538464	0	
1578	Q8C2A1	0.342857143	Cd34	Hematopoietic progenitor cell antigen CD34	Q8C2A1	18.778217	17.666702	18.316524	0	
1587	A0A1B0GRJ0	0.342857143	Rab11fip1	Rab11 family-interacting protein 1	A0A1B0GRI0	24.412333	25.205799	24.3976437	25.202623	
1595	Q3U055	0.342857143	Plod2	Procollagen-lysine,2-oxoglutarate 5-dioxygenas	Q3U055	18.769312	17.90712	18.1953428	0	
1596	A0A0N4SUZ3	0.342857143	Lsm8	U6 snRNA-associated Sm-like protein LSm8	A0A0N4SUZ3	19.066804	18.289682	18.9616483	0	
1599	A0A0G2JDK2	0.342857143	Gba	Glucosylceramidase	A0A0G2JDK2	19.454306	17.821615	19.1856426	0	
1602	P48771	0.342857143	Cox7a2	Cytochrome c oxidase subunit 7A2, mitochondri	P48771	21.858616	17.828327	21.8205459	0	
1637	Q3T9X3	0.342857143	Dnm2;Dnm3	Dynamin-2;Dynamin-3	Q3T9X3	17.502936	18.225017	15.9064393	0	
1639	Q3V471	0.328205128	Lgals3	Galectin;Galectin-3	Q3V471	21.388243	19.384716	21.2207656	18.205821	
1659	Q8BH78	0.342857143	Rtn4	Reticulon	Q8BH78	21.013207	18.393274	20.9816369	0	
1675	Q3TN35	0.342857143	Sgta	Small glutamine-rich tetratricopeptide repeat-co	Q3TN35	20.436626	18.164748	20.3873163	0	
1694	Q6PJ87	0.342857143	Csnk1a1	Casein kinase I isoform alpha	Q6PJ87	19.280004	17.565931	19.2151678	0	
1728	Q9JJY3	0.342857143	Smpd3	Sphingomyelin phosphodiesterase 3	Q9JJY3	19.331191	17.830153	18.9670429	0	
1746	Q543P7	0.342857143	Arl3	ADP-ribosylation factor-like protein 3	Q543P7	18.789999	17.785016	18.4778	0	
1761	Q64735-2	0.342857143	Cr1l	Complement component receptor 1-like protein	Q64735-2	18.797846	18.244091	18.1049962	0	
1767	Q80WW9	0.342857143	Ddrgk1	DDRGK domain-containing protein 1	Q80WW9	18.661076	18.351221	17.9131614	0	
1794	P43275	0.328205128	Hist1h1a	Histone H1.1	P43275	21.70921	20.926165	21.6792782	20.457482	
1796	E9Q5H2	0.342857143	Anp32e	Acidic leucine-rich nuclear phosphoprotein 32 fa	E9Q5H2	18.868235	18.153562	18.6908593	0	
1822	A2A4J3	0.342857143	Psme3	Proteasome activator complex subunit 3	A2A4J3	19.828822	18.210222	19.4464556	0	
1832	Q8VEI6	0.342857143	Asna1	ATPase Asna1	Q8VEI6	19.89881	17.79896	19.659397	0	
1852	A0A1B0GSY1	0.342857143	Pgpep1	Pyroglutamyl-peptidase 1	A0A1B0GSY1	18.711236	17.429958	18.2222094	0	
1898	A0A0N4SUH8	0.342857143	Nfu1	NFU1 iron-sulfur cluster scaffold homolog, mitoc	A0A0N4SUH8	20.850514	18.255529	20.7633529	0	
1908	A0A0R4J0T8	0.342857143	Arfgap3	ADP-ribosylation factor GTPase-activating prote	A0A0R4J0T8	18.880394	17.74924	18.0138127	0	
1909	D3Z0Z4	0.342857143	Abcg3	ATP-binding cassette sub-family G member 3	D3Z0Z4	19.544432	17.810591	19.297328	0	
1911	Q3TLX9	0.342857143	Smpdl3b	Acid sphingomyelinase-like phosphodiesterase	Q3TLX9	18.497176	17.184045	18.1371573	0	
1919	Q921W0	0.342857143	Chmp1a	Charged multivesicular body protein 1a	Q921W0	20.075737	19.03986	19.7909244	18.206056	
1938	A0A087WSK7	0.342857143	Gm28778;Orm	ORM1-like protein 2;ORM1-like protein 1	A0A087WSK7	17.768417	18.047309	16.5934482	0	
1943	Q3UWX7	0.328205128	Tcea1;Tcea2	Transcription elongation factor A protein 1;Tran	Q3UWX7	19.491331	18.830313	18.9991918	0	
1953	D3Z568	0.342857143	Ndufb5	NADH dehydrogenase [ubiquinone] 1 beta subc	D3Z568	19.393119	18.07385	19.3931191	0	
1954	A0A087WSN6	0.342857143	Fn1		A0A087WSN6	19.95999	17.820086	19.8006396	0	
1960	A0A087WS31	0.342857143	Dnpep	Aspartyl aminopeptidase	A0A087WS31	18.962527	17.76158	18.7834945	0	
1969	Q8C1D0	0.342857143	Spcs3		Q8C1D0	20.481921	17.988259	20.4312857	0	
1980	Q9CPQ1	0.342857143	Cox6c	Cytochrome c oxidase subunit 6C	Q9CPQ1	20.515822	17.473371	20.3798101	0	
1981	Q8R5H1-5	0.342857143	Usp15	Ubiquitin carboxyl-terminal hydrolase 15	Q8R5H1-5	18.869811	18.272583	18.8698114	0	
1992	Q8K273	0.342857143	Mmgt1	Membrane magnesium transporter 1	Q8K273	19.590649	18.121091	19.4399486	16.743229	
2068	Q3UQA0	0.342857143	Il6st;Il6st	Interleukin-6 receptor subunit beta	Q3UQA0	19.224579	17.86942	18.9457104	0	
2129	Q3U8P8	0.342857143	Jak1	Tyrosine-protein kinase;Tyrosine-protein kinase	Q3U8P8	17.881007	18.906634	0	18.507183	
179	Q60972	0.347357919	Rbbp4	Histone-binding protein RBBP4	Q60972	20.53034	20.492909	19.7574941	19.830446	
181	Q561N4	0.347357919	Ube2l3	Ubiquitin-conjugating enzyme E2 L3	Q561N4	22.496491	20.614557	22.4201625	19.924775	
437	P99024	0.351411522	Tubb5	Tubulin beta-5 chain	P99024	25.790849	24.737052	25.7781699	24.703287	
711	Q545F0	0.351411522	Mif	Macrophage migration inhibitory factor	Q545F0	24.278736	24.919609	24.2383309	24.891626	
456	Q3U8F5	0.353454493	Psmd12	26S proteasome non-ATPase regulatory subuni	Q3U8F5	21.788745	22.669651	21.556891	22.33305	
901	Q2YDW1	0.360808603	Eif3j;Eif3j1;Eif	Eukaryotic translation initiation factor 3 subunit	Q2YDW1	22.419206	20.56381	22.3039218	18.770434	
1333	Q9CQ19	0.373099055	Myl9	Myosin regulatory light polypeptide 9	Q9CQ19	26.825202	26.72849	26.8202777	26.714688	
194	Q8BH64	0.377685311	Ehd2	EH domain-containing protein 2	Q8BH64	20.815126	20.350042	20.5076069	19.658809	
1483	Q3TVJ8	0.377685311	Ssr4	Translocon-associated protein subunit delta	Q3TVJ8	22.087135	20.457361	22.0109185	19.652984	
1505	Q3UJ76	0.377685311	Ap2s1	AP-2 complex subunit sigma	Q3UJ76	20.503591	19.667732	20.1455154	17.948616	
779	A0A0G2JEC4	0.382284382	Sh3glb1	Endophilin-B1	A0A0G2JEC4	19.580729	19.40723	18.3117878	18.446784	
898	Q6GQT9	0.382284382	Nomo1	Nodal modulator 1	Q6GQT9	19.460094	19.3844	18.6257873	17.928375	
993	Q3U944	0.380896465	Lman1	Protein ERGIC-53	Q3U944	21.350669	20.59744	21.0059154	19.543237	
1171	Q3UE99	0.382284382	Lgmn	Legumain	Q3UE99	19.925091	19.285825	19.4297521	0	
1274	Q8BXY4	0.382284382	Rab2a;Rab2b;	Ras-related protein Rab-2A;Ras-related protein	Q8BXY4	21.786465	20.750919	21.6321447	20.442976	
1441	Q4KML7	0.382284382	Ezr	Ezrin	Q4KML7	20.445256	19.425008	20.0766892	18.328593	
1474	Q3UPB9	0.382284382	Snx18	Sorting nexin;Sorting nexin-18	Q3UPB9	20.711732	19.160592	20.5184937	0	
1627	Q8VDX8	0.382284382	Hars;Hars2	Histidine--tRNA ligase, cytoplasmic;Probable his	Q8VDX8	20.214322	19.216103	20.0490659	16.714782	
1672	Q3V0L4	0.382284382	Rbms1;Rbms3	RNA-binding motif, single-stranded-interacting p	Q3V0L4	21.083629	20.37643	20.9535612	19.949748	
1881	Q99LC3	0.382284382	Ndufa10	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q99LC3	20.499735	18.979066	20.2666643	0	
2105	A0A140LHA2	0.382284382	Bub3	Mitotic checkpoint protein BUB3	A0A140LHA2	20.785006	20.87474	20.5582513	20.48007	
76	E0CZ72	0.386372964	Kif2a;Kif2c	Kinesin-like protein;Kinesin-like protein KIF2A;K	E0CZ72	20.487069	19.925868	20.2696631	19.107908	
1563	Q0PD38	0.395580536	Rab18	Ras-related protein Rab-18	Q0PD38	22.815852	22.351777	22.6668786	22.14502	
736	Q5SUH7	0.402401515	Clint1	Clathrin interactor 1	Q5SUH7	20.435031	19.960091	20.1656248	18.595982	
903	Q8VDP9	0.407115953	Fhl2	Four and a half LIM domains protein 2	Q8VDP9	22.986288	22.158625	22.9111116	21.864149	
144	A2BFF8	0.409510398	Dync1i2	Cytoplasmic dynein 1 intermediate chain 2	A2BFF8	21.433605	20.43887	21.2760714	19.874922	
239	Q6PDY2	0.409510398	Ado	2-aminoethanethiol dioxygenase	Q6PDY2	21.012522	22.233503	20.7562395	22.126909	
1102	Q9WVH9	0.409510398	Fbln5	Fibulin-5	Q9WVH9	20.173123	20.061932	19.4034049	19.202924	
1590	Q3U2X7	0.409510398	Nov	Protein NOV homolog	Q3U2X7	20.95779	19.554977	20.7107741	0	
1461	Q8R326	0.423019989	Pspc1	Paraspeckle component 1	Q8R326	21.243231	19.940279	20.9869983	0	
1124	Q8C3Z4	0.428629048	Prkar1a;Prkar	cAMP-dependent protein kinase type I-alpha re	Q8C3Z4	22.273793	21.014873	22.032358	20.518607	
57	Q3UXP2	0.430768443	Ruvbl2	RuvB-like 2	Q3UXP2	22.084836	21.187806	21.7884157	20.062777	
323	E9Q5Y2	0.441802642	U2af1l4;U2af	Splicing factor U2AF 26 kDa subunit;Splicing fac	E9Q5Y2	21.056494	20.226516	20.8985283	19.794151	
575	Q3UAS4	0.44283318	Psap	Prosaposin	Q3UAS4	21.681845	19.842256	21.5433833	0	
1223	Q3U281	0.441802642	Sar1b	GTP-binding protein SAR1b	Q3U281	19.898926	19.290474	19.3325579	0	
1265	Q9QYE6	0.441802642	Golga5	Golgin subfamily A member 5	Q9QYE6	19.964993	18.863782	19.7887694	0	
1323	Q9DAU1	0.441802642	Cnpy3	Protein canopy homolog 3	Q9DAU1	20.354355	19.162089	20.0894462	0	
1380	B7ZN33	0.441802642	Nisch	Nischarin	B7ZN33	19.982024	18.880599	19.5914505	0	
1385	Q80YG4	0.441802642	Sec63	Translocation protein SEC63 homolog	Q80YG4	19.246769	19.496988	17.8276156	19.084906	
1449	P14733	0.44283318	Lmnb1	Lamin-B1	P14733	21.068958	19.712013	20.9252713	0	
1470	A0A0J9YUB6	0.441802642	Ociad1	OCIA domain-containing protein 1	A0A0J9YUB6	19.623644	19.084718	19.2438473	17.151882	
1565	Q5SXA5	0.441802642	Tom1l2	TOM1-like protein 2	Q5SXA5	20.257791	19.638854	19.8867118	18.803487	
1680	Q9D6U8	0.441802642	Fam162a	Protein FAM162A	Q9D6U8	19.900423	19.014521	19.2442189	0	
2111	Q5M9N6	0.441802642	Rpl37a	60S ribosomal protein L37a	Q5M9N6	22.492027	21.783227	22.4480382	21.718937	
2140	V9GXQ2	0.441802642	Gm17087		V9GXQ2	26.183667	25.602241	26.1797836	25.589388	
974	Q99KV1	0.456638266	Dnajb11	DnaJ homolog subfamily B member 11	Q99KV1	24.116395	25.089779	23.973778	25.029772	
1929	Q5SUF2	0.461920478	Luc7l3	Luc7-like protein 3	Q5SUF2	20.041192	25.584322	19.52014	25.577846	

22	Q3THC1	0.477575258	Psmd9	26S proteasome non-ATPase regulatory subuni	Q3THC1	20.740948	19.779242	20.2398315	0	
42	Q9Z1P6	0.485714286	Ndufa7	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q9Z1P6	20.037296	18.190415	19.8044703	0	
141	A2ATP5	0.485714286	Myef2	Myelin expression factor 2	A2ATP5	18.144247	17.838884	17.9686489	0	
250	Q8CDJ8	0.485714286	Ston1	Stonin-1	Q8CDJ8	19.635551	17.756376	19.4383942	0	
560	E9QP49	0.485714286	Ehbp1l1	EH domain-binding protein 1-like protein 1	E9QP49	19.399012	17.921082	19.145386	0	
569	Q6NV59	0.485714286	Loxl4	Lysyl oxidase homolog 4	Q6NV59	19.077687	17.93981	18.9233854	0	
603	Q9WTP6-2	0.485714286	Ak2	Adenylate kinase 2, mitochondrial;Adenylate ki	Q9WTP6-2	19.609159	18.367689	19.5167566	0	
607	A0A0A6YWM5	0.485714286	Rab3gap2	Rab3 GTPase-activating protein non-catalytic su	A0A0A6YWM5	21.617272	21.178586	21.5705003	20.96334	
650	Q6NVF2	0.485714286	Pld1	Phospholipase D1	Q6NVF2	18.111198	18.102	17.6349599	0	
783	Q545F4	0.467673627	Hspb1	Heat shock protein beta-1	Q545F4	22.90987	22.550902	22.8193557	22.411907	
803	Q6IRT4	0.485714286	Eif3f	Eukaryotic translation initiation factor 3 subunit	Q6IRT4	20.388857	18.394938	20.2520422	16.361699	
858	E9Q3F3	0.485714286	Efemp2	EGF-containing fibulin-like extracellular matrix p	E9Q3F3	18.424402	18.323827	18.0856093	17.305252	
863	Q05D44	0.477575258	Eif5b	Eukaryotic translation initiation factor 5B	Q05D44	21.068783	21.150019	20.9343889	20.764327	
925	Q64519	0.485714286	Sdc3	Syndecan-3	Q64519	19.53061	18.029084	19.2537833	0	
958	F8WHP8	0.485714286	Atp5j2	ATP synthase subunit f, mitochondrial	F8WHP8	21.222763	19.295121	21.1866532	18.715198	
966	D3Z249	0.485714286	Myl12a		D3Z249	19.035607	17.675431	18.7790396	0	
968	Q6XLQ8	0.485714286	Calu		Q6XLQ8	21.964473	20.546605	21.9510476	20.469243	
998	Q0QER9	0.485714286	Idh1	Isocitrate dehydrogenase [NADP];Isocitrate deh	Q0QER9	19.389401	18.16021	19.0755975	0	
1021	Q06G06	0.485714286	Pla2g4b;Gm2	Cytosolic phospholipase A2 beta	Q06G06	18.324748	17.990517	17.9077612	0	
1086	U3M993	0.485714286	Nsdhl	Sterol-4-alpha-carboxylate 3-dehydrogenase, d	U3M993	19.742088	18.072554	19.6080386	0	
1190	Q3UW53	0.478942395	Fam129a	Protein Niban	Q3UW53	21.707292	21.957953	21.3089439	21.519002	
1197	E9QPX6	0.485714286	AW554918;Ki	Uncharacterized protein KIAA1328	E9QPX6	19.536874	17.769177	19.536874	0	
1199	Q4VAI2	0.485714286	Acp1	Low molecular weight phosphotyrosine protein	Q4VAI2	21.71476	17.989508	21.7000961	0	
1202	E9Q1H3	0.485714286	Aldh7a1	Alpha-aminoadipic semialdehyde dehydrogena	E9Q1H3	18.323901	17.921251	17.5384378	0	
1381	Q8R1F1	0.485714286	Fam129b	Niban-like protein 1	Q8R1F1	19.676918	17.898918	19.5986696	0	
1442	Q7TNV0	0.485714286	Dek	Protein DEK	Q7TNV0	19.516912	18.591501	19.318561	18.148887	
1453	Q9CRX2	0.485714286	Laptm4a	Lysosomal-associated transmembrane protein 4	Q9CRX2	19.97515	17.845624	19.770279	0	
1532	Q6P9J9	0.485714286	Ano6	Anoctamin-6	Q6P9J9	19.65897	18.839996	19.6589701	18.295556	
1603	G3UY21	0.478942395	Slit2	Slit homolog 2 protein;Slit homolog 2 protein N-	G3UY21	21.489054	21.845937	20.8868494	21.200921	
1648	D3Z479	0.485714286	Snx15	Sorting nexin-15	D3Z479	18.686386	17.844788	18.3654406	0	
1657	Q3UJC3	0.485714286	Nmt1	Glycylpeptide N-tetradecanoyltransferase;Glycy	Q3UJC3	18.915549	17.931048	18.5485887	0	
1669	F6VQ81	0.485714286	Tpd52l2	Tumor protein D54	F6VQ81	19.892445	17.869045	19.783805	0	
1685	Q3V1X1	0.485714286	Mycbp	C-Myc-binding protein	Q3V1X1	19.010941	17.886685	18.6813078	0	
1690	P70245	0.485714286	Ebp	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomer	P70245	19.741941	17.737526	19.6440657	0	
1722	Q9JMH6-2	0.485714286	Txnrd1	Thioredoxin reductase 1, cytoplasmic	Q9JMH6-2	19.858329	18.572532	19.6179801	0	
1734	A0A0R3P9C8	0.485714286	Ndufa9	NADH dehydrogenase [ubiquinone] 1 alpha sub	A0A0R3P9C8	19.887834	18.150043	19.8878342	0	
1747	F6Z6F4	0.485714286	Smarca1;Sma	Probable global transcription activator SNF2L1;S	F6Z6F4	18.745966	17.855914	17.8004652	0	
1755	C6EQH3	0.485714286	Suclg2	Succinyl-CoA ligase subunit beta;Succinyl-CoA li	C6EQH3	18.997521	17.943773	18.6464924	0	
1806	Q9CX78	0.485714286	Ndufa5	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q9CX78	18.84268	18.257055	18.84268	0	
1878	Q6GU23	0.485714286	Stat3	Signal transducer and activator of transcription;	Q6GU23	18.525132	17.529045	17.9957776	0	
1966	Q6A0A2-2	0.485714286	Larp4b	La-related protein 4B	Q6A0A2-2	18.304205	18.150682	17.8044976	0	
1985	Q8CBS7	0.485714286	Tulp3	Tubby-related protein 3	Q8CBS7	24.479273	23.749756	24.477195	23.742953	
2006	A0A0K6GHA8	0.485714286			A0A0K6GHA8	19.009768	21.908306	15.7528385	21.760056	
2038	D3YXU7	0.485714286	Myo10	Unconventional myosin-X	D3YXU7	17.444527	18.909039	0	18.659573	
2051	Q8CGB6-2	0.485714286	Tns2;Tns1;Tn	Tensin-2;Tensin-3	Q8CGB6-2	18.350842	18.866283	0	18.148066	
2060	G3X9V2	0.485714286	Ctnnd1	Catenin delta-1	G3X9V2	20.170787	17.743189	20.0393996	0	
2081	Q1XG81	0.485714286	Higd1a	HIG1 domain family member 1A, mitochondrial	Q1XG81	19.053303	17.740266	18.6806833	0	
2085	Q3UDS4	0.477575258	Sqrdl	Sulfide:quinone oxidoreductase, mitochondrial	Q3UDS4	21.105338	20.25148	20.826283	19.229453	
2096	E9QPX1	0.485714286	Col18a1	Collagen alpha-1(XVIII) chain;Endostatin	E9QPX1	19.661442	19.795497	19.5721767	19.487452	
2116	Q9D1T2	0.485714286	Mxra7	Matrix-remodeling-associated protein 7	Q9D1T2	19.604269	17.573244	19.6042691	0	
2123	Q91Z73	0.485714286	B2m	Beta-2-microglobulin	Q91Z73	20.558487	18.025359	20.4354747	0	
2127	Q91VT9	0.485714286	Pvrl2	Nectin-2	Q91VT9	18.363456	18.180473	17.6814734	0	
2132	A0A0G2JE32	0.485714286	Ube2d3;Ube2	Ubiquitin-conjugating enzyme E2 D2;Ubiquitin-c	A0A0G2JE32	19.332865	17.719609	19.1090767	0	
834	Q61081	0.490915916	Cdc37	Hsp90 co-chaperone Cdc37;Hsp90 co-chaperon	Q61081	21.981212	21.092647	21.8274926	20.619802	
240	Q91ZH2	0.493199174	Pcna	Proliferating cell nuclear antigen	Q91ZH2	22.425164	21.313471	22.1830301	20.554561	
62	Q9CWK0	0.494523496	Rpl14;Rpl14-p	60S ribosomal protein L14	Q9CWK0	24.341756	23.662519	24.3240027	23.596525	
980	Q99M31	0.494518992	Hspa14	Heat shock 70 kDa protein 14	Q99M31	19.58499	19.279672	19.3836592	0	
44	A2AI08	0.505361305	Tprn	Taperin	A2AI08	20.178188	19.598026	19.8015939	18.808629	
96	Q6PB44-2	0.505361305	Ptpn23	Tyrosine-protein phosphatase non-receptor typ	Q6PB44-2	19.878844	19.116483	19.5946797	0	
710	Q9CXW3	0.505361305	Cacybp	Calcyclin-binding protein	Q9CXW3	19.440153	19.697254	18.5025988	19.0359	
1043	O35558	0.505361305	Erk2;Mapk1;I	Mitogen-activated protein kinase;Mitogen-acti	O35558	20.604233	19.155442	20.5035814	0	
1243	Q5SV80	0.505361305	Myo19	Unconventional myosin-XIX	Q5SV80	20.331969	19.603225	20.0200836	0	
1263	Q61935	0.505361305	Myl3;Myl1	Myosin light chain 1/3, skeletal muscle isoform;	Q61935	25.367045	25.311357	25.3576775	25.305298	
1372	Q8K411-3	0.505361305	Pitrm1	Presequence protease, mitochondrial	Q8K411-3	20.718085	19.771956	20.5960853	18.870824	
1827	Q9D1E6	0.505361305	Tbcb	Tubulin-folding cofactor B	Q9D1E6	19.992661	18.77797	19.7900541	0	
345	Q3TJD0	0.513722581	Sec61a1;Sec6	Protein transport protein Sec61 subunit alpha iso	Q3TJD0	22.409938	21.975336	22.236766	21.78983	
594	Q9D6R2	0.514748002	Idh3a	Isocitrate dehydrogenase [NAD] subunit alpha,	Q9D6R2	21.578228	20.720224	21.2239385	20.176215	
1509	Q571B0	0.514748002	Tm9sf3	Transmembrane 9 superfamily member 3	Q571B0	21.223823	20.437129	20.8566471	19.510219	
2102	Q7M754	0.514748002	Gm5409;Try10		Q7M754	27.592584	28.955306	27.5901685	28.954363	
731	B1AXI9	0.536802024	Cep131	Centrosomal protein of 131 kDa	B1AXI9	22.200743	22.488012	21.7257246	22.063626	
91	Q5FWB7	0.537476216	Aldoa;Aldoart	Fructose-bisphosphate aldolase;Fructose-bisph	Q5FWB7	24.351608	24.368097	24.2251719	24.257851	
327	Q9EPL8	0.53914969	Ipo7	Importin-7	Q9EPL8	21.239138	20.619233	21.0190132	19.718	
351	Q3TJD7	0.546823018	Pdlim7	PDZ and LIM domain protein 7	Q3TJD7	22.830835	22.657923	22.6975413	22.547303	
1998	P61092	0.546823018	Siah1a	E3 ubiquitin-protein ligase SIAH1A	P61092	23.127971	22.802646	23.0419567	22.594767	
1174	Q9CQU0	0.551167166	Txndc12	Thioredoxin domain-containing protein 12	Q9CQU0	21.184209	19.369289	21.086934	0	
1447	P97333	0.551167166	Nrp1	Neuropilin-1	P97333	20.320999	19.780441	20.1290529	18.519696	
2071	Q6A0A9	0.551167166	FAM120A;Fa	Constitutive coactivator of PPAR-gamma-like pr	Q6A0A9	21.676421	20.791853	21.5197496	20.323635	
721	P70333	0.563236627	Hnrnph2	Heterogeneous nuclear ribonucleoprotein H2	P70333	20.245715	19.465886	20.0610277	0	
109	Q99LC5	0.564102006	Etfa	Electron transfer flavoprotein subunit alpha, mit	Q99LC5	21.248531	20.747894	20.9184094	20.123538	
426	Q3U7I9	0.564831637	Ctsd	Cathepsin D	Q3U7I9	22.073309	21.340835	21.8014876	20.632092	
573	Q4TVN0	0.564831637	Tbc1d15	TBC1 domain family member 15	Q4TVN0	21.462747	20.522852	21.1063728	17.909758	
1226	Q8K1K2	0.564102006	Psmc5	26S protease regulatory subunit 8	Q8K1K2	20.634707	20.582427	19.9031429	19.274518	
506	O35405	0.573737374	Pld3	Phospholipase D3	O35405	20.846287	19.27485	20.5839064	17.815979	
933	Q9CY58-2	0.573737374	Serbp1	Plasminogen activator inhibitor 1 RNA-binding p	Q9CY58-2	19.447245	20.485835	18.5305228	20.162208	
1030	Q9DBN5	0.573737374	Lonp2	Lon protease homolog 2, peroxisomal	Q9DBN5	19.018316	20.16934	0	19.75003	
1122	Q3U6M5	0.573737374	Ranbp1;Htf9-a	Ran-specific GTPase-activating protein	Q3U6M5	21.972169	20.61965	21.8632574	20.157312	
1148	A2A815	0.573737374	Park7;Dj1	Protein deglycase DJ-1	A2A815	21.247252	19.78039	21.1563215	19.249497	
1222	Q6ZWR6-4	0.573737374	Syne1	Nesprin-1	Q6ZWR6-4	26.949	28.606441	26.9436965	28.605454	

1706	Q9ERS2	0.573737374	Ndufa13	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q9ERS2	20.129059	19.570406	19.780425	18.41544	
1739	Q3US65	0.573737374	Srpk1	SRSF protein kinase 1	Q3US65	20.017791	18.865195	19.8792625	0	
588	Q9DCN2-2	0.587992654	Cyb5r3	NADH-cytochrome b5 reductase 3;NADH-cytocl	Q9DCN2-2	23.792451	23.006322	23.7180646	22.908235	
897	P27046	0.589875732	Man2a1	Alpha-mannosidase 2	P27046	20.057139	19.991266	18.7141818	18.399384	
970	Q80U35	0.589875732	Arhgef17	Rho guanine nucleotide exchange factor 17	Q80U35	20.054514	20.190516	18.9089459	19.393157	
1392	Q3UIA5	0.589875732	Cul1	Cullin-1	Q3UIA5	20.683019	20.436523	20.2090837	19.369567	
77	Q58EA6	0.615563482	Rps25	40S ribosomal protein S25	Q58EA6	24.923236	24.793106	24.9074284	24.77275	
1040	P63028	0.629725504	Tpt1	Translationally-controlled tumor protein	P63028	22.614426	21.151558	22.5102751	20.85476	
253	Q8BLD7	0.645376845	Acaa1a;Acaa1	3-ketoacyl-CoA thiolase A, peroxisomal;3-ketoa	Q8BLD7	20.205577	18.928531	20.1517739	0	
294	A2RSB1	0.645376845	Nap1l4	Nucleosome assembly protein 1-like 4	A2RSB1	20.10824	19.393575	19.6692935	18.478525	
483	F6VQ19	0.645376845	Lrch2	Leucine-rich repeat and calponin homology dom	F6VQ19	20.794922	19.326098	20.7646547	0	
961	Q8C7V6	0.645376845	Abcc1	Multidrug resistance-associated protein 1	Q8C7V6	19.301086	18.991052	18.7279999	0	
1027	Q0VGU9	0.645376845	Rbm39	RNA-binding protein 39	Q0VGU9	19.708332	19.560836	19.0629489	18.793976	
1072	Q3UWH6	0.645376845	Ctsl	Cathepsin L1;Cathepsin L1 heavy chain;Catheps	Q3UWH6	21.321048	20.124138	21.2200088	19.514326	
1130	G3UXW9	0.645376845	Gps1	COP9 signalosome complex subunit 1	G3UXW9	20.280043	19.692637	20.0547566	19.001609	
1160	P10493	0.645376845	Nid1	Nidogen-1	P10493	19.175121	19.329724	18.5622595	17.030859	
1316	A0A1B0GT04	0.645376845	Glrx3	Glutaredoxin-3	A0A1B0GT04	20.182538	19.75593	19.7358999	19.354988	
1378	Q3TI06	0.645376845	Ugdh	UDP-glucose 6-dehydrogenase	Q3TI06	20.544321	20.59413	20.3104766	20.034808	
1730	Q3TRL2	0.645376845	Pdpk1	3-phosphoinositide-dependent protein kinase 1	Q3TRL2	20.021385	18.976961	19.7551769	0	
2079	E9Q4G8	0.645376845	Alcam	CD166 antigen	E9Q4G8	20.311348	19.030202	19.9103414	0	
5	Q3ULG4	0.685714286	Psmd4	26S proteasome non-ATPase regulatory subuni	Q3ULG4	19.434005	18.6378	19.3677045	17.953407	
11	A0A097BW21	0.670659533	Postn	Periostin	A0A097BW21	20.679988	21.312613	20.0821756	20.87055	
97	Q8BPS5	0.685714286	Slc16a1	Monocarboxylate transporter 1	Q8BPS5	18.415115	18.879888	0	18.369944	
120	Q545A2	0.675265662	Slc25a5	ADP/ATP translocase 2;ADP/ATP translocase 2,	Q545A2	25.36845	25.586571	25.3522731	25.573614	
147	Q8VE99	0.685714286	Ccdc115	Coiled-coil domain-containing protein 115	Q8VE99	18.032088	17.601317	17.3171264	0	
197	Q3UP74	0.685714286	Anpep	Aminopeptidase N	Q3UP74	19.202474	18.073585	18.9550275	0	
305	Q3TN93	0.685714286	Ubqln1	Ubiquilin-1	Q3TN93	18.813472	17.867621	18.3401212	0	
341	Q3UIW3	0.685714286	Lamp2	Lysosome-associated membrane glycoprotein 2	Q3UIW3	20.698463	19.061598	20.6744362	18.801413	
431	P40240	0.685714286	Cd9	CD9 antigen	P40240	19.087901	17.660995	18.7817616	0	
519	Q8C2S9	0.685714286	Mvp	Major vault protein	Q8C2S9	18.018695	19.097438	0	18.780483	
667	Q3UQL2	0.685714286	Cops3	COP9 signalosome complex subunit 3	Q3UQL2	19.50277	17.999927	19.3553737	0	
769	Q3V222	0.685714286	Sec24c		Q3V222	18.858989	18.233295	18.6127429	0	
795	Q8BMJ3	0.685714286	Eif1ax;Eif1a;G	Eukaryotic translation initiation factor 1A, X-chro	Q8BMJ3	21.852083	20.565252	21.8368632	20.542057	
896	Q3U5V2	0.685714286	Cnbp	Cellular nucleic acid-binding protein	Q3U5V2	20.262576	18.557348	20.1601066	17.854688	
923	Q99L05	0.685714286	Sdc2	Syndecan;Syndecan-2	Q99L05	20.598647	20.119292	20.4514476	19.884029	
949	F6V084	0.685714286	Tmx1	Thioredoxin-related transmembrane protein 1	F6V084	19.793072	17.905706	19.6361716	0	
1005	X1WI15	0.685714286	BC017643	Uncharacterized protein C17orf62 homolog	X1WI15	19.01316	17.653616	18.8698049	0	
1029	Q8CH18-3	0.685714286	Ccar1	Cell division cycle and apoptosis regulator prote	Q8CH18-3	18.657893	17.568649	18.4927549	0	
1164	A2AWT6	0.685714286	Ubtf	Nucleolar transcription factor 1	A2AWT6	18.074627	18.949505	0	18.335783	
1177	E9QNH6	0.685714286	Myo1b		E9QNH6	21.008269	18.189689	21.0082688	0	
1179	B1AU76	0.685714286	Nasp	Nuclear autoantigenic sperm protein	B1AU76	19.75846	17.961808	19.6916846	0	
1186	A0A0N4SV15	0.685714286	Kcmf1	E3 ubiquitin-protein ligase KCMF1	A0A0N4SV15	18.72526	18.05114	18.5728177	0	
1218	Q9QY73	0.685714286	Tmem59	Transmembrane protein 59	Q9QY73	18.996981	18.463604	18.5467723	0	
1237	Q0PD66	0.685714286	Rab1b	Ras-related protein Rab-1B	Q0PD66	21.136756	21.708336	21.0956027	21.695552	
1248	E9Q0Z7	0.685714286	Sumf1	Sulfatase-modifying factor 1	E9Q0Z7	19.079769	18.035266	18.9321787	0	
1253	Q3UD86	0.685714286	Mov10	Putative helicase MOV-10	Q3UD86	19.409529	18.188042	19.1319396	0	
1258	Q6A022	0.685714286	mKIAA0668;S	SUN domain-containing protein 2	Q6A022	19.949656	19.55628	19.8739595	19.430636	
1278	Q9CR41	0.685714286	Hypk	Huntingtin-interacting protein K	Q9CR41	18.781949	17.723267	18.3959484	0	
1299	D3Z6I0	0.685714286	Prkd3	Protein kinase C;Serine/threonine-protein kinas	D3Z6I0	18.279401	17.912297	17.7694273	0	
1301	Q5FW91	0.685714286	Tuba3a	Tubulin alpha-3 chain	Q5FW91	19.350047	18.086931	19.0876671	0	
1326	Q7TMW3	0.685714286	Bgn	Biglycan	Q7TMW3	21.31782	20.643181	21.3044544	20.59249	
1339	Q9EP69	0.685714286	Sacm1l	Phosphatidylinositide phosphatase SAC1	Q9EP69	18.546011	18.327502	18.0721171	0	
1395	Q9CYA0	0.685714286	Creld2	Cysteine-rich with EGF-like domain protein 2	Q9CYA0	18.448272	18.322758	0	17.515488	
1399	Q8C073	0.685714286	Sh3bgrl2	SH3 domain-binding glutamic acid-rich-like prot	Q8C073	18.853634	18.180968	18.5121989	0	
1402	Q9JLB9	0.685714286	Pvrl3	Nectin-3	Q9JLB9	18.996939	17.806971	18.8496818	0	
1416	Q8C0E2	0.685714286	Vps26b	Vacuolar protein sorting-associated protein 26B	Q8C0E2	18.748551	18.376711	18.4953893	0	
1428	Q8BQ47	0.685714286	Cnpy4	Protein canopy homolog 4	Q8BQ47	18.917015	18.713274	18.3058715	17.941626	
1457	O08797	0.685714286	Serpinb9;Serpinb9g;Serpinb9d;Serpinb9e;Serpinb9f;Gm11397		O08797	19.544183	17.937709	19.1517834	0	
1522	Q5SVD0	0.685714286	Fam101b	Filamin-interacting protein FAM101B	Q5SVD0	20.00447	18.11534	19.8351383	0	
1523	Q2TBF9	0.685714286	Map1lc3b	Microtubule-associated proteins 1A/1B light cha	Q2TBF9	21.22191	18.142841	21.1630376	0	
1534	Q921U7	0.685714286	Cast	Calpastatin	Q921U7	18.858614	18.401554	18.6257902	0	
1539	Q3TJS8	0.685714286	Chchd6	MICOS complex subunit Mic25	Q3TJS8	20.116068	17.907871	19.9850901	0	
1548	P62867	0.685714286	Fau;fau	40S ribosomal protein S30	P62867	24.054424	23.141711	24.044912	23.118638	
1575	Q3U316	0.685714286	Tnpo2	Transportin-2	Q3U316	19.044683	18.274685	18.392091	17.637729	
1586	A0A0J9YUI1	0.685714286	Ube2k	Ubiquitin-conjugating enzyme E2 K	A0A0J9YUI1	19.469454	18.22709	19.132695	0	
1604	P56391	0.685714286	Cox6b1	Cytochrome c oxidase subunit 6B1	P56391	21.622852	19.138235	21.5648058	18.385952	
1649	Q8K567	0.685714286	Atp6ap1	V-type proton ATPase subunit S1	Q8K567	19.580868	18.202929	19.3293947	0	
1653	Q9D7M1	0.685714286	Gid8	Glucose-induced degradation protein 8 homolog	Q9D7M1	18.578445	17.992844	18.1170495	0	
1666	Q497K3	0.685714286	Snrpf	Small nuclear ribonucleoprotein F	Q497K3	19.978643	18.062319	19.821746	0	
1704	Q9DCC4	0.685714286	Pycrl	Pyrroline-5-carboxylate reductase 3	Q9DCC4	18.52853	18.063404	17.7081743	17.019375	
1724	Q05DE0	0.685714286	Arfgap1	ADP-ribosylation factor GTPase-activating prote	Q05DE0	19.450976	18.644501	19.3025076	17.577077	
1738	O88983	0.685714286	Stx8	Syntaxin-8	O88983	18.617245	17.679092	18.2457615	0	
1743	Q9D7B7	0.685714286	Gpx8	Probable glutathione peroxidase 8	Q9D7B7	18.227185	17.914248	0	17.004892	
1760	Q7TMM9	0.685714286	Tubb2a	Tubulin beta-2A chain	Q7TMM9	21.305546	19.739409	21.305546	19.624966	
1785	Q3THK3	0.685714286	Gtf2f1	General transcription factor II F subunit 1	Q3THK3	18.378092	17.758451	17.9137613	0	
1798	A8XY17	0.685714286	Vps25	Vacuolar protein-sorting-associated protein 25	A8XY17	18.823807	18.271504	18.3876629	0	
1844	D3Z6N3	0.685714286	Mcm7	DNA helicase;DNA replication licensing factor M	D3Z6N3	17.933674	18.711021	0	18.21031	
1847	Q3TUQ7	0.685714286	Prkaa1	5-AMP-activated protein kinase catalytic subun	Q3TUQ7	19.569847	17.9942	19.256292	0	
1850	E9QAF9	0.685714286	Tanc1	Protein TANC1	E9QAF9	19.265226	18.25089	19.0246917	0	
1855	Q3U950	0.685714286	Mogs	Mannosyl-oligosaccharide glucosidase	Q3U950	18.843616	18.025838	18.4575531	0	
1858	Q3UKN6	0.685714286	Nucb2	Nucleobindin-2;Nesfatin-1	Q3UKN6	19.676502	18.105972	19.3709203	0	
1870	B7ZBY6	0.685714286	Ube2v1;Ube2	Ubiquitin-conjugating enzyme E2 variant 2;Ubic	B7ZBY6	21.089539	19.374291	21.0414412	18.899949	
1872	P52480-2	0.685714286	Pkm	Pyruvate kinase PKM	P52480-2	19.665838	18.618611	19.4335071	18.204652	
1890	Q3UHF8	0.685714286	Reps2;Reps1	RalBP1-associated Eps domain-containing prote	Q3UHF8	19.396892	17.753171	19.223625	0	
1895	F7CVJ5	0.685714286	Ahnak2		F7CVJ5	19.160369	18.234638	18.7566283	0	
1902	Q60902-3	0.685714286	Eps15l1	Epidermal growth factor receptor substrate 15-l	Q60902-3	19.107346	17.845087	19.1073465	0	
1918	A2API8	0.685714286	Akap2	A-kinase anchor protein 2	A2API8	18.009215	19.431424	0	19.21449	
1931	U5TIY7	0.685714286			U5TIY7	21.147941	24.950918	21.0428357	24.942631	

1997	Q5I0U7	0.685714286	Tbca	Tubulin-specific chaperone A	Q5I0U7	20.925268	19.292989	20.8011056	17.783381	
2008	Q9Z0Y1	0.685714286	Dctn3	Dynactin subunit 3	Q9Z0Y1	18.71947	18.064501	18.242243	0	
2011	Q9CSH0	0.685714286	Hnrnpll	Heterogeneous nuclear ribonucleoprotein L-like	Q9CSH0	19.311541	17.909408	19.0489423	0	
2080	A0A077S9N1	0.685714286	Lyz1	Lysozyme;Lysozyme C-1	A0A077S9N1	20.046705	20.103399	19.9872827	19.945603	
2093	Q8C1Y3	0.685714286	H1f0	Histone H1.0;Histone H1.0, N-terminally proces	Q8C1Y3	19.964756	18.042664	19.8559034	0	
2114	Q9CQ75	0.685714286	Ndufa2	NADH dehydrogenase [ubiquinone] 1 alpha sub	Q9CQ75	19.494356	18.157028	19.4943564	0	
2145	Q8K4L4	0.685714286	Pof1b	Protein POF1B	Q8K4L4	20.175235	20.538545	20.0403658	20.421348	
1231	D3Z7T4	0.696258329	Hdac10	Histone deacetylase 10	D3Z7T4	21.007402	20.816814	20.4059979	20.100152	
162	P49817-2	0.720901321	Cav1	Caveolin-1;Caveolin	P49817-2	20.573653	22.8955	20.2696481	22.832878	
172	B2CY77	0.715275339	Rpsa	40S ribosomal protein SA	B2CY77	25.883904	25.279569	25.850548	25.203653	
379	Q3U5U5	0.720901321	Ahcy	Adenosylhomocysteinase	Q3U5U5	21.275219	20.216888	21.1026134	19.522002	
383	Q9JL12	0.712989487	Tcirg1	V-type proton ATPase subunit a	Q9JL12	20.440753	19.266463	20.364749	0	
458	O55234	0.720901321	Psmb5	Proteasome subunit beta type-5	O55234	20.362107	20.729354	20.0687524	20.552172	
678	Q9Z175	0.720901321	Loxl3	Lysyl oxidase homolog 3	Q9Z175	19.353754	19.777213	0	19.273112	
696	Q5BJ94	0.720901321	Kifc1;Kifc5b	Kinesin-like protein;Kinesin-like protein KIFC1	Q5BJ94	20.247802	19.185544	20.0444772	0	
823	Q3UJ36	0.720901321	Actg2	Actin, gamma-enteric smooth muscle	Q3UJ36	23.03805	22.023696	23.0089899	21.907791	
910	Q5SUR3	0.712535076	Skp1a;Skp1	S-phase kinase-associated protein 1	Q5SUR3	21.757128	21.723206	21.5079322	21.599622	
957	Q9DCT2	0.720901321	Ndufs3	NADH dehydrogenase [ubiquinone] iron-sulfur p	Q9DCT2	20.131921	20.536556	19.2084344	20.286014	
1023	A2AN08-3	0.720901321	Ubr4	E3 ubiquitin-protein ligase UBR4	A2AN08-3	19.803353	19.006308	19.451812	0	
1358	Q5XJY5	0.712535076	Arcn1	Coatomer subunit delta	Q5XJY5	20.491657	20.519023	20.1085928	20.052217	
1482	Q99KF5	0.720901321	Hprt;Hprt1	Hypoxanthine-guanine phosphoribosyltransfera	Q99KF5	21.316919	19.573824	21.1814431	18.438863	
1513	Q9JKB1	0.720901321	Uchl3	Ubiquitin carboxyl-terminal hydrolase isozyme L	Q9JKB1	20.325579	19.052423	20.0963912	0	
1560	Q6PFZ4	0.720901321	Smc3	Structural maintenance of chromosomes protein	Q6PFZ4	19.889659	19.054041	19.6170773	0	
1606	Q8BNS0	0.720901321	Drg1	Developmentally-regulated GTP-binding protein	Q8BNS0	20.293522	19.206868	20.2113154	0	
1630	Q3TXV7	0.720901321	Hexa	Beta-hexosaminidase;Beta-hexosaminidase sub	Q3TXV7	20.917745	19.330021	20.6518571	0	
1753	Q78P93	0.720901321	Asah1	Acid ceramidase;Acid ceramidase subunit alpha	Q78P93	20.177005	19.854749	19.9307062	19.138753	
1788	A0A0R4J293	0.720901321	Tgm1	Protein-glutamine gamma-glutamyltransferase	A0A0R4J293	21.24142	22.734359	20.9752018	22.718909	
1823	Q3U417	0.712535076	Ppfibp1	Liprin-beta-1	Q3U417	20.905334	20.390769	20.6597296	19.544824	
1851	P15626	0.720901321	Gstm2;Gstm7	Glutathione S-transferase Mu 2;Glutathione S-t	P15626	21.393631	18.901773	21.297396	0	
1859	Q58E49	0.712535076	Hdac1;Gm100	Histone deacetylase;Histone deacetylase 1;Hist	Q58E49	21.037689	20.616245	20.6371601	20.22788	
1910	Q9ES03-2	0.720901321	Tbx20	T-box transcription factor TBX20	Q9ES03-2	21.111215	24.69251	20.8112279	24.688087	
1950	Q3UMT7	0.712535076	Hnrnpl	Heterogeneous nuclear ribonucleoprotein L	Q3UMT7	21.244589	20.321005	21.1696043	19.009269	
2035	E9PZ69	0.720901321	Tm9sf2	Transmembrane 9 superfamily member 2	E9PZ69	19.619253	18.803646	19.2016797	0	
504	Q542H2	0.723975274	Psma7;Psma8	Proteasome subunit alpha type;Proteasome sub	Q542H2	22.374285	20.604509	22.237418	18.833551	
79	Q3UMM1	0.73602349	Tubb6	Tubulin beta-6 chain	Q3UMM1	22.551806	22.222136	22.2925416	22.102081	
1843	Q8BWQ4	0.752537727	Cmtr2	Cap-specific mRNA (nucleoside-2-O-)-methyltra	Q8BWQ4	20.673159	20.320364	20.3339456	19.978915	
337	Q4FJL0	0.755284828	Rab10	Ras-related protein Rab-10	Q4FJL0	22.485553	21.368332	22.441755	21.118843	
864	Q80U79	0.755284828	Ckap5	Cytoskeleton-associated protein 5	Q80U79	20.665793	19.548988	20.4861063	0	
1249	Q99KD5	0.780443152	Unc45a	Protein unc-45 homolog A	Q99KD5	21.245348	21.129291	20.9560277	20.604094	
278	S4R2E1	0.798445998	Abcg2	ATP-binding cassette sub-family G member 2	S4R2E1	20.549796	19.566534	20.4275528	18.338447	
352	G3UWR2	0.798445998	Gigyf2	PERQ amino acid-rich with GYF domain-containi	G3UWR2	19.533967	19.213575	19.4139916	18.075577	
544	Q3U3C4	0.794837674	Sh3gl1;Sh3gl2	Endophilin-A2;Endophilin-A1	Q3U3C4	20.694189	19.644449	20.5915906	0	
726	Q3TZI3	0.798745339	Sf1	Splicing factor 1	Q3TZI3	20.858833	19.751687	20.6194129	0	
732	Q8BI72	0.798445998	Cdkn2aip	CDKN2A-interacting protein	Q8BI72	19.4319	18.990335	18.8940737	0	
828	Q71V27	0.798745339	Csrp2	Cysteine and glycine-rich protein 2	Q71V27	21.296719	20.146312	21.0302972	19.342723	
944	Q3UDE2	0.798445998	Ttll12	Tubulin--tyrosine ligase-like protein 12	Q3UDE2	20.108931	19.352621	19.9592481	0	
978	Q71FD7	0.798445998	Fblim1	Filamin-binding LIM protein 1	Q71FD7	20.141398	20.123265	19.5856163	19.610515	
1092	Q3UI56	0.798445998	Anxa6	Annexin;Annexin A6	Q3UI56	19.99042	20.225033	19.0187055	19.917708	
1105	Q3U1N3	0.798745339	Rras	Ras-related protein R-Ras	Q3U1N3	21.225851	20.428576	21.0342703	19.449692	
1289	Q3U9G9	0.798445998	Lbr	Lamin-B receptor	Q3U9G9	21.094135	20.347329	21.0139246	19.725307	
1366	Q3TE85	0.798445998	Hmga1;mCG_	High mobility group protein HMG-I/HMG-Y	Q3TE85	20.720368	20.019687	20.421675	19.657714	
1401	Q9QYT9	0.798445998	Prnpb;Prnp	Major prion protein	Q9QYT9	21.929142	22.138507	21.8789329	22.05109	
1436	O55222	0.798745339	Ilk	Integrin-linked protein kinase	O55222	20.350238	20.046557	19.8153861	18.689613	
1538	A2AEG2	0.798745339	Ofd1	Oral-facial-digital syndrome 1 protein homolog	A2AEG2	20.355495	19.802982	20.048336	18.141656	
1552	A0A0N4SVQ1	0.798445998	Ndufa4	Cytochrome c oxidase subunit NDUFA4	A0A0N4SVQ1	22.813602	23.297461	22.7990554	23.259315	
1768	Q8K1J6-2	0.798445998	Trnt1	CCA tRNA nucleotidyltransferase 1, mitochondri	Q8K1J6-2	20.377276	20.072943	19.996981	19.342297	
1774	Q3TIL5	0.798445998	Stx12	Syntaxin-12	Q3TIL5	20.11202	19.188974	19.8648789	17.570485	
1982	Q8VDM6-2	0.798445998	Hnrnpul1	Heterogeneous nuclear ribonucleoprotein U-like	Q8VDM6-2	19.857969	18.82671	19.5687081	0	
2124	E9Q715	0.798445998	Luc7l2;Luc7l	Putative RNA-binding protein Luc7-like 2;Putativ	E9Q715	20.882679	19.709027	20.5901076	18.741981	
1099	Q8BLV4	0.799407187	Lrrfip1	Leucine-rich repeat flightless-interacting protein	Q8BLV4	21.902052	21.671068	21.6442889	21.336204	
1307	P97350	0.809125348	Pkp1	Plakophilin-1	P97350	21.764216	22.08911	21.5223494	21.833357	
965	Q3U8S5	0.814269221	Capn2	Calpain-2 catalytic subunit	Q3U8S5	21.919602	21.732762	21.6086641	21.070981	
226	D3YVN7	0.826169907	Gm9755;Tufm	Elongation factor Tu;Elongation factor Tu, mitoc	D3YVN7	22.760651	23.135445	22.5693604	22.892263	
1618	P62274	0.833514213	Rps29	40S ribosomal protein S29	P62274	20.999775	21.531084	20.7337015	21.356389	
74	A0A0R4J0T7	0.838059801	Gas1	Growth arrest-specific protein 1	A0A0R4J0T7	20.99014	20.982094	20.5933488	20.488251	
103	B7ZCJ0	0.838059801	Arhgap21	Rho GTPase-activating protein 21	B7ZCJ0	20.941172	20.569545	20.7191308	18.973029	
94	Q3UE51	0.84283599	Hk1	Hexokinase;Hexokinase-1	Q3UE51	20.696564	19.72937	20.4748439	0	
551	Q8BMZ7	0.84283599	Vps28	Vacuolar protein sorting-associated protein 28 H	Q8BMZ7	20.53929	20.192352	19.9355974	18.532272	
620	Q3UPL0-2	0.84283599	Sec31a	Protein transport protein Sec31A	Q3UPL0-2	20.330818	19.945002	19.6791233	18.31079	
99	Q7TMN7	0.883056978	Anxa4	Annexin;Annexin A4	Q7TMN7	21.880645	21.016763	21.6303828	20.360265	
100	O35864	0.885714286	Cops5	COP9 signalosome complex subunit 5	O35864	17.428105	17.706314	16.5041549	0	
199	Q3UDW8-2	0.885714286	Hgsnat	Heparan-alpha-glucosaminide N-acetyltransfera	Q3UDW8-2	19.243429	18.103825	18.8290178	0	
201	Q8BNY6	0.885714286	Ncs1	Neuronal calcium sensor 1	Q8BNY6	19.615787	17.925508	19.6157867	0	
422	P61967	0.885714286	Ap1s1	AP-1 complex subunit sigma-1A	P61967	19.217312	18.285337	18.4032556	0	
474	Q9CRT8	0.885714286	Xpot	Exportin-T	Q9CRT8	17.730332	17.735855	17.0973886	0	
515	Q3U716	0.885714286	Cars	Cysteine--tRNA ligase, cytoplasmic	Q3U716	18.740005	18.842857	18.3796314	18.243197	
524	A0A1B0GRB6	0.885714286			A0A1B0GRB6	21.733556	22.375538	21.6805752	22.346648	
546	Q61391	0.885714286	Mme	Neprilysin	Q61391	17.9904	18.930641	0	18.456367	
609	Q52L78	0.885714286	Cryab	Alpha-crystallin B chain	Q52L78	19.554999	17.89898	19.3070551	0	
631	B2RRE8	0.885714286	Cd2ap	CD2-associated protein	B2RRE8	20.366515	19.055553	20.1217401	0	
724	P68181-2	0.885714286	Prkacb	cAMP-dependent protein kinase catalytic subun	P68181-2	18.257358	17.363144	17.7368117	0	
776	Q3UDZ1	0.885714286	Rhog	Rho-related GTP-binding protein RhoG	Q3UDZ1	19.860015	18.557769	19.633485	0	
867	Q9CWD7	0.885714286	Wdyhv1	Protein N-terminal glutamine amidohydrolase	Q9CWD7	18.524134	17.899159	17.9754574	0	
888	Q3ULL5	0.861987059	Eif2s2	Eukaryotic translation initiation factor 2 subunit	Q3ULL5	21.94705	21.571701	21.7481488	20.978119	
893	Q8K0E2	0.885714286	Exoc3	Exocyst complex component 3	Q8K0E2	18.350977	17.585548	17.8260104	0	
900	Q8R0W6	0.878477078	Ndfip1	NEDD4 family-interacting protein 1	Q8R0W6	20.745808	19.254057	20.6336525	0	
1015	E9PYT3	0.885714286	Atl3	Atlastin-3	E9PYT3	18.607178	17.933212	18.2218196	0	
1056	F6UFG6	0.878477078	Anp32a	Acidic leucine-rich nuclear phosphoprotein 32 fa	F6UFG6	20.470121	18.872745	20.2715834	0	

1076	Q3TW01	0.885714286	Rab32	Ras-related protein Rab-32	Q3TW01	17.947313	18.259283	16.6573183	0	
1079	Q9CTI8	0.885714286	Atp6v1d	V-type proton ATPase subunit D	Q9CTI8	19.438524	17.903697	19.2403771	0	
1152	P97384	0.885714286	Anxa11	Annexin A11;Annexin	P97384	21.542118	20.378299	21.5043517	20.203224	
1189	B2RXP1	0.878477078	Lrch3	Leucine-rich repeat and calponin homology dom	B2RXP1	19.625281	19.817169	18.8724175	18.727322	
1234	A0A1B0GT68	0.885714286	Ifitm2	Interferon-induced transmembrane protein 2	A0A1B0GT68	20.181428	19.487939	20.1814275	19.341628	
1240	Q9JMJ6	0.878477078	Stx7	Syntaxin-7	Q9JMJ6	20.179734	19.157069	20.0189487	0	
1246	A0A0G2JFQ5	0.885714286	Mccc1	Methylcrotonoyl-CoA carboxylase subunit alpha	A0A0G2JFQ5	19.269318	18.337181	18.9680829	0	
1256	Q3TG71	0.885714286	Cndp2	Cytosolic non-specific dipeptidase	Q3TG71	19.059784	17.686178	18.8263286	0	
1297	B1AX98	0.885714286	Lrrc47	Leucine-rich repeat-containing protein 47	B1AX98	23.958692	18.482116	23.9529116	0	
1327	F6T4V9	0.885714286	Mycbp2	E3 ubiquitin-protein ligase MYCBP2	F6T4V9	18.712953	18.268396	18.0379157	0	
1355	Q9ESR1	0.878477078	Rheb	GTP-binding protein Rheb	Q9ESR1	20.36546	19.392202	19.902438	0	
1359	Q04857	0.885714286	Col6a1	Collagen alpha-1(VI) chain	Q04857	19.158231	18.66586	18.7382637	18.019922	
1390	Q9CPT4	0.885714286	Mydgf	Myeloid-derived growth factor	Q9CPT4	19.81449	18.164726	19.5811507	0	
1462	Q3TIU7	0.885714286	Ndufs1	NADH-ubiquinone oxidoreductase 75 kDa subun	Q3TIU7	19.859159	19.414778	19.4030612	18.276436	
1507	Q9CVJ6	0.885714286	Tram1	Translocating chain-associated membrane prote	Q9CVJ6	18.187053	17.9356	17.3772791	0	
1543	Q8R105	0.885714286	Vps37c	Vacuolar protein sorting-associated protein 37C	Q8R105	18.22057	18.377899	17.8308614	0	
1551	Q569X3	0.885714286	Snrpc	U1 small nuclear ribonucleoprotein C	Q569X3	20.054014	17.797065	19.9356435	0	
1568	D3YYD5	0.885714286	Vps29	Vacuolar protein sorting-associated protein 29	D3YYD5	19.705481	19.039241	19.3120764	18.690369	
1593	Q9D8V0-3	0.885714286	Hm13;H13	Minor histocompatibility antigen H13	Q9D8V0-3	18.709458	17.734529	18.470301	0	
1607	Q00780	0.885714286	Col8a1	Collagen alpha-1(VIII) chain;Vastatin	Q00780	20.533827	21.536445	20.4709896	21.506022	
1673	Q9WTS9	0.885714286	Rbms2	RNA-binding motif, single-stranded-interacting p	Q9WTS9	19.022478	18.704037	18.6646372	18.187554	
1682	Q545I6	0.885714286	Ctsz	Cathepsin Z	Q545I6	18.013797	18.387005	17.5464626	0	
1720	Q8BKI9	0.885714286	Naa25	N-alpha-acetyltransferase 25, NatB auxiliary sub	Q8BKI9	19.041684	17.742351	18.7734151	0	
1729	Q3UL78	0.878477078	Cdc42		Q3UL78	20.613384	19.416433	20.4761836	17.904295	
1772	Q3UYQ0	0.885714286	Prps1l3;Prps1	Ribose-phosphate pyrophosphokinase 1;Ribose	Q3UYQ0	18.238067	17.860957	17.7547288	0	
1784	Q7JD02	0.885714286	Atp8;mt-Atp8	ATP synthase protein 8	Q7JD02	20.35583	18.210713	20.2475118	0	
1829	Q8VCE7	0.885714286	Tmed5	Transmembrane emp24 domain-containing pro	Q8VCE7	20.855776	18.73079	20.8557763	0	
1871	Q059U9	0.885714286	Akap8	A-kinase anchor protein 8	Q059U9	18.566679	17.918994	18.1452239	0	
1873	Q641N8	0.885714286	Nedd4l	E3 ubiquitin-protein ligase NEDD4-like	Q641N8	20.090552	19.228078	20.053308	18.676677	
1879	E9QMH7	0.885714286	Ikbip	Inhibitor of nuclear factor kappa-B kinase-intera	E9QMH7	20.105343	19.060091	20.0550977	18.740468	
1880	Q8K335	0.885714286	Grwd1	Glutamate-rich WD repeat-containing protein 1	Q8K335	19.118677	17.914957	18.8076164	0	
1901	P32020-2	0.885714286	Scp2	Non-specific lipid-transfer protein	P32020-2	18.71607	18.006314	18.2481682	0	
1903	Q8BWK1	0.885714286	Ccdc93	Coiled-coil domain-containing protein 93	Q8BWK1	21.163829	18.135858	21.0569952	0	
1916	Q3UEN1	0.885714286	Upp2	Uridine phosphorylase;Uridine phosphorylase 2	Q3UEN1	22.929988	23.907711	22.9032765	23.894047	
1930	B1AT24	0.885714286	Dusp14	Dual specificity protein phosphatase 14	B1AT24	18.844565	18.246284	18.4454093	0	
1952	A0A0N4SVH2	0.885714286	Clasp1	CLIP-associating protein 1	A0A0N4SVH2	19.024171	18.064415	19.0241714	0	
1965	Q6Y642	0.885714286	Gm20498;Syn	Synaptojanin-2-binding protein	Q6Y642	19.537487	18.153487	19.5374867	0	
2001	E9Q7G1	0.885714286	Tmed7		E9Q7G1	19.983721	19.347293	19.8062203	19.075648	
2005	G5E8R4	0.885714286	Ppp6r3	Serine/threonine-protein phosphatase 6 regulat	G5E8R4	18.905555	18.372246	18.3482264	0	
2017	Q8K2P1	0.885714286	Lurap1l	Leucine rich adaptor protein 1-like	Q8K2P1	17.906893	17.851427	17.1970636	0	
2039	A0A0N4SUM7	0.885714286	Tmem176b	Transmembrane protein 176B	A0A0N4SUM7	18.570817	18.113916	17.9966468	0	
2050	Q3TV50	0.885714286	Baiap2	Brain-specific angiogenesis inhibitor 1-associate	Q3TV50	19.405077	17.908391	19.2593959	0	
2052	D3Z2X5	0.885714286	Pafah1b3	Platelet-activating factor acetylhydrolase IB sub	D3Z2X5	18.361205	17.510198	17.7490922	0	
2053	Q9QXS1-6	0.885714286	Plec	Plectin	Q9QXS1-6	18.27895	19.399058	0	19.067935	
2058	B9EHV2	0.885714286	Spata13		B9EHV2	23.722822	18.116982	23.7127505	0	
2059	J3QJX0	0.878477078	Cul4b	Cullin-4B	J3QJX0	19.794285	19.024397	19.3698855	0	
2077	Q62261-2	0.885714286	Sptbn1	Spectrin beta chain, non-erythrocytic 1	Q62261-2	20.299345	20.256648	20.1051857	20.209779	
2078	Q6ZVL3	0.885714286			Q6ZVL3	26.553217	29.01101	26.5532167	29.010768	
2084	F6TFN2	0.885714286	Lmo7		F6TFN2	20.748364	20.840783	20.6615448	20.737082	
2104	Q62379	0.885714286	Snrpb2	U2 small nuclear ribonucleoprotein B	Q62379	19.503005	18.465289	19.1026032	0	
2110	Q9D6K2	0.885714286	Tagln3	Transgelin;Transgelin-3	Q9D6K2	19.699397	17.971147	19.5092848	0	
2112	Q8R527	0.885714286	Rhoq	Rho-related GTP-binding protein RhoQ	Q8R527	19.694832	18.155425	19.4742232	0	
2115	Q61838	0.885714286	A2m	Alpha-2-macroglobulin;Alpha-2-macroglobulin 1	Q61838	18.143065	19.497903	0	19.259288	
2125	Q9D1I6	0.885714286	Mrpl14	39S ribosomal protein L14, mitochondrial	Q9D1I6	19.815432	19.605206	19.6986295	19.528834	
2135	Q8BH59	0.878477078	Slc25a12	Calcium-binding mitochondrial carrier protein Ar	Q8BH59	20.157365	20.100676	19.8598053	19.557135	
2137	Q9Z2L2	0.885714286	Hspb1		Q9Z2L2	21.08349	21.281966	21.0397606	21.257894	
2141	Q9CT54	0.885714286	Fam21	WASH complex subunit FAM21	Q9CT54	19.791593	18.325593	19.7915935	0	
2144	E9Q4B9	0.885714286	Arpp19;Ensa	cAMP-regulated phosphoprotein 19;Alpha-endo	E9Q4B9	19.220919	17.83294	18.9577534	0	
254	Q61033	0.887385935	Tmpo	Lamina-associated polypeptide 2, isoforms alph	Q61033	20.258817	20.382482	19.6166138	19.835623	
1547	D3Z5I1	0.887385935	Zc3hav1	Zinc finger CCCH-type antiviral protein 1	D3Z5I1	20.312118	19.826864	19.9885623	0	
812	Q642L7	0.890108936	Rps27a	Ubiquitin-40S ribosomal protein S27a;Ubiquitin;	Q642L7	26.660372	26.151179	26.6418469	26.121497	
128	E9PYF4	0.923134334	Lmo7		E9PYF4	24.978144	25.346445	24.7838289	25.182061	
14	P62878	0.932300503	Rbx1	E3 ubiquitin-protein ligase RBX1;E3 ubiquitin-pro	P62878	20.777219	19.745832	20.3686139	17.528182	
874	Q80XJ7	0.932300503	Akr1a1	Alcohol dehydrogenase [NADP(+)]	Q80XJ7	20.598994	20.180679	19.5957576	19.281791	
1826	A0A0G2JG10	0.932300503	Dhx15	Pre-mRNA-splicing factor ATP-dependent RNA H	A0A0G2JG10	20.695354	19.990888	20.556203	17.78452	
810	P54071	0.939908596	Idh2	Isocitrate dehydrogenase [NADP], mitochondria	P54071	21.838924	21.569823	21.6864869	21.276006	
533	Q8BX02	0.943010334	Kank2	KN motif and ankyrin repeat domain-containing	Q8BX02	21.355638	21.591279	20.7956026	21.342722	
196	E9PWY9	0.946695469	Farsa	Phenylalanine--tRNA ligase alpha subunit	E9PWY9	22.067967	22.470285	21.7577551	22.271625	
177	A2AE89	0.959129759	Gstm1	Glutathione S-transferase Mu 1	A2AE89	20.549787	19.447585	20.2366367	0	
372	Q8VI75	0.959129759	Ipo4	Importin-4	Q8VI75	19.746342	19.297296	19.3659095	0	
531	Q0VE46	0.959129759	Myadm	Myeloid-associated differentiation marker	Q0VE46	24.655034	23.535303	24.6442361	23.529163	
835	P50543	0.959129759	S100a11	Protein S100-A11	P50543	22.296492	20.559095	22.2570614	20.319503	
1018	Q9Z1Z0	0.959129759	Uso1	General vesicular transport factor p115	Q9Z1Z0	19.923695	19.298582	19.6986368	0	
1173	A0A0R4J2D2	0.959129759	Sulf1	Extracellular sulfatase Sulf-1	A0A0R4J2D2	19.412677	19.611173	18.3023093	19.022929	
1408	Q3UKT3	0.959129759	Oat	Ornithine aminotransferase, mitochondrial	Q3UKT3	19.525849	19.449242	19.23468	18.353639	
1520	Q3UBU9	0.959129759	Fkbp3	Peptidyl-prolyl cis-trans isomerase;Peptidyl-pro	Q3UBU9	20.746403	19.040127	20.549725	0	
1811	Q3TML0	0.959129759	Pdia6		Q3TML0	22.005779	21.598467	21.9749263	21.508349	
75	Q78T54-2	1	Vma21	Vacuolar ATPase assembly integral membrane	Q78T54-2	19.517279	17.809767	19.3279323	0	
139	B2RSW8	0.978489208	Pcm1	Pericentriolar material 1 protein	B2RSW8	25.353819	25.279256	25.2111379	25.117832	
155	Q05BC8	1	Nbr1	Next to BRCA1 gene 1 protein	Q05BC8	19.421588	18.622682	19.0470763	0	
644	Q3UYR3	1	Cbfb	Core-binding factor subunit beta	Q3UYR3	19.370581	18.163425	18.9824781	0	
654	Q3UV17	1	Krt76	Keratin, type II cytoskeletal 2 oral	Q3UV17	25.371742	22.407262	25.3643826	22.345255	
715	Q3TU85	0.976965249	Hspa1b;Hspa1	Heat shock 70 kDa protein 1A;Heat shock 70 kD	Q3TU85	21.235373	20.615046	21.1586014	19.569357	
718	Q8R138	1	Tmem119	Transmembrane protein 119	Q8R138	19.783526	17.917181	19.6611058	0	
786	Q5SX46	1	Slc25a11	Mitochondrial 2-oxoglutarate/malate carrier pro	Q5SX46	18.905402	17.900092	18.7099663	0	
820	O88840	1	Fbn1	Fibrillin-1	O88840	18.29617	18.591387	0	17.54387	
891	Q8BSL7	1	Arf2	ADP-ribosylation factor 2	Q8BSL7	18.763853	19.253628	18.763853	19.103782	
907	B2RRC5	1	Rabgap1;Rabg	Rab GTPase-activating protein 1	B2RRC5	18.119788	17.812583	17.8211068	0	

1013	A0A1B0GS22	0.977402191	Rcn3	Reticulocalbin-3	A0A1B0GS22	21.511215	20.852482	21.2841155	20.417781	
1048	Q9CQ60	1	Pgls	6-phosphogluconolactonase	Q9CQ60	18.565124	17.959741	18.2651989	0	
1065	O89054	1	Actb		O89054	23.550438	23.024728	23.5353275	23.003821	
1085	Q9D1M7	1	Fkbp11	Peptidyl-prolyl cis-trans isomerase FKBP11	Q9D1M7	20.096697	17.601203	19.9909769	0	
1095	F6UK66	1	Ccdc50	Coiled-coil domain-containing protein 50	F6UK66	18.379694	18.015952	18.0617942	0	
1140	Q3V199	1	Bzw2	Basic leucine zipper and W2 domain-containing	Q3V199	19.244772	18.149069	19.010163	0	
1142	A2AWN8	1	Ythdf1;Ythdf3	YTH domain-containing family protein 1;YTH do	A2AWN8	19.195819	18.071585	18.9459788	0	
1216	Q8BML1	1	Mical2	Protein-methionine sulfoxide oxidase MICAL2	Q8BML1	20.118239	20.369084	19.3181538	19.647646	
1275	Q91VA7	1	Idh3b	Isocitrate dehydrogenase [NAD] subunit, mitoch	Q91VA7	19.360263	18.040145	19.1434163	0	
1294	Q6NXY1	1	Tbc1d31	TBC1 domain family member 31	Q6NXY1	20.928682	19.681591	20.7202292	0	
1314	Q8BHC2	1	Ist1	IST1 homolog	Q8BHC2	19.143274	17.99961	18.9622188	0	
1373	D6RFQ0	1	Dcps	m7GpppX diphosphatase	D6RFQ0	18.420462	17.916724	17.6091092	0	
1433	P52503	1	Ndufs6	NADH dehydrogenase [ubiquinone] iron-sulfur p	P52503	19.652706	18.150604	19.6139868	0	
1472	Q0VGR5	0.977402191	Cdk5rap2	CDK5 regulatory subunit-associated protein 2	Q0VGR5	20.177686	20.508018	18.6812314	19.707063	
1519	Q9D077	1	Mcm4	DNA replication licensing factor MCM4	Q9D077	19.769987	18.319042	19.7004621	0	
1592	Q3UGN9	1	Stam	Signal transducing adapter molecule 1	Q3UGN9	19.200078	18.395751	19.0903548	0	
1660	Q8R0W0	1	Eppk1	Epiplakin	Q8R0W0	19.288482	18.972412	18.4804208	18.396495	
1701	Q9D9C1	1	Atad1	ATPase family AAA domain-containing protein 1	Q9D9C1	19.929494	18.530559	19.8005624	0	
1731	G3UZW7	1	Magohb;Mag	Protein mago nashi homolog 2;Protein mago na	G3UZW7	20.450624	17.968865	20.368859	0	
1735	Q3U754	1	Samhd1	Deoxynucleoside triphosphate triphosphohydro	Q3U754	19.376927	18.921091	19.275937	17.373216	
1741	Q3TLE2	0.977402191	Fermt2	Fermitin family homolog 2	Q3TLE2	24.230589	20.654551	24.2006472	19.678394	
1766	Q3U3G2	1	Rnf126	E3 ubiquitin-protein ligase RNF126	Q3U3G2	19.271774	18.111145	18.9154924	0	
1845	Q80X54	1	Snx4	Sorting nexin-4	Q80X54	17.543177	18.428217	0	18.115321	
1863	Q69ZZ8	1	Adgrf5		Q69ZZ8	18.993829	17.831177	18.4876641	0	
1866	U5TAU7	1			U5TAU7	20.788299	20.380371	20.7485453	20.157659	
1882	A2A5R8	1	Stau1	Double-stranded RNA-binding protein Staufen h	A2A5R8	19.612734	18.47634	19.4598873	0	
1884	Q545I9	1	S100a6	Protein S100;Protein S100-A6	Q545I9	23.70364	23.354437	23.6625589	23.298348	
1889	Q8BJ71	1	Nup93	Nuclear pore complex protein Nup93	Q8BJ71	20.227173	19.661117	19.7309847	18.927497	
1925	Q9WTI7-2	1	Myo1c	Unconventional myosin-Ic	Q9WTI7-2	19.11445	19.12925	18.8979318	18.76289	
1944	Q9D0M5	1	Dynll2	Dynein light chain 2, cytoplasmic	Q9D0M5	19.27716	18.11035	19.1374922	0	
1962	A0A0N4SW38	1	Olfr704;Olfr694	Olfr700;Olfr699;Olfr1395;Olfr697;Olfr703	A0A0N4SW38	18.064154	22.685162	0	22.669079	
2030	F8WGM8	1	Actg1		F8WGM8	20.194637	17.792782	20.1031018	0	
2072	B6YY24	1	D15Ertd621e;	Protein FAM91A1	B6YY24	18.210474	17.981421	17.5902921	0	
2073	Q3UGS4	1	Fam195b	Protein FAM195B	Q3UGS4	19.155351	18.4303	18.9691263	0	
2103	Q545H7	1	S100a13	Protein S100-A13	Q545H7	19.966037	19.413892	19.9660373	19.236226	
2126	Q9DAK9	1	Phpt1	14 kDa phosphohistidine phosphatase	Q9DAK9	19.124996	17.983022	18.7840576	0	
2134	Q9CR60	1	Golt1b	Vesicle transport protein GOT1B	Q9CR60	19.562137	17.956436	19.3373692	0	
1	E0CY46		Ipo9	Importin-9	E0CY46					
149	Q3TMB8		Adsl	Adenylosuccinate lyase	Q3TMB8					
158	Q6PE06		Klhdc4	Kelch domain-containing protein 4	Q6PE06					
165	Q80Y83-12		Dixdc1	Dixin	Q80Y83-12					
180	G5E8G6		Myo5b	Unconventional myosin-Vb	G5E8G6	17.948639	22.869228	0	22.858354	
193	Q3UDG2		Wars	Tryptophan--tRNA ligase, cytoplasmic;T1-TrpRS	Q3UDG2					
216	O88860		Clca3a1;Clca3a2		O88860					
220	Q4VA93		Prkca;Pkca;Pr	Protein kinase C;Protein kinase C alpha type;Pro	Q4VA93					
264	Q80UN7		Prkcd	Protein kinase C delta type;Protein kinase C delt	Q80UN7					
286	Q922Y1		Ubxn1	UBX domain-containing protein 1	Q922Y1					
291	Q8BH24		Tm9sf4	Transmembrane 9 superfamily member 4	Q8BH24					
303	Q7TT37		Ikbkap	Elongator complex protein 1	Q7TT37					
329	Q9CQU5		Zwint	ZW10 interactor	Q9CQU5					
354	Q78IK2		Usmg5	Up-regulated during skeletal muscle growth pro	Q78IK2					
374	Q9Z0N2		Eif2s3y	Eukaryotic translation initiation factor 2 subunit	Q9Z0N2					
412	Q3UPP3		mt-Nd4;ND4;I	NADH-ubiquinone oxidoreductase chain 4	Q3UPP3					
418	Q3UZL8		Srp72	Signal recognition particle subunit SRP72	Q3UZL8					
419	P97807-2		Fh;Fh1	Fumarate hydratase, mitochondrial	P97807-2					
420	Q3UM45		Ppp1r7	Protein phosphatase 1 regulatory subunit 7	Q3UM45					
436	Q8BG92-2		Clvs2	Clavesin-2	Q8BG92-2					
457	Q5D0D8		Gcnt1	Beta-1,3-galactosyl-O-glycosyl-glycoprotein bet	Q5D0D8					
468	A0A1B0GR11		Taldo1	Transaldolase	A0A1B0GR11	25.882074	18.218805	25.8795777	0	
479	D6REU0		Fam188a	Protein FAM188A	D6REU0					
486	Q3ULS2		Smc2	Structural maintenance of chromosomes protein	Q3ULS2					
489	P48759		Ptx3	Pentraxin-related protein PTX3	P48759					
500	B7ZCU2		Abi1	Abl interactor 1	B7ZCU2					
516	Q3UJV1		Ccdc61	Coiled-coil domain-containing protein 61	Q3UJV1					
522	Q9WTI7		Myo1c	Unconventional myosin-Ic	Q9WTI7					
540	P47857		Pfkm	ATP-dependent 6-phosphofructokinase, muscle	P47857					
572	Q9JKK7-3		Tmod2	Tropomodulin-2	Q9JKK7-3					
579	Q3UG63		Ascc2	Activating signal cointegrator 1 complex subunit	Q3UG63					
580	D3YUP1		Carm1	Histone-arginine methyltransferase CARM1	D3YUP1					
598	G3UXY0		Psme1	Proteasome activator complex subunit 1	G3UXY0					
615	D3YTS4		Lamtor2	Ragulator complex protein LAMTOR2	D3YTS4					
622	Q9WV42		Ncoa4;Gm6768		Q9WV42					
637	E9Q7P0		Dnah17	Dynein heavy chain 17, axonemal	E9Q7P0	18.770228	18.380982	18.7702277	0	
675	A0A0N4SV16		Antxr1	Anthrax toxin receptor 1	A0A0N4SV16					
720	E9QKX5		Vps13b	Vacuolar protein sorting-associated protein 13B	E9QKX5	21.665604	17.971743	21.5308263	0	
725	E9Q6F4		Npepps	Puromycin-sensitive aminopeptidase	E9Q6F4					
742	E9PZ92		Exoc5	Exocyst complex component 5	E9PZ92					
773	G3UZ26		Shmt1	Serine hydroxymethyltransferase;Serine hydrox	G3UZ26					
778	Q8BK67		Rcc2	Protein RCC2	Q8BK67					
791	A1IGX7		Dsg4	Desmoglein-4	A1IGX7					
793	D3YUK4		Ndufb10	NADH dehydrogenase [ubiquinone] 1 beta subc	D3YUK4					
816	H9BUG1				H9BUG1	19.47824	17.59114	19.0873812	0	
825	G0XXC3				G0XXC3	19.81912	18.875898	19.3921912	0	
833	Q3TDD1		Ldlr	Low-density lipoprotein receptor	Q3TDD1					
842	Q3TVS1		App	Amyloid beta A4 protein;N-APP;Soluble APP-alp	Q3TVS1					
849	Q3UHN1		Slit3	Slit homolog 3 protein	Q3UHN1					
854	A2AAN2		Srp68	Signal recognition particle subunit SRP68	A2AAN2					
870	Q3TXM9		Srsf9	Serine/arginine-rich splicing factor 9	Q3TXM9					

875	B7U582		Hspa2	Heat shock-related 70 kDa protein 2	B7U582					
878	Q8BGX0-3		Trim23	E3 ubiquitin-protein ligase TRIM23	Q8BGX0-3					
909	F8VPX1		Usp7	Ubiquitin carboxyl-terminal hydrolase;Ubiquitin	F8VPX1					
916	Q9CXZ9		Galk1	Galactokinase	Q9CXZ9					
940	D3YVV1		Tmem30a	Cell cycle control protein 50A	D3YVV1					
948	A2AWI9		Sh3glb2	Endophilin-B2	A2AWI9					
951	Q8R3B1		Plcd1	1-phosphatidylinositol 4,5-bisphosphate phosph	Q8R3B1					
977	Q3TPW0		Trub1	Probable tRNA pseudouridine synthase 1	Q3TPW0					
986	Q543K9		Pnp;Pnp2	Purine nucleoside phosphorylase	Q543K9					
988	A0A087WRV4		Srgap2	SLIT-ROBO Rho GTPase-activating protein 2	A0A087WRV4					
994	Q9D6J6-2		Ndufv2	NADH dehydrogenase [ubiquinone] flavoprotein	Q9D6J6-2					
1000	Q3TIJ9		Actb		Q3TIJ9	18.132256	17.68026	17.8415791	0	
1009	Q62422		Ostf1	Osteoclast-stimulating factor 1	Q62422					
1011	Q3UJS6		Cpne3;Cpne9;	Copine-3;Copine-2;Copine-9;Copine-6;Copine-4	Q3UJS6					
1024	Q99J29		Scpep1	Retinoid-inducible serine carboxypeptidase	Q99J29					
1026	Q2TBE6		Pi4k2a	Phosphatidylinositol 4-kinase type 2-alpha	Q2TBE6					
1034	Q04447		Ckb	Creatine kinase B-type	Q04447					
1046	Q9QY23		Pkp3	Plakophilin-3	Q9QY23					
1054	Q8K2Q0		Commd9	COMM domain-containing protein 9	Q8K2Q0					
1059	Q3UIJ3		Actc1		Q3UIJ3					
1061	Q9ET22		Dpp7	Dipeptidyl peptidase 2	Q9ET22					
1062	Q6ZPX7		mKIAA1101;C	Serine/threonine-protein kinase OSR1	Q6ZPX7					
1064	Q3UAA9		Actb		Q3UAA9	17.799701	20.255365	0	20.033416	
1071	E9Q9H2		Dnajc2	DnaJ homolog subfamily C member 2;DnaJ hom	E9Q9H2					
1088	Q541Z2		Fnta	Protein farnesyltransferase/geranylgeranyltran	Q541Z2					
1093	Q3UPG1		Hmbs	Porphobilinogen deaminase	Q3UPG1					
1094	Q3V1G6		Kat6a	Histone acetyltransferase;Histone acetyltransfe	Q3V1G6					
1096	A0A0R4J0H7		Ncapd2	Condensin complex subunit 1	A0A0R4J0H7					
1107	Q80ZP8		Manf	Mesencephalic astrocyte-derived neurotrophic	Q80ZP8					
1111	H3BJE0		Grn	Granulins;Acrogranin;Granulin-1;Granulin-2;Gra	H3BJE0					
1114	A0A140LIS2		Tbc1d17	TBC1 domain family member 17	A0A140LIS2					
1118	Q8BH99		Nck1	Cytoplasmic protein NCK1	Q8BH99					
1120	B7ZN40		Arl15	ADP-ribosylation factor-like protein 15	B7ZN40					
1123	Q9QX52		Pex14	Peroxisomal membrane protein PEX14	Q9QX52					
1145	Q68FG2		Sptbn2		Q68FG2					
1156	Q3UE75		Gpnmb	Transmembrane glycoprotein NMB	Q3UE75	18.440333	17.559427	18.2475858	0	
1158	A0A0N4SVV4		Rarres2	Retinoic acid receptor responder protein 2	A0A0N4SVV4					
1168	B2RSV4		Sf3b3	Splicing factor 3B subunit 3	B2RSV4					
1172	P97346		Nxn	Nucleoredoxin	P97346					
1180	Q4FK16		Plscr1	Phospholipid scramblase 1	Q4FK16					
1182	Q9D051		Pdhb	Pyruvate dehydrogenase E1 component subuni	Q9D051					
1212	D3Z3F8		Spg20	Spartin	D3Z3F8					
1213	P24527		Lta4h	Leukotriene A-4 hydrolase	P24527					
1215	P99028		Uqcrh	Cytochrome b-c1 complex subunit 6, mitochond	P99028					
1224	G3X956		Supt16;Supt1	FACT complex subunit SPT16	G3X956					
1225	A0A0R4IZY0		Thop1	Thimet oligopeptidase	A0A0R4IZY0					
1229	E9Q8N1		Ttn	Titin	E9Q8N1	22.225327	18.065513	22.2113535	0	
1230	E9Q1F5		Myo5c		E9Q1F5					
1232	Q80Y35		Numa1		Q80Y35	23.125768	17.519677	23.1069499	0	
1233	A2AQB2		Neb		A2AQB2	19.89878	19.410533	19.398886	0	
1239	Q6P5I3		Adh5	S-(hydroxymethyl)glutathione dehydrogenase;	Q6P5I3					
1259	Q3U9N0		Slc12a4	Solute carrier family 12 member 4	Q3U9N0					
1267	Q9MQW4		COX3;mt-Co3	Cytochrome c oxidase subunit 3	Q9MQW4					
1272	Q8BVK3		Uap1l1	UDP-N-acetylhexosamine pyrophosphorylase-li	Q8BVK3					
1277	Q8CHH8		Rb1cc1	RB1-inducible coiled-coil protein 1	Q8CHH8					
1282	Q9QYI6		Dnajb9	DnaJ homolog subfamily B member 9	Q9QYI6					
1287	Q4FJY3		Tgoln1;Tgoln2	Trans-Golgi network integral membrane protein	Q4FJY3					
1293	Q3UAF7		Actb		Q3UAF7	19.298301	18.182938	18.7000644	0	
1313	P11531		Dmd	Dystrophin	P11531	18.304926	17.524249	17.628341	0	
1317	E9Q9M5		Usp19	Ubiquitin carboxyl-terminal hydrolase;Ubiquitin	E9Q9M5					
1321	Q3UQD0		Tomm40	Mitochondrial import receptor subunit TOM40 H	Q3UQD0					
1334	Q62084		Ppp1r14b	Protein phosphatase 1 regulatory subunit 14B	Q62084					
1340	Q8CE08		Acpp	Prostatic acid phosphatase	Q8CE08					
1345	Q8BKH0		Rap1gds1		Q8BKH0					
1357	H3BIV7		Ppp2r1b	Serine/threonine-protein phosphatase 2A 65 kD	H3BIV7					
1376	B8JJ53		Trim26	Tripartite motif-containing protein 26	B8JJ53					
1396	E9Q197		Glod4	Glyoxalase domain-containing protein 4	E9Q197					
1397	E9QPJ5		D5Ertd579e;K	Uncharacterized protein KIAA0232	E9QPJ5					
1398	F6XLV1		E030010N08Rik		F6XLV1					
1403	Q8WTY4		Ciapin1	Anamorsin	Q8WTY4					
1404	Q3UJ02		Hbs1l;Gm992	HBS1-like protein	Q3UJ02					
1406	Q8C2K3		Gm1821		Q8C2K3					
1407	Q78KL9		Rexo2	Oligoribonuclease, mitochondrial	Q78KL9					
1409	V9GX53		Erc1;Erc2	ELKS/Rab6-interacting/CAST family member 1;	V9GX53					
1411	G3UYF9		Pfdn6	Prefoldin subunit 6	G3UYF9					
1420	Q9D4J1		Efhd1	EF-hand domain-containing protein D1	Q9D4J1					
1421	Q64337-2		Sqstm1	Sequestosome-1	Q64337-2					
1422	Q6GXA8		Timp2	Metalloproteinase inhibitor 2	Q6GXA8					
1425	Q542H7		Fabp4	Fatty acid-binding protein, adipocyte	Q542H7					
1431	Q3TML7		mt-Co1;COX1;	Cytochrome c oxidase subunit 1	Q3TML7					
1434	Q9D9R3		Spata9	Spermatogenesis-associated protein 9	Q9D9R3	25.175755	17.938969	25.1719478	0	
1437	Q3V314		Ick	Serine/threonine-protein kinase ICK	Q3V314	24.283759	24.439419	24.2801639	24.433003	
1464	E9QPH0		Prkg2	cGMP-dependent protein kinase;cGMP-depend	E9QPH0					
1465	Q9CQK7		Rwdd1	RWD domain-containing protein 1	Q9CQK7					
1466	Q9QZH3		Ppie	Peptidyl-prolyl cis-trans isomerase E	Q9QZH3					
1480	Q61781		Krt14	Keratin, type I cytoskeletal 14	Q61781					
1493	A2RSP9		Reck	Reversion-inducing cysteine-rich protein with Ka	A2RSP9					
1501	Q8C7K6		Pcyox1l	Prenylcysteine oxidase-like	Q8C7K6					

1503	B9EIE9		Adss	Adenylosuccinate synthetase isozyme 2	B9EIE9					
1524	Q91VR7		Map1lc3a	Microtubule-associated proteins 1A/1B light cha	Q91VR7					
1525	A2AI57		Trpm7	Transient receptor potential cation channel subf	A2AI57					
1528	Q8C8R3		Ank2	Ankyrin-2	Q8C8R3	25.680699	17.909983	25.6751461	0	
1535	Q9D1P4		Chordc1	Cysteine and histidine-rich domain-containing p	Q9D1P4					
1542	Q8K2C9		Hacd3	Very-long-chain (3R)-3-hydroxyacyl-CoA dehyd	Q8K2C9					
1554	R4GML2		Rad21l;Rad21	Double-strand-break repair protein rad21-like p	R4GML2					
1557	B1AT92		Grb2	Growth factor receptor-bound protein 2	B1AT92					
1559	Q8C4N0		Herpud1	Homocysteine-responsive endoplasmic reticulu	Q8C4N0					
1572	Q9EST1		Gsdma;Gsdm	Gasdermin-A;Gasdermin-A2;Gasdermin-A3	Q9EST1					
1576	Q4FJX9		Sod2	Superoxide dismutase;Superoxide dismutase [M	Q4FJX9					
1580	Q8VED9		Lgalsl	Galectin-related protein	Q8VED9					
1583	B1AWT7		Ube2j1	Ubiquitin-conjugating enzyme E2 J1	B1AWT7					
1605	A2RS68		Snrnp70	U1 small nuclear ribonucleoprotein 70 kDa	A2RS68					
1609	A2RSX9		Arfip1		A2RSX9					
1612	Q8R0M2		Ugp2	UTP--glucose-1-phosphate uridylyltransferase	Q8R0M2					
1615	Q8BGH7		Cdc42se2	CDC42 small effector protein 2	Q8BGH7					
1617	Q5M9P7		Carhsp1	Calcium-regulated heat stable protein 1	Q5M9P7					
1623	G3X9U9		Fis1	Mitochondrial fission 1 protein	G3X9U9					
1625	A0A087WP81		Uchl5	Ubiquitin carboxyl-terminal hydrolase isozyme L	A0A087WP81					
1629	Q9JIX0		Eny2	Transcription and mRNA export factor ENY2	Q9JIX0					
1631	D3Z037		Tcf3		D3Z037	21.696642	23.076337	21.6506638	23.031426	
1635	Q99LJ0		Cttnbp2nl	CTTNBP2 N-terminal-like protein	Q99LJ0					
1641	Q3V424		Snrpg	Small nuclear ribonucleoprotein G	Q3V424					
1645	P62774		Mtpn	Myotrophin	P62774					
1656	A2AUK7		Epb4.1l1;Epb4	Band 4.1-like protein 1	A2AUK7					
1658	A0A0N4SV61		Bcat1	Branched-chain-amino-acid aminotransferase;E	A0A0N4SV61					
1663	F6X8H7		Depdc5	DEP domain-containing protein 5	F6X8H7					
1681	Q3TTE3		Ube2h	Ubiquitin-conjugating enzyme E2 H	Q3TTE3					
1686	Q3TCL2		Akr1b3;Akr1b	Aldose reductase	Q3TCL2					
1693	Q3TFB5		Nans		Q3TFB5					
1697	Q8BMD8		Slc25a24	Calcium-binding mitochondrial carrier protein SC	Q8BMD8					
1700	Q8R3F6		Mms22l	Protein MMS22-like	Q8R3F6					
1710	B1AV14		Apool	MICOS complex subunit Mic27	B1AV14					
1714	A2AQE4		Cops2	COP9 signalosome complex subunit 2	A2AQE4					
1719	Q9D6H6		Ndufb3	NADH dehydrogenase [ubiquinone] 1 beta subc	Q9D6H6					
1723	Q3UFV3		Gcc2	GRIP and coiled-coil domain-containing protein 2	Q3UFV3					
1727	A0A1B0GT81		Bax	Apoptosis regulator BAX	A0A1B0GT81					
1732	Q544R1		Prosc	Proline synthase co-transcribed bacterial homol	Q544R1					
1736	P70124		Serpinb5	Serpin B5	P70124					
1748	Q8VC72		Ndufs8	NADH dehydrogenase [ubiquinone] iron-sulfur p	Q8VC72					
1749	Q9D2Q3			Uncharacterized protein C10orf88 homolog	Q9D2Q3					
1750	Q6ZQC2		Nupl1	Nucleoporin p58/p45	Q6ZQC2					
1751	Q5HZI6		Tsta3	GDP-L-fucose synthase	Q5HZI6					
1754	D3Z368		Camk1	Calcium/calmodulin-dependent protein kinase t	D3Z368					
1756	E9Q190		Ttc23l	Tetratricopeptide repeat protein 23-like	E9Q190	26.769206	18.730197	26.7672047	0	
1759	A0A087WRM2		Itm2c	Integral membrane protein 2C;CT-BRI3	A0A087WRM2					
1770	Q9D0B6		Pbdc1	Protein PBDC1	Q9D0B6					
1773	Q545R3		Ndrg1	Protein NDRG1	Q545R3					
1775	A0A0A0MQN1		Krtcap2	Keratinocyte-associated protein 2	A0A0A0MQN1					
1776	Q3TV29		Plk2	Serine/threonine-protein kinase PLK;Serine/thr	Q3TV29					
1778	O54692		Zw10	Centromere/kinetochore protein zw10 homolog	O54692					
1780	Q6P069-2		Sri	Sorcin	Q6P069-2					
1786	A0A1B0GS27		Ddhd2	Phospholipase DDHD2	A0A1B0GS27					
1787	Q8C783		Gjc1	Gap junction gamma-1 protein	Q8C783					
1791	F7CB97		Clip1	CAP-Gly domain-containing linker protein 1	F7CB97					
1793	Q543U7		Il17rb	Interleukin-17 receptor B	Q543U7	25.276301	23.026359	25.2747153	22.999705	
1795	P0C6V0				P0C6V0					
1797	U5THC6				U5THC6					
1799	L7N1Z9		Vmn2r93		L7N1Z9	24.623958	17.908002	24.6182551	0	
1801	Q3TSL7		Ptn	Pleiotrophin	Q3TSL7					
1803	Q80TQ3		Nefh	Neurofilament heavy polypeptide	Q80TQ3					
1809	Q9Z2L6		Minpp1	Multiple inositol polyphosphate phosphatase 1	Q9Z2L6					
1810	Q80X95		Rraga	Ras-related GTP-binding protein A	Q80X95					
1815	Q3TGJ9		Gsn		Q3TGJ9					
1816	A0A0J9YVH9		Psph	Phosphoserine phosphatase	A0A0J9YVH9					
1821	Q3UKH3		Acaa2	3-ketoacyl-CoA thiolase, mitochondrial	Q3UKH3					
1830	Q3U955		Tmpo		Q3U955					
1837	G5E814		Ndufa11	NADH dehydrogenase [ubiquinone] 1 alpha sub	G5E814					
1839	F7CIB0		Cep128	Centrosomal protein of 128 kDa	F7CIB0					
1846	P70677		Casp3	Caspase-3;Caspase-3 subunit p17;Caspase-3 su	P70677					
1848	Q7TT42		trypsinogen;	1810009J06Rik;Gm2663	Q7TT42					
1854	Q3UEB3-3		Puf60	Poly(U)-binding-splicing factor PUF60	Q3UEB3-3					
1860	Q544T5		Dhfr	Dihydrofolate reductase	Q544T5	28.491597	27.980705	28.4913685	27.980449	
1861	Q99MX0		Tktl1	Transketolase-like protein 1	Q99MX0					
1865	Q7TMW5		Hdc	Histidine decarboxylase	Q7TMW5	25.96905	24.974951	25.9596036	24.963166	
1867	Q3V3U0		Pygb	Alpha-1,4 glucan phosphorylase;Glycogen phos	Q3V3U0					
1869	Q3ULE5		Ubxn4	UBX domain-containing protein 4	Q3ULE5					
1874	E9PUM4		Tln2	Talin-2	E9PUM4					
1876	F7CBN2		Stag3	Cohesin subunit SA-3	F7CBN2					
1877	E9Q8I9-3		Fry	Protein furry homolog	E9Q8I9-3					
1886	Q91XU0		Wrnip1	ATPase WRNIP1	Q91XU0	26.447326	24.841125	26.4456948	24.835892	
1891	Q8BTU6		Eif4a2	Eukaryotic initiation factor 4A-II;Eukaryotic initi	Q8BTU6					
1896	Q6A028		Swap70	Switch-associated protein 70	Q6A028					
1904	Q62232		Six2	Homeobox protein SIX2	Q62232	27.98216	26.720082	27.9821598	26.717069	
1906	Q91XQ0-2		Dnah8	Dynein heavy chain 8, axonemal	Q91XQ0-2	23.379974	25.741479	23.3071295	25.736558	
1912	Q80X72		Lrrc15	Leucine-rich repeat-containing protein 15	Q80X72					
1917	U5TNU7				U5TNU7	19.359046	17.992251	19.1128936	0	

1920	S4R255		Nos1;NOS-I	Nitric oxide synthase;Nitric oxide synthase, brain	S4R255	25.555818	18.128048	25.5521677	0	
1921	Q5SWJ8		Sec61g	Protein transport protein Sec61 subunit gamma	Q5SWJ8					
1923	Q9R060		Nubp1	Cytosolic Fe-S cluster assembly factor NUBP1	Q9R060					
1927	Q3USX5		Irak4	Interleukin-1 receptor-associated kinase 4	Q3USX5	22.941842	21.135038	22.9296091	21.023743	
1932	Q3V1H1		Ckap2	Cytoskeleton-associated protein 2	Q3V1H1	25.076519	24.89372	25.0743755	24.892288	
1934	Q69ZS8-3				Q69ZS8-3					
1935	Q3TXY0		Crmp1	Dihydropyrimidinase-related protein 1	Q3TXY0					
1936	M0QWX7		Cox4i1		M0QWX7					
1939	A2AED9		Spata6	Spermatogenesis-associated protein 6	A2AED9	19.534606	17.82304	19.3361978	0	
1940	J3QMC5		Mdn1		J3QMC5					
1942	Q6P7V9		Smc4	Structural maintenance of chromosomes protein	Q6P7V9	18.422359	20.213424	0	20.162532	
1946	B1ATY9		Hgs	Hepatocyte growth factor-regulated tyrosine kin	B1ATY9					
1951	Q7TQG5		Neo1	Neogenin	Q7TQG5					
1956	Q3U4H0		Mapre1	Microtubule-associated protein RP/EB family m	Q3U4H0					
1958	Q6PDL0		Dync1li2	Cytoplasmic dynein 1 light intermediate chain 2	Q6PDL0					
1970	D3Z383		Mest	Mesoderm-specific transcript protein	D3Z383					
1974	Q80ZM5		H1fx		Q80ZM5					
1976	F8WHW3		Tmem55b	Type 1 phosphatidylinositol 4,5-bisphosphate 4-	F8WHW3					
1988	A2A7Z4		Btf3l4	Transcription factor BTF3;Transcription factor BT	A2A7Z4					
1989	P57080		Usp25	Ubiquitin carboxyl-terminal hydrolase 25	P57080	22.088143	17.720877	22.0484348	0	
1990	Q545E6		Tsn	Translin	Q545E6					
1993	Q2M4H1		Onecut1	Hepatocyte nuclear factor 6	Q2M4H1	22.718058	24.575513	22.7100552	24.566414	
1994	Q8C9B9		Dido1	Death-inducer obliterator 1	Q8C9B9	21.424224	23.213578	21.3579502	23.150917	
1995	Q9D5W3		4921513D11Rik		Q9D5W3	22.357852	17.631198	22.3578518	0	
2002	Q0VBU4		Ccdc47	Coiled-coil domain-containing protein 47	Q0VBU4					
2004	Q3U3B9		Syap1	Synapse-associated protein 1	Q3U3B9					
2010	P16546		Sptan1	Spectrin alpha chain, non-erythrocytic 1	P16546					
2024	Q9D0T1		Nhp2l1	NHP2-like protein 1;NHP2-like protein 1, N-term	Q9D0T1					
2026	Q8R0J7		Vps37b	Vacuolar protein sorting-associated protein 37B	Q8R0J7					
2027	A0A0R4J1L2		Eef1d	Elongation factor 1-delta	A0A0R4J1L2					
2028	O55091		Impact	Protein IMPACT	O55091					
2032	Q0PD48		Rab9;Rab9a	Ras-related protein Rab-9A	Q0PD48					
2034	Q9CUR1		Ccdc178	Coiled-coil domain-containing protein 178	Q9CUR1	23.935498	24.192368	23.9354981	24.175511	
2037	D6RFN5		Ninj1	Ninjurin-1	D6RFN5					
2040	D6RI64		1110004F10Rik	Small acidic protein	D6RI64					
2041	Q9CSL5		Dusp10	Dual specificity protein phosphatase 10	Q9CSL5	21.461319	17.790129	21.3698534	0	
2046	A0A140LHY3		Mcoln1	Mucolipin-1	A0A140LHY3					
2047	E9Q467		Abcc4		E9Q467					
2054	Q3TQ66		Hnrnpm		Q3TQ66	17.759847	20.207861	0	20.101842	
2056	Q3UGE5		Pdcd6	Programmed cell death protein 6	Q3UGE5					
2061	Q792Y0		Pabpc2	Polyadenylate-binding protein	Q792Y0					
2065	L7N1Y0		Dnah7b		L7N1Y0	27.198614	30.515488	27.1979378	30.515454	
2067	Q8R3C7		Ddx19b;Ddx19	ATP-dependent RNA helicase DDX19A	Q8R3C7					
2070	E9PXC0		Srp54c;Srp54	Signal recognition particle 54 kDa protein	E9PXC0					
2083	Q9D281		Fam114a1	Protein Noxp20	Q9D281					
2087	A0A140LHR4				A0A140LHR4					
2091	P31254		Uba1y	Ubiquitin-like modifier-activating enzyme 1 Y	P31254					
2094	Q4U4S6		Xirp2	Xin actin-binding repeat-containing protein 2	Q4U4S6					
2095	A2AUV5		Procr	Endothelial protein C receptor	A2AUV5					
2097	Q61425		Hadh	Hydroxyacyl-coenzyme A dehydrogenase, mito	Q61425					
2100	Q9D927		1810009N02Rik		Q9D927					
2106	D3Z795		Psmg1	Proteasome assembly chaperone 1	D3Z795					
2109	Q3THH0		Cmtm7	CKLF-like MARVEL transmembrane domain-cor	Q3THH0					
2117	G3UZX6		Sumo3;Sumo	Small ubiquitin-related modifier;Small ubiquitin-	G3UZX6					
2118	Q9R1R9		Rdh11	Retinol dehydrogenase 11	Q9R1R9					
2119	Q3USZ5		Nop56	Nucleolar protein 56	Q3USZ5					
2120	Q99KD0		Fn1		Q99KD0					
2122	Q80TA9		Epg5	Ectopic P granules protein 5 homolog	Q80TA9					
2131	Q9JJC6		Rilpl1	RILP-like protein 1	Q9JJC6					
2136	P60521		Gabarapl2	Gamma-aminobutyric acid receptor-associated	P60521					
2143	Q6A0D1		Emc2	ER membrane protein complex subunit 2	Q6A0D1					
2149	Q6PI12		Tnfrsf11b	Tumor necrosis factor receptor superfamily mer	Q6PI12					
2151	A0A0G2JDW1		Agap1;Agap3	Arf-GAP with GTPase, ANK repeat and PH doma	A0A0G2JDW1					
2152	Q921Q7		Rin1	Ras and Rab interactor 1	Q921Q7					

Appendix B

A summary of SMCHD1 mutations and their effects on protein function

TABLE B.1: **A summary of SMCHD1 mutations and their effects on protein function.**

SNPs	Smchd1 domain	Associated disease	Observed functional change	Pubmed ID
Arg34Pro	UBL	FSHD2	N/A	31243061
Asn104Ser	UBL	FSHD2	N/A	31243061
Leu107Pro	UBL-ATPase linker	FSHD2, BAMS	N/A	31243061, 29980640, 28067909
Ala110Thr	UBL-ATPase linker	FSHD2	N/A	31243061, 25370034
Met129Arg	ATPase	BAMS	N/A	31243061
Met129Lys	ATPase	BAMS	N/A	31243061, 28067909
Ala134Ser	ATPase	BAMS	Enhanced ATPase activity	31243061, 28067911
Ser135Asn	ATPase	BAMS	No change in ATPase activity	31243061, 28067909, 28067911
Ser135Cys	ATPase	BAMS	Enhanced ATPase activity, altered nuclear localisation	31243061, 28067909, 28067911
Ser135Ile	ATPase	BAMS	N/A	31243061, 28067909, 28067911
Glu136Asp	ATPase	BAMS	N/A	31243061, 28067909
Glu136Gly	ATPase	BAMS	Enhanced ATPase activity	31243061, 28067911
Gly137Glu	ATPase	FSHD2, BAMS	Enhanced ATPase activity	31243061, 28067909, 25256356
Asn139His	ATPase	BAMS	Enhanced ATPase activity	31243061, 28067909
Leu141Phe	ATPase	BAMS	N/A	31243061, 28067909, 28067911
Glu147Ala	ATPase	N/A	Reduced ATPase activity, altered cellular localisation	27059856, 26391951
Asp150His	ATPase	FSHD2	N/A	31243061
Phe171Val	ATPase	BAMS	N/A	31243061, 28067909
Gly188Arg	ATPase	FSHD2	N/A	31243061
Met189Val	ATPase	FSHD2	N/A	31243061
Leu194Phe	ATPase	FSHD2	Reduced ATPase activity	31243061, 25256356
Lys204Glu	ATPase	FSHD2	N/A	31243061, 29980640
Ala242Gly	ATPase	BAMS	N/A	31243061, 28067909
Ala242Thr	ATPase	FSHD2	N/A	31243061, 29980640
His263Asp	ATPase	FSHD2	Reduced ATPase activity	31243061, 25256356
Glu264Lys	ATPase	FSHD2	N/A	31243061
Tyr283Cys	ATPase	FSHD2	No change in ATPase activity	31243061, 27061275
Trp324Ser	ATPase	BAMS	Reduced ATPase activity	31243061, 28067911
Arg344Gln	ATPase	FSHD2	N/A	31243061, 29980640
Gln345Arg	ATPase	BAMS	Enhanced ATPase activity	31243061, 28067909
His348Arg	ATPase	BAMS	No change in ATPase activity	31243061, 28067909, 28067911
Tyr353Cys	ATPase	FSHD2	Reduced ATPase activity	31243061, 23143600
Gln400Leu	Transducer	BAMS	N/A	31243061, 28067909
Asp420Val	Transducer	BAMS	Reduced ATPase activity	31243061, 28067909, 28067911
Gly425Arg	Transducer	FSHD2	N/A	31243061, 25256356
Arg428Cys	Transducer	FSHD2	N/A	31243061
Glu473Gln	Transducer	BAMS	Reduced ATPase activity	31243061, 28067909
Gly478Glu	Transducer	FSHD2	Reduced ATPase activity	31243061, 25370034
Arg479Pro	Transducer	FSHD2	N/A	31243061, 23143600
Arg479Leu	Transducer	FSHD2	N/A	31243061, 28744936
Arg479Gln	Transducer	FSHD2	N/A	31243061
Cys492Arg	Transducer	FSHD2	N/A	31243061, 23143600
Lys518Glu	Transducer	BAMS	Enhanced ATPase activity	31243061, 28067911
Phe519Ser	Transducer	FSHD2	N/A	31243061, 29980640
Thr523Lys	Transducer	BAMS	Reduced ATPase activity	31243061, 28067909
Asn524Ser	Transducer	BAMS	N/A	31243061, 28067909
Thr527Met	Transducer	FSHD2	Reduced ATPase activity	31243061, 24075187
Gln551Arg	Transducer	FSHD2	N/A	31243061
Arg552Gln	Transducer	BAMS	No change in ATPase activity	31243061, 28067909, 28067911
Trp596Gly	linker (strand prediction)	FSHD2	N/A	31243061
Val615Asp	linker (strand prediction)	FSHD2	N/A	31243061, 25370034
Pro622Leu	linker (helical prediction)	FSHD2	N/A	31243061
Val641Leu	linker (strand prediction)	FSHD2	N/A	31243061
Ala667Glu	linker (strand prediction)	N/A	No change in ATPase activity, altered nuclear localisation	N/A
Pro690Ser	linker (loop prediction)	FSHD2	Reduced ATPase activity	31243061, 23143600
Leu748Pro	linker (strand prediction)	FSHD2	N/A	31243061, 25256356
Tyr774Cys	linker (loop prediction)	FSHD2	N/A	31243061
Asp849Asn	linker (strand prediction)	FSHD2	N/A	31243061, 23143600
Leu923Pro	linker (strand prediction)	FSHD2	N/A	31243061
Leu978His	linker (loop prediction)	FSHD2	N/A	31243061
Tyr981Asp	linker (strand prediction)	FSHD2	N/A	31243061, 27153398
Gly1063Arg	linker (disorder prediction)	FSHD2	N/A	31243061
Leu1108Pro	linker (disorder prediction)	FSHD2	N/A	31243061
Val1114Ile	linker (disorder prediction)	FSHD2	N/A	31243061
Val1271Leu	linker (strand prediction)	FSHD2	N/A	31243061
Ile1300Lys	linker (disorder prediction)	FSHD2	N/A	31243061
Gln1463Pro	linker (strand prediction)	FSHD2	N/A	31243061, 25370034
Met1468Ile	linker (strand prediction)	FSHD2	N/A	31243061, 25256356
Pro1485Leu	linker (disorder prediction)	FSHD2	N/A	31243061, 25370034
Phe1554Ser	linker (helical prediction)	FSHD2	N/A	31243061, 23143600
Lys1718Ala	SMC hinge	N/A	No change in affinity for DNA	32546545
Arg1719Ala	SMC hinge	N/A	No change in affinity for DNA	32546545
Asp1750Gly	SMC hinge	FSHD2	Altered nuclear localisation, enhanced affinity for DNA	31243061
Asp1750Val	SMC hinge	FSHD2	Altered nuclear localisation, enhanced affinity for DNA	31243061
Arg1762Ala	SMC hinge	N/A	Reduced affinity for DNA	32546545
Tyr1765Ala	SMC hinge	N/A	No change in affinity for DNA	32546545
Arg1771Ala	SMC hinge	N/A	Reduced affinity for DNA	32546545
Lys1789Ala	SMC hinge	N/A	Reduced affinity for DNA	32546545
Arg1790Ala	SMC hinge	N/A	Reduced affinity for DNA, no change in cellular localisation	32546545
Arg1796Ala	SMC hinge	N/A	Reduced affinity for DNA, no change in cellular localisation	32546545
Lys1799Ala	SMC hinge	N/A	Reduced affinity for DNA	32546545
Asp1842Ala	SMC hinge	N/A	No change in affinity for DNA or cellular localisation	32546545
Arg1848Ala	SMC hinge	N/A	Reduced affinity for DNA and altered cellular localisation	32546545
Tyr1846Cys	SMC hinge	FSHD2	N/A	32546545
Arg1866Gly	SMC hinge	FSHD2	Reduced affinity for DNA and altered cellular localisation	31243061, 27153398
Arg1866Gln	SMC hinge	FSHD2	N/A	31243061
Arg1869Ala	SMC hinge	N/A	Reduced affinity for DNA, no change in cellular localisation	32546545
Gly1872Ala	SMC hinge	N/A	No change in affinity for DNA	32546545
Lys1873Ala	SMC hinge	N/A	Reduced affinity for DNA, no change in cellular localisation	32546545
Phe1874Ala	SMC hinge	N/A	No change in affinity for DNA	32546545
Lys1880Ala	SMC hinge	N/A	No change in affinity for DNA	32546545

Bibliography

[1] Karolin Luger, Armin W Mäder, Robin K Richmond, David F Sargent, and Timothy J Richmond. "Crystal structure of the nucleosome core particle at 2.8 Å resolution". In: *Nature* 389.6648 (1997), pp. 251–260.

[2] Rudolf Jaenisch and Adrian Bird. "Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals". In: *Nature genetics* 33.3 (2003), pp. 245–254.

[3] Tony Kouzarides. "Chromatin modifications and their function". In: *Cell* 128.4 (2007), pp. 693–705.

[4] Andrew S Belmont, Michael B Braunfeld, John W Sedat, and David A Agard. "Large-scale chromatin structural domains within mitotic and interphase chromosomes in vivo and in vitro". In: *Chromosoma* 98.2 (1989), pp. 129–143.

[5] Jan-Michael Peters, Antonio Tedeschi, and Julia Schmitz. "The cohesin complex and its roles in chromosome biology". In: *Genes & development* 22.22 (2008), pp. 3089–3114.

[6] Lydia C Green, Paul Kalitsis, Tsz M Chang, Miri Cipetic, Ji Hun Kim, Owen Marshall, Lynne Turnbull, Cynthia B Whitchurch, Paola Vagnarelli, Kumiko Samejima, et al. "Contrasting roles of condensin I and condensin II in mitotic chromosome formation". In: *Journal of cell science* 125.6 (2012), pp. 1591–1604.

[7] Stanislau Yatskevich, James Rhodes, and Kim Nasmyth. "Organization of Chromosomal DNA by SMC Complexes". In: *Annual Review of Genetics* 53 (2019), pp. 445–482.

[8] Chloe M Rivera and Bing Ren. "Mapping human epigenomes". In: *Cell* 155.1 (2013), pp. 39–55.

[9] Huda Y Zoghbi and Arthur L Beaudet. "Epigenetics and human disease". In: *Cold Spring Harbor perspectives in biology* 8.2 (2016), a019497.

[10] Robert J Klose and Adrian P Bird. "Genomic DNA methylation: the mark and its mediators". In: *Trends in biochemical sciences* 31.2 (2006), pp. 89–97.

[11] Di Cui and Xiangru Xu. "DNA methyltransferases, DNA methylation, and age-associated cognitive function". In: *International journal of molecular sciences* 19.5 (2018), p. 1315.

[12] Robert S Illingworth and Adrian P Bird. "CpG islands–'a rough guide'". In: *FEBS letters* 583.11 (2009), pp. 1713–1720.

[13] Lisa D Moore, Thuc Le, and Guoping Fan. "DNA methylation and its basic function". In: *Neuropsychopharmacology* 38.1 (2013), pp. 23–38.

[14] Ali Shilatifard. "Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression". In: *Annu. Rev. Biochem.* 75 (2006), pp. 243–269.

[15] Andrew J Bannister and Tony Kouzarides. "Regulation of chromatin by histone modifications". In: *Cell research* 21.3 (2011), pp. 381–395.

[16] Olivier Binda, Gary LeRoy, Dennis J Bua, Benjamin A Garcia, Or Gozani, and Stéphane Richard. "Trimethylation of histone H3 lysine 4 impairs methylation of histone H3 lysine 9: regulation of lysine methyltransferases by physical interaction with their substrates". In: *Epigenetics* 5.8 (2010), pp. 767–775.

[17] Helena Santos-Rosa, Robert Schneider, Andrew J Bannister, Julia Sherriff, Bradley E Bernstein, NC Tolga Emre, Stuart L Schreiber, Jane Mellor, and Tony Kouzarides. "Active genes are tri-methylated at K4 of histone H3". In: *Nature* 419.6905 (2002), pp. 407–411.

[18] Diego Pasini, Klaus H Hansen, Jesper Christensen, Karl Agger, Paul AC Cloos, and Kristian Helin. "Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2". In: *Genes & development* 22.10 (2008), pp. 1345–1355.

[19] Anjanabha Saha, Jacqueline Wittmeyer, and Bradley R Cairns. "Chromatin remodelling: the industrial revolution of DNA around histones". In: *Nature reviews Molecular cell biology* 7.6 (2006), pp. 437–447.

[20] Cedric R Clapier and Bradley R Cairns. "The biology of chromatin remodeling complexes". In: *Annual review of biochemistry* 78 (2009), pp. 273–304.

[21] Diana C Hargreaves and Gerald R Crabtree. "ATP-dependent chromatin remodeling: genetics, genomics and mechanisms". In: *Cell research* 21.3 (2011), pp. 396–420.

[22] Swetansu K Hota and Blaine Bartholomew. "Diversity of operation in ATP-dependent chromatin remodelers". In: *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1809.9 (2011), pp. 476–487.

[23] Jane Mellor and Antonin Morillon. "ISWI complexes in Saccharomyces cerevisiae". In: *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression* 1677.1-3 (2004), pp. 100–112.

[24] Bradley R Cairns. "Chromatin remodeling: insights and intrigue from single-molecule studies". In: *Nature structural & molecular biology* 14.11 (2007), pp. 989–996.

[25] Maheshi Udugama, Abdellah Sabri, and Blaine Bartholomew. "The INO80 ATP-dependent chromatin remodeling complex is a nucleosome spacing factor". In: *Molecular and cellular biology* 31.4 (2011), pp. 662–673.

[26] Chris Stockdale, Andrew Flaus, Helder Ferreira, and Tom Owen-Hughes. "Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes". In: *Journal of Biological Chemistry* 281.24 (2006), pp. 16279–16288.

[27] Martin Zofall, Jim Persinger, Stefan R Kassabov, and Blaine Bartholomew. "Chromatin remodeling by ISW2 and SWI/SNF requires DNA translocation inside the nucleosome". In: *Nature structural & molecular biology* 13.4 (2006), pp. 339–346.

[28] Alexander Gimelbrant, John N Hutchinson, Benjamin R Thompson, and Andrew Chess. "Widespread monoallelic expression on human autosomes". In: *Science* 318.5853 (2007), pp. 1136–1140.

[29] Mélanie A Eckersley-Maslin and David L Spector. "Random monoallelic expression: regulating gene expression one allele at a time". In: *Trends in Genetics* 30.6 (2014), pp. 237–244.

[30] Jeannie T Lee and Marisa S Bartolomei. "X-inactivation, imprinting, and long noncoding RNAs in health and disease". In: *Cell* 152.6 (2013), pp. 1308–1323.

[31] Neil Brockdorff, Alan Ashworth, Graham F Kay, Penny Cooper, Sandy Smith, Veronica M McCabe, Dominic P Norris, Graeme D Penny, Dipika Patel, and Sohaila Rastan. "Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome". In: *Nature* 351.6324 (1991), pp. 329–331.

[32] Stephen DM Brown. "XIST and the mapping of the X chromosome inactivation centre". In: *Bioessays* 13.11 (1991), pp. 607–612.

[33] Janet Rossant and Patrick PL Tam. "Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse". In: *Development* 136.5 (2009), pp. 701–713.

[34] Chun-Kan Chen, Mario Blanco, Constanza Jackson, Erik Aznauryan, Noah Ollikainen, Christine Surka, Amy Chow, Andrea Cerase, Patrick McDonel, and Mitchell Guttman. "Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing". In: *Science* 354.6311 (2016), pp. 468–472.

[35] Yesu Jeon, Kavitha Sarma, and Jeannie T Lee. "New and Xisting regulatory mechanisms of X chromosome inactivation". In: *Current opinion in genetics & development* 22.2 (2012), pp. 62–71.

[36] Denise P Barlow and Marisa S Bartolomei. "Genomic imprinting in mammals". In: *Cold Spring Harbor perspectives in biology* 6.2 (2014), a018382.

[37] Merlin G Butler. "Genomic imprinting disorders in humans: a mini-review". In: *Journal of assisted reproduction and genetics* 26.9-10 (2009), pp. 477–486.

[38] Jeannie T Lee and Marisa S Bartolomei. "X-inactivation, imprinting, and long noncoding RNAs in health and disease". In: *Cell* 152.6 (2013), pp. 1308–1323.

[39] Mary AM Cleaton, Carol A Edwards, and Anne C Ferguson-Smith. "Phenotypic outcomes of imprinted gene models in mice: elucidation of pre-and postnatal functions of imprinted genes". In: *Annual review of genomics and human genetics* 15 (2014), pp. 93–126.

[40] P Jelinic and P Shaw. "Loss of imprinting and cancer". In: *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* 211.3 (2007), pp. 261–268.

[41] Denise P Barlow and Marisa S Bartolomei. "Genomic imprinting in mammals". In: *Cold Spring Harbor perspectives in biology* 6.2 (2014), a018382.

[42] Marnie E Blewitt, Nicola K Vickaryous, Sarah J Hemley, Alyson Ashe, Timothy J Bruxner, Jost I Preis, Ruth Arkell, and Emma Whitelaw. "An N-ethyl-N-nitrosourea screen for genes involved in variegation in the mouse". In: *Proceedings of the National Academy of Sciences* 102.21 (2005), pp. 7629–7634.

[43] Marnie E Blewitt, Anne-Valerie Gendrel, Zhenyi Pang, Duncan B Sparrow, Nadia Whitelaw, Jeffrey M Craig, Anwyn Apedaile, Douglas J Hilton, Sally L Dunwoodie, Neil Brockdorff, et al. "SmcHD1, containing a structural-maintenance- of-chromosomes hinge domain, has a critical role in X inactivation". In: *Nature genetics* 40.5 (2008), p. 663.

[44] Arne W Mould, Zhenyi Pang, Miha Pakusch, Ian D Tonks, Mitchell Stark, Dianne Carrie, Pamela Mukhopadhyay, Annica Seidel, Jonathan J Ellis, Janine Deakin, et al. "Smchd1 regulates a subset of autosomal genes subject to monoallelic expression in addition to being critical for X inactivation". In: *Epigenetics & chromatin* 6.1 (2013), p. 19.

[45] Anne-Valerie Gendrel, Y Amy Tang, Masako Suzuki, Jonathan Godwin, Tatyana B Nesterova, John M Greally, Edith Heard, and Neil Brockdorff. "Epigenetic functions of smchd1 repress gene clusters on the inactive X chromosome and on autosomes". In: *Molecular and cellular biology* 33.16 (2013), pp. 3150–3165.

[46] Christopher T Gordon, Shifeng Xue, Gökhan Yigit, Hicham Filali, Kelan Chen, Nadine Rosin, Koh-ichiro Yoshiura, Myriam Oufadem, Tamara J Beck, Ruth McGowan, et al. "De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development". In: *Nature genetics* 49.2 (2017), p. 249.

[47] Natalie D Shaw, Harrison Brand, Zachary A Kupchinsky, Hemant Bengani, Lacey Plummer, Takako I Jones, Serkan Erdin, Kathleen A Williamson, Joe Rainger, Alexei Stortchevoi, et al. "SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome". In: *Nature genetics* 49.2 (2017), p. 238.

[48] Richard JLF Lemmers, Rabi Tawil, Lisa M Petek, Judit Balog, Gregory J Block, Gijs WE Santen, Amanda M Amell, Patrick J Van Der Vliet, Rowida Almomani, Kirsten R Straasheijm, et al. "Digenic inheritance of an SMCHD1 mutation and an FSHD-permissive D4Z4 allele causes facioscapulohumeral muscular dystrophy type 2". In: *Nature genetics* 44.12 (2012), p. 1370.

[49] David Garrick, Steven Fiering, David IK Martin, and Emma Whitelaw. "Repeat-induced gene silencing in mammals". In: *Nature genetics* 18.1 (1998), pp. 56–59.

[50] Anne-Valerie Gendrel, Anwyn Apedaile, Heather Coker, Ausma Termanis, Ilona Zvetkova, Jonathan Godwin, Y Amy Tang, Derek Huntley, Giovanni Montana, Steven Taylor, et al. "Smchd1-dependent and-independent pathways determine developmental dynamics of CpG island methylation on the inactive X chromosome". In: *Developmental cell* 23.2 (2012), pp. 265–279.

[51] Ryu-Suke Nozawa, Koji Nagao, Ken-Taro Igami, Sachiko Shibata, Natsuko Shirai, Naohito Nozaki, Takashi Sado, Hiroshi Kimura, and Chikashi Obuse. "Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBiX1 pathway". In: *Nature structural & molecular biology* 20.5 (2013), p. 566.

[52] Natasha Jansz, Andrew Keniry, Marie Trussart, Heidi Bildsoe, Tamara Beck, Ian D Tonks, Arne W Mould, Peter Hickey, Kelsey Breslin, Megan Iminitoff, et al. "Smchd1 regulates long-range chromatin interactions on the inactive X chromosome and at Hox clusters". In: *Nature structural & molecular biology* 25.9 (2018), pp. 766–777.

[53] Chen-Yu Wang, Teddy Jégu, Hsueh-Ping Chu, Hyun Jung Oh, and Jeannie T Lee. "SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X". In: *Cell* 174.2 (2018), pp. 406–421.

[54] Michal R Gdula, Tatyana B Nesterova, Greta Pintacuda, Jonathan Godwin, Ye Zhan, Hakan Ozadam, Michael Mcclellan, Daniella Moralli, Felix Krueger, Catherine M Green, et al. "The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome". In: *Nature communications* 10.1 (2019), pp. 1–14.

[55] Natasha Jansz, Tatyana Nesterova, Andrew Keniry, Megan Iminitoff, Peter F Hickey, Greta Pintacuda, Osamu Masui, Simon Kobelke, Niall Geoghegan, Kelsey

A Breslin, et al. "Smchd1 targeting to the inactive X is dependent on the Xist-HnrnpK-PRC1 pathway". In: *Cell reports* 25.7 (2018), pp. 1912–1923.

[56] Anand Minajigi, John E Froberg, Chunyao Wei, Hongjae Sunwoo, Barry Kesner, David Colognori, Derek Lessing, Bernhard Payer, Myriam Boukhali, Wilhelm Haas, et al. "A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation". In: *Science* 349.6245 (2015), aab2276.

[57] Shunsuke Toyoda, Masahumi Kawaguchi, Toshihiro Kobayashi, Etsuko Tarusawa, Tomoko Toyama, Masaki Okano, Masaaki Oda, Hiromitsu Nakauchi, Yumiko Yoshimura, Makoto Sanbo, et al. "Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity". In: *Neuron* 82.1 (2014), pp. 94–108.

[58] Kelan Chen, Jiang Hu, Darcy L Moore, Ruijie Liu, Sarah A Kessans, Kelsey Breslin, Isabelle S Lucet, Andrew Keniry, Huei San Leong, Clare L Parish, et al. "Genome-wide binding and mechanistic analyses of Smchd1-mediated epigenetic regulation". In: *Proceedings of the National Academy of Sciences* 112.27 (2015), E3535–E3544.

[59] Iromi Wanigasuriya, Quentin Gouil, Sarah A Kinkel, Andrés Tapia del Fierro, Tamara Beck, Ellise EA Roper, Kelsey Breslin, Jessica Stringer, Karla Hutt, Heather J Lee, et al. "Smchd1 is a maternal effect gene required for autosomal imprinting". In: *BioRxiv* (DOI: 10.1101/2020.01.20.913376).

[60] P Dolle and D Duboule. "Two gene members of the murine HOX-5 complex show regional and cell-type specific expression in developing limbs and gonads." In: *The EMBO journal* 8.5 (1989), pp. 1507–1515.

[61] Kelan Chen, Renwick CJ Dobson, Isabelle S Lucet, Samuel N Young, F Grant Pearce, Marnie E Blewitt, and James M Murphy. "The epigenetic regulator Smchd1 contains a functional GHKL-type ATPase domain". In: *Biochemical Journal* 473.12 (2016), pp. 1733–1744.

[62] Kelan Chen, Peter E Czabotar, Marnie E Blewitt, and James M Murphy. "The hinge domain of the epigenetic repressor Smchd1 adopts an unconventional homodimeric configuration". In: *Biochemical Journal* 473.6 (2016), pp. 733–742.

[63] Richard JLF Lemmers, Nienke van der Stoep, Patrick J van der Vliet, Steven A Moore, David San Leon Granado, Katherine Johnson, Ana Topf, Volker Straub, Teresinha Evangelista, Tahseen Mozaffar, et al. "SMCHD1 mutation spectrum for facioscapulohumeral muscular dystrophy type 2 (FSHD2) and Bosma arhinia microphthalmia syndrome (BAMS) reveals disease-specific localisation of variants in the ATPase domain". In: *Journal of medical genetics* 56.10 (2019), pp. 693–700.

[64] Frank Uhlmann. "SMC complexes: from DNA to chromosomes". In: *Nature reviews Molecular cell biology* 17.7 (2016), p. 399.

[65] Vasso Makrantoni and Adele L Marston. "Cohesin and chromosome segregation". In: *Current Biology* 28.12 (2018), R688–R693.

[66] Jesse R Dixon, David U Gorkin, and Bing Ren. "Chromatin domains: the unit of chromosome organization". In: *Molecular cell* 62.5 (2016), pp. 668–680.

[67] Kim Nasmyth and Christian H Haering. "Cohesin: its roles and mechanisms". In: *Annual review of genetics* 43 (2009), pp. 525–558.

[68] Tatsuya Hirano. "Condensin-based chromosome organization from bacteria to vertebrates". In: *Cell* 164.5 (2016), pp. 847–857.

[69] Johan H Gibcus, Kumiko Samejima, Anton Goloborodko, Itaru Samejima, Natalia Naumova, Johannes Nuebler, Masato T Kanemaki, Linfeng Xie, James R Paulson, William C Earnshaw, et al. "A pathway for mitotic chromosome formation". In: *Science* 359.6376 (2018), eaao6135.

[70] Sara Cuylen, Jutta Metz, Andrea Hruby, and Christian H Haering. "Entrapment of chromosomes by condensin rings prevents their breakage during cytokinesis". In: *Developmental cell* 27.4 (2013), pp. 469–478.

[71] Esther Toselli-Mollereau, Xavier Robellet, Lydia Fauque, Sébastien Lemaire, Christoph Schiklenk, Carlo Klein, Clémence Hocquet, Pénélope Legros, Lia N'Guyen, Léo Mouillard, et al. "Nucleosome eviction in mitosis assists condensin loading and chromosome condensation". In: *The EMBO journal* 35.14 (2016), pp. 1565–1581.

[72] Xavier Robellet, Vincent Vanoosthuyse, and Pascal Bernard. "The loading of condensin in the context of chromatin". In: *Current genetics* 63.4 (2017), pp. 577–589.

[73] Nike Walther, M Julius Hossain, Antonio Z Politi, Birgit Koch, Moritz Kueblbeck, Øyvind Ødegård-Fougner, Marko Lampe, and Jan Ellenberg. "A quantitative map of human Condensins provides new insights into mitotic chromosome architecture". In: *Journal of Cell Biology* 217.7 (2018), pp. 2309–2328.

[74] Tatsuya Hirano. "At the heart of the chromosome: SMC proteins in action". In: *Nature reviews Molecular cell biology* 7.5 (2006), pp. 311–322.

[75] Michiko Hirano and Tatsuya Hirano. "Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions". In: *Molecular cell* 21.2 (2006), pp. 175–186.

[76] Christian H Haering, Ana-Maria Farcas, Prakash Arumugam, Jean Metson, and Kim Nasmyth. "The cohesin ring concatenates sister DNA molecules". In: *Nature* 454.7202 (2008), pp. 297–301.

[77] Dmitri Ivanov and Kim Nasmyth. "A topological interaction between cohesin rings and a circular minichromosome". In: *Cell* 122.6 (2005), pp. 849–860.

[78] Markus Hassler, Indra A Shaltiel, and Christian H Haering. "Towards a unified model of SMC complex function". In: *Current Biology* 28.21 (2018), R1266–R1281.

[79] Tsuyoshi Terakawa, Shveta Bisht, Jorine M Eeftens, Cees Dekker, Christian H Haering, and Eric C Greene. "The condensin complex is a mechanochemical motor that translocates along DNA". In: *Science* 358.6363 (2017), pp. 672–676.

[80] Mahipal Ganji, Indra A Shaltiel, Shveta Bisht, Eugene Kim, Ana Kalichava, Christian H Haering, and Cees Dekker. "Real-time imaging of DNA loop extrusion by condensin". In: *Science* 360.6384 (2018), pp. 102–105.

[81] Eugene Kim, Jacob Kerssemakers, Indra A Shaltiel, Christian H Haering, and Cees Dekker. "DNA-loop extruding condensin complexes can traverse one another". In: *Nature* 579.7799 (2020), pp. 438–442.

[82] Markus Hassler, Indra A Shaltiel, Marc Kschonsak, Bernd Simon, Fabian Merkel, Lena Thärichen, Henry J Bailey, Jakub Macošek, Sol Bravo, Jutta Metz, et al. "Structural basis of an asymmetric condensin ATPase cycle". In: *Molecular cell* 74.6 (2019), pp. 1175–1188.

[83] John F Marko, Paolo De Los Rios, Alessandro Barducci, and Stephan Gruber. "DNA-segment-capture model for loop extrusion by structural maintenance of chromosome (SMC) protein complexes". In: *Nucleic acids research* 47.13 (2019), pp. 6956–6972.

[84] Iain F Davidson, Daniela Goetz, Maciej P Zaczek, Maxim I Molodtsov, Florian Weissmann, Gabriele Litos, David A Cisneros, Maria Ocampo-Hafalla, Rene Ladurner, Frank Uhlmann, et al. "Rapid movement and transcriptional re-localization of human cohesin on DNA". In: *The EMBO journal* 35.24 (2016), pp. 2671–2685.

[85] Johannes Stigler, Gamze Ö Çamdere, Douglas E Koshland, and Eric C Greene. "Single-molecule imaging reveals a collapsed conformational state for DNA-bound cohesin". In: *Cell reports* 15.5 (2016), pp. 988–998.

[86] Mai Kanke, Eri Tahara, Tomoko Nishiyama, et al. "Cohesin acetylation and Wapl-Pds5 oppositely regulate translocation of cohesin along DNA". In: *The EMBO journal* 35.24 (2016), pp. 2686–2698.

[87] Yoori Kim, Zhubing Shi, Hongshan Zhang, Ilya J Finkelstein, and Hongtao Yu. "Human cohesin compacts DNA by loop extrusion". In: *Science* 366.6471 (2019), pp. 1345–1349.

[88] Iain F Davidson, Benedikt Bauer, Daniela Goetz, Wen Tang, Gordana Wutz, and Jan-Michael Peters. "DNA loop extrusion by human cohesin". In: *Science* 366.6471 (2019), pp. 1338–1345.

[89] Rafal Ciosk, Masaki Shirayama, Anna Shevchenko, Tomoyuki Tanaka, Attila Toth, Andrej Shevchenko, and Kim Nasmyth. "Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins". In: *Molecular cell* 5.2 (2000), pp. 243–254.

[90] William CH Chao, Yasuto Murayama, Sofia Muñoz, Alessandro Costa, Frank Uhlmann, and Martin R Singleton. "Structural studies reveal the functional modularity of the Scc2-Scc4 cohesin loader". In: *Cell reports* 12.5 (2015), pp. 719–725.

[91] Stephen M Hinshaw, Vasso Makrantoni, Alastair Kerr, Adele L Marston, and Stephen C Harrison. "Structural evidence for Scc4-dependent localization of cohesin loading". In: *Elife* 4 (2015), e06057.

[92] Gamze Ö Çamdere, Kristian K Carlborg, and Douglas Koshland. "Intermediate step of cohesin's ATPase cycle allows cohesin to entrap DNA". In: *Proceedings of the National Academy of Sciences* 115.39 (2018), pp. 9732–9737.

[93] Masashi Minamino, Torahiko L Higashi, Céline Bouchoux, and Frank Uhlmann. "Topological in vitro loading of the budding yeast cohesin ring onto DNA". In: *Life science alliance* 1.5 (2018).

[94] Yan Li, Kyle W Muir, Matthew W Bowler, Jutta Metz, Christian H Haering, and Daniel Panne. "Structural basis for Scc3-dependent cohesin recruitment to chromatin". In: *Elife* 7 (2018), e38356.

[95] Vania Parelho, Suzana Hadjur, Mikhail Spivakov, Marion Leleu, Stephan Sauer, Heather C Gregson, Adam Jarmuz, Claudia Canzonetta, Zoe Webster, Tatyana Nesterova, et al. "Cohesins functionally associate with CTCF on mammalian chromosome arms". In: *Cell* 132.3 (2008), pp. 422–433.

[96] Eric D Rubio, David J Reiss, Piri L Welcsh, Christine M Disteche, Galina N Filippova, Nitin S Baliga, Ruedi Aebersold, Jeffrey A Ranish, and Anton Krumm. "CTCF physically links cohesin to chromatin". In: *Proceedings of the National Academy of Sciences* 105.24 (2008), pp. 8309–8314.

[97] Georg A Busslinger, Roman R Stocsits, Petra Van Der Lelij, Elin Axelsson, Antonio Tedeschi, Niels Galjart, and Jan-Michael Peters. "Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl". In: *Nature* 544.7651 (2017), pp. 503–507.

[98] Maria Ocampo-Hafalla, Sofia Muñoz, Catarina P Samora, and Frank Uhlmann. "Evidence for cohesin sliding along budding yeast chromosomes". In: *Open biology* 6.6 (2016), p. 150178.

[99] Rinku Dutta and Masayori Inouye. "GHKL, an emergent ATPase/kinase superfamily". In: *Trends in biochemical sciences* 25.1 (2000), pp. 24–28.

[100] Philippe Meyer, Chrisostomos Prodromou, Bin Hu, Cara Vaughan, S Mark Roe, Barry Panaretou, Peter W Piper, and Laurence H Pearl. "Structural and functional analysis of the middle segment of hsp90: implications for ATP hydrolysis and client protein and cochaperone interactions". In: *Molecular cell* 11.3 (2003), pp. 647–658.

[101] Kevin D Corbett and James M Berger. "Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution". In: *The EMBO Journal* 22.1 (2003), pp. 151–163.

[102] Julie Papillon, Jean-Francois Menetret, Claire Batisse, Reynald Helye, Patrick Schultz, Noelle Potier, and Valerie Lamour. "Structural insight into negative DNA supercoiling by DNA gyrase, a bacterial type 2A DNA topoisomerase". In: *Nucleic acids research* 41.16 (2013), pp. 7815–7827.

[103] Flora S Groothuizen, Ines Winkler, Michele Cristóvão, Alexander Fish, Herrie HK Winterwerp, Annet Reumer, Andreas D Marx, Nicolaas Hermans, Robert A Nicholls, Garib N Murshudov, et al. "MutS/MutL crystal structure reveals that the MutS sliding clamp loads MutL onto DNA". In: *Elife* 4 (2015), e06744.

[104] Philippe Meyer, Chrisostomos Prodromou, Chunyan Liao, Bin Hu, S Mark Roe, Cara K Vaughan, Ignacija Vlasic, Barry Panaretou, Peter W Piper, and Laurence H Pearl. "Structural basis for recruitment of the ATPase activator Aha1 to the Hsp90 chaperone machinery". In: *The EMBO journal* 23.3 (2004), pp. 511–519.

[105] Joanna L Holmes, Swee Y Sharp, Steve Hobbs, and Paul Workman. "Silencing of HSP90 cochaperone AHA1 expression decreases client protein activation and increases cellular sensitivity to the HSP90 inhibitor 17-allylamino-17-demethoxygeldanamycin". In: *Cancer research* 68.4 (2008), pp. 1188–1197.

[106] Florian H Schopf, Maximilian M Biebl, and Johannes Buchner. "The HSP90 chaperone machinery". In: *Nature reviews Molecular cell biology* 18.6 (2017), p. 345.

[107] Klaus Richter, Paul Muschler, Otmar Hainzl, and Johannes Buchner. "Coordinated ATP hydrolysis by the Hsp90 dimer". In: *Journal of Biological Chemistry* 276.36 (2001), pp. 33689–33696.

[108] Klaus Richter, Jochen Reinstein, and Johannes Buchner. "N-terminal residues regulate the catalytic efficiency of the Hsp90 ATPase cycle". In: *Journal of Biological Chemistry* 277.47 (2002), pp. 44905–44910.

[109] Chrisostomos Prodromou, Barry Panaretou, Shahzad Chohan, Giuliano Siligardi, Ronan O'Brien, John E Ladbury, S Mark Roe, Peter W Piper, and Laurence H Pearl. "The ATPase cycle of Hsp90 drives a molecular 'clamp'via transient

dimerization of the N-terminal domains". In: *The EMBO journal* 19.16 (2000), pp. 4383–4392.

[110] Chrisostomos Prodromou. "The 'active life'of Hsp90 complexes". In: *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1823.3 (2012), pp. 614–623.

[111] Klaus Richter, Sandra Moser, Franz Hagn, Rainer Friedrich, Otmar Hainzl, Markus Heller, Sandra Schlee, Horst Kessler, Jochen Reinstein, and Johannes Buchner. "Intrinsic inhibition of the Hsp90 ATPase activity". In: *Journal of Biological Chemistry* 281.16 (2006), pp. 11301–11311.

[112] Martin Hessling, Klaus Richter, and Johannes Buchner. "Dissection of the ATP-induced conformational cycle of the molecular chaperone Hsp90". In: *Nature structural & molecular biology* 16.3 (2009), pp. 287–293.

[113] Julia M Eckl and Klaus Richter. "Functions of the Hsp90 chaperone system: lifting client proteins to new heights". In: *International journal of biochemistry and molecular biology* 4.4 (2013), p. 157.

[114] Sebastian Karl Wandinger, Klaus Richter, and Johannes Buchner. "The Hsp90 chaperone machinery". In: *Journal of Biological Chemistry* 283.27 (2008), pp. 18473–18477.

[115] Christopher H Douse, Stuart Bloor, Yangci Liu, Maria Shamin, Iva A Tchasovnikarova, Richard T Timms, Paul J Lehner, and Yorgo Modis. "Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms". In: *Nature communications* 9.1 (2018), pp. 1–15.

[116] C Jake Harris, Dylan Husmann, Wanlu Liu, Farid El Kasmi, Haifeng Wang, Ashot Papikian, William A Pastor, Guillaume Moissiard, Ajay A Vashisht, Jeffery L Dangl, et al. "Arabidopsis AtMORC4 and AtMORC7 form nuclear bodies and repress a large number of protein-coding genes". In: *PLoS genetics* 12.5 (2016).

[117] Natasha E Weiser, Danny X Yang, Suhua Feng, Natallia Kalinava, Kristen C Brown, Jayshree Khanikar, Mallory A Freeberg, Martha J Snyder, Györgyi Csankovszki, Raymond C Chan, et al. "MORC-1 integrates nuclear RNAi and transgenerational chromatin architecture to promote germline immortality". In: *Developmental cell* 41.4 (2017), pp. 408–423.

[118] Iva A Tchasovnikarova, Richard T Timms, Christopher H Douse, Rhys C Roberts, Gordon Dougan, Robert E Kingston, Yorgo Modis, and Paul J Lehner. "Hyperactivation of HUSH complex function by Charcot–Marie–Tooth disease mutation in MORC2". In: *Nature genetics* 49.7 (2017), p. 1035.

[119] Lakshminarayan M Iyer, Saraswathi Abhiman, and L Aravind. "MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases". In: *Biology direct* 3.1 (2008), p. 8.

[120] Sisi Li, Linda Yen, William A Pastor, Jonathan B Johnston, Jiamu Du, Colin J Shew, Wanlu Liu, Jamie Ho, Bryan Stender, Amander T Clark, et al. "Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin". In: *Proceedings of the National Academy of Sciences* 113.35 (2016), E5108–E5116.

[121] Yasuhiro Mimura, Keiko Takahashi, Kiyo Kawata, Takashi Akazawa, and Norimitsu Inoue. "Two-step colocalization of MORC3 with PML nuclear bodies". In: *Journal of cell science* 123.12 (2010), pp. 2014–2024.

[122] William A Pastor, Hume Stroud, Kevin Nee, Wanlu Liu, Dubravka Pezic, Sergei Manakov, Serena A Lee, Guillaume Moissiard, Natasha Zamudio, Déborah Bourc'his, et al. "MORC1 represses transposable elements in the mouse male germline". In: *Nature communications* 5 (2014), p. 5795.

[123] HyeongJun Kim, Linda Yen, Somsakul P Wongpalee, Jessica A Kirshner, Nicita Mehta, Yan Xue, Jonathan B Johnston, Alma L Burlingame, John K Kim, Joseph J Loparo, et al. "The Gene-Silencing Protein MORC-1 Topologically Entraps DNA and Forms Multimeric Assemblies to Cause DNA Compaction". In: *Molecular cell* 75.4 (2019), pp. 700–710.

[124] Gui-Ling Wang, Chun-Yu Wang, Xin-Ze Cai, Wei Chen, Xiao-Hui Wang, and Feng Li. "Identification and expression analysis of a novel CW-type zinc finger protein MORC2 in cancer cells". In: *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology* 293.6 (2010), pp. 1002–1009.

[125] Yangguang Shao, Yan Li, Jian Zhang, Di Liu, Furong Liu, Yue Zhao, Tao Shen, and Feng Li. "Involvement of histone deacetylation in MORC2-mediated down-regulation of carbonic anhydrase IX". In: *Nucleic acids research* 38.9 (2010), pp. 2813–2824.

[126] Da-Qiang Li, Sujit S Nair, Kazufumi Ohshiro, Anupam Kumar, Vasudha S Nair, Suresh B Pakala, Sirigiri Divijendra Natha Reddy, Rajendra P Gajula, Jeyanthy Eswaran, L Aravind, et al. "MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response". In: *Cell reports* 2.6 (2012), pp. 1657–1669.

[127] Forest H Andrews, Qiong Tong, Kelly D Sullivan, Evan M Cornett, Yi Zhang, Muzaffar Ali, JaeWoo Ahn, Ahway Pandey, Angela H Guo, Brian D Strahl, et al. "Multivalent chromatin engagement and inter-domain crosstalk regulate MORC3 ATPase". In: *Cell reports* 16.12 (2016), pp. 3195–3207.

[128] Keiko Takahashi, Naofumi Yoshida, Naoko Murakami, Kiyo Kawata, Hiroyuki Ishizaki, Miki Tanaka-Okamoto, Jun Miyoshi, Andrew R Zinn, Hiroaki Shime, and Norimitsu Inoue. "Dynamic regulation of p53 subnuclear localization and senescence by MORC3". In: *Molecular biology of the cell* 18.5 (2007), pp. 1701–1709.

[129] Barry Panaretou, Chrisostomos Prodromou, S Mark Roe, Ronan O'Brien, John E Ladbury, Peter W Piper, and Laurence H Pearl. "ATP binding and hydrolysis are essential to the function of the Hsp90 molecular chaperone in vivo". In: *The EMBO journal* 17.16 (1998), pp. 4829–4836.

[130] Christian N Cunningham, Daniel R Southworth, Kristin A Krukenberg, and David A Agard. "The conserved arginine 380 of Hsp90 is not a catalytic residue, but stabilizes the closed conformation required for ATP hydrolysis". In: *Protein Science* 21.8 (2012), pp. 1162–1171.

[131] Nicholas J Brideau, Heather Coker, Anne-Valerie Gendrel, C Alistair Siebert, Karel Bezstarosti, Jeroen Demmers, Raymond A Poot, Tatyana B Nesterova, and Neil Brockdorff. "Independent mechanisms target SMCHD1 to trimethylated histone H3 lysine 9-modified chromatin and the inactive X chromosome". In: *Molecular and cellular biology* 35.23 (2015), pp. 4053–4068.

[132] S Mark Roe, Chrisostomos Prodromou, Ronan O'Brien, John E Ladbury, Peter W Piper, and Laurence H Pearl. "Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin". In: *Journal of medicinal chemistry* 42.2 (1999), pp. 260–266.

[133] Lars C Pedersen, Kaoru Inoue, Susan Kim, Lalith Perera, and Natalie D Shaw. "A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1". In: *Communications biology* 2.1 (2019), pp. 1–10.

[134] Kevin D Corbett and James M Berger. "Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI". In: *Structure* 13.6 (2005), pp. 873–882.

[135] Rabi Tawil and Silvère M Van Der Maarel. "Facioscapulohumeral muscular dystrophy". In: *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine* 34.1 (2006), pp. 1–15.

[136] Silvère M van der Maarel and Rune R Frants. "The D4Z4 Repeat–Mediated Pathogenesis of Facioscapulohumeral Muscular Dystrophy". In: *The American Journal of Human Genetics* 76.3 (2005), pp. 375–386.

[137] Richard JLF Lemmers, Jelle J Goeman, Patrick J van der Vliet, Merlijn P van Nieuwenhuizen, Judit Balog, Marianne Vos-Versteeg, Pilar Camano, Maria Antonia Ramos Arroyo, Ivonne Jerico, Mark T Rogers, et al. "Inter-individual differences in CpG methylation at D4Z4 correlate with clinical variability in FSHD1 and FSHD2". In: *Human molecular genetics* 24.3 (2015), pp. 659–669.

[138] Linda N Geng, Zizhen Yao, Lauren Snider, Abraham P Fong, Jennifer N Cech, Janet M Young, Silvere M van der Maarel, Walter L Ruzzo, Robert C Gentleman, Rabi Tawil, et al. "DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy". In: *Developmental cell* 22.1 (2012), pp. 38–51.

[139] Zizhen Yao, Lauren Snider, Judit Balog, Richard JLF Lemmers, Silvère M Van Der Maarel, Rabi Tawil, and Stephen J Tapscott. "DUX4-induced gene expression is the major molecular signature in FSHD skeletal muscle". In: *Human molecular genetics* 23.20 (2014), pp. 5342–5352.

[140] Valeria Kowaljow, Aline Marcowycz, Eugénie Ansseau, Cecilia B Conde, Sébastien Sauvage, Christel Mattéotti, Cristina Arias, E Daniel Corona, Nicolás G Nuñez, Oberdan Leo, et al. "The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein". In: *Neuromuscular Disorders* 17.8 (2007), pp. 611–623.

[141] Mirjam Larsen, Simone Rost, Nady El Hajj, Andreas Ferbert, Marcus Deschauer, Maggie C Walter, Benedikt Schoser, Pawel Tacik, Wolfram Kress, and Clemens R Müller. "Diagnostic approach for FSHD revisited: SMCHD1 mutations cause FSHD2 and act as modifiers of disease severity in FSHD1". In: *European Journal of Human Genetics* 23.6 (2015), pp. 808–816.

[142] Silvère M van der Maarel, Rabi Tawil, and Stephen J Tapscott. "Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence". In: *Trends in molecular medicine* 17.5 (2011), pp. 252–258.

[143] Petra GM van Overveld, Richard JFL Lemmers, Lodewijk A Sandkuijl, Leo Enthoven, Sara T Winokur, Floor Bakels, George W Padberg, Gert-Jan B van Ommen, Rune R Frants, and Silvère M van der Maarel. "Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy". In: *Nature genetics* 35.4 (2003), pp. 315–317.

[144] Weihua Zeng, Jessica C De Greef, Yen-Yun Chen, Richard Chien, Xiangduo Kong, Heather C Gregson, Sara T Winokur, April Pyle, Keith D Robertson, John A Schmiesing, et al. "Specific loss of histone H3 lysine 9 trimethylation and HP1γ/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD)". In: *PLoS genetics* 5.7 (2009).

[145] Sabrina Sacconi, Richard JLF Lemmers, Judit Balog, Patrick J Van Der Vliet, Pauline Lahaut, Merlijn P Van Nieuwenhuizen, Kirsten R Straasheijm, Rashmie D Debipersad, Marianne Vos-Versteeg, Leonardo Salviati, et al. "The FSHD2 gene SMCHD1 is a modifier of disease severity in families affected by FSHD1". In: *The American Journal of Human Genetics* 93.4 (2013), pp. 744–751.

[146] Alexandra D Gurzau, Kelan Chen, Shifeng Xue, Weiwen Dai, Isabelle S Lucet, Thanh Thao Nguyen Ly, Bruno Reversade, Marnie E Blewitt, and James M Murphy. "FSHD2-and BAMS-associated mutations confer opposing effects on SMCHD1 function". In: *Journal of Biological Chemistry* 293.25 (2018), pp. 9841–9853.

[147] JF Bosma, RI Henkin, RL Christiansen, and JR Herdt. "Hypoplasia of the nose and eyes, hyposmia, hypogeusia, and hypogonadotrophic hypogonadism in two males." In: *Journal of craniofacial genetics and developmental biology* 1.2 (1981), pp. 153–184.

[148] Benjamin Brasseur, Cindy M Martin, Zuzan Cayci, Lynn Burmeister, and Lisa A Schimmenti. "Bosma arhinia microphthalmia syndrome: clinical report and review of the literature". In: *American Journal of Medical Genetics Part A* 170.5 (2016), pp. 1302–1307.

[149] Karlien Mul, Richard JLF Lemmers, Marjolein Kriek, Patrick J van der Vliet, Marlinde L van den Boogaard, Umesh A Badrising, John M Graham, Angela E Lin, Harrison Brand, Steven A Moore, et al. "FSHD type 2 and Bosma arhinia microphthalmia syndrome: Two faces of the same mutation". In: *Neurology* 91.6 (2018), e562–e570.

[150] Alexander Peter Murphy and Volker Straub. "The classification, natural history and treatment of the limb girdle muscular dystrophies". In: *Journal of neuromuscular diseases* 2.s2 (2015), S7–S19.

[151] Suzanne B Cassidy and Daniel J Driscoll. "Prader–willi syndrome". In: *European journal of human genetics* 17.1 (2009), pp. 3–13.

[152] Carla S D'Angelo, José A Da Paz, Chong A Kim, Débora R Bertola, Claudia IE Castro, Monica C Varela, and Célia P Koiffmann. "Prader-Willi-like phenotype: investigation of 1p36 deletion in 41 patients with delayed psychomotor development, hypotonia, obesity and/or hyperphagia, learning disabilities and behavioral problems". In: *European journal of medical genetics* 49.6 (2006), pp. 451–460.

[153] Tayfun Özçelik, Stuart Leff, Wendy Robinson, Tim Donlon, Marc Lalande, Elvira Sanjines, Albert Schinzel, and Uta Francke. "Small nuclear ribonucleoprotein polypeptide N (SNRPN), an expressed gene in the Prader–Willi syndrome critical region". In: *Nature genetics* 2.4 (1992), pp. 265–269.

[154] Martha L Reed and Stuart E Leff. "Maternal imprinting of human SNRPN, a gene deleted in Prader–Willi syndrome". In: *Nature genetics* 6.2 (1994), pp. 163–167.

[155] Anthony P Goldstone. "Prader-Willi syndrome: advances in genetics, pathophysiology and treatment". In: *Trends in Endocrinology & Metabolism* 15.1 (2004), pp. 12–20.

[156] Ross A Dickins, Michael T Hemann, Jack T Zilfou, David R Simpson, Ingrid Ibarra, Gregory J Hannon, and Scott W Lowe. "Probing tumor phenotypes using

stable and regulated synthetic microRNA precursors". In: *Nature genetics* 37.11 (2005), pp. 1289–1295.

[157] Sarah A Kinkel, Roman Galeev, Christoffer Flensburg, Andrew Keniry, Kelsey Breslin, Omer Gilan, Stanley Lee, Joy Liu, Kelan Chen, Linden J Gearing, et al. "Jarid2 regulates hematopoietic stem cell function by acting with polycomb repressive complex 2". In: *Blood, The Journal of the American Society of Hematology* 125.12 (2015), pp. 1890–1900.

[158] Ian J Majewski, Marnie E Blewitt, Carolyn A de Graaf, Edward J McManus, Melanie Bahlo, Adrienne A Hilton, Craig D Hyland, Gordon K Smyth, Jason E Corbin, Donald Metcalf, et al. "Polycomb repressive complex 2 (PRC2) restricts hematopoietic stem cell activity". In: *PLoS biology* 6.4 (2008).

[159] Jeffrey J Babon and James M Murphy. "In vitro JAK kinase activity and inhibition assays". In: *JAK-STAT Signalling*. Springer, 2013, pp. 39–55.

[160] Carolina Perez-Iratxeta and Miguel A Andrade-Navarro. "K2D2: estimation of protein secondary structure from circular dichroism spectra". In: *BMC structural biology* 8.1 (2008), pp. 1–5.

[161] Emma J Petrie, Jarrod J Sandow, Annette V Jacobsen, Brian J Smith, Michael DW Griffin, Isabelle S Lucet, Weiwen Dai, Samuel N Young, Maria C Tanzer, Ahmad Wardak, et al. "Conformational switching of the pseudokinase domain promotes human MLKL tetramerization and cell death by necroptosis". In: *Nature communications* 9.1 (2018), pp. 1–15.

[162] Kelan Chen, Richard W Birkinshaw, Alexandra D Gurzau, Iromi Wanigasuriya, Ruoyun Wang, Megan Iminitoff, Jarrod J Sandow, Samuel N Young, Patrick J Hennessy, Tracy A Willson, et al. "Crystal structure of the hinge domain of Smchd1 reveals its dimerization mode and nucleic acid–binding residues". In: *Science signaling* 13.636 (2020).

[163] Julia J Griese, Gregor Witte, and Karl-Peter Hopfner. "Structure and DNA binding activity of the mouse condensin hinge domain highlight common and diverse features of SMC proteins". In: *Nucleic acids research* 38.10 (2010), pp. 3454–3465.

[164] Jürgen Cox and Matthias Mann. "MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification". In: *Nature biotechnology* 26.12 (2008), pp. 1367–1372.

[165] Jurgen Cox, Nadin Neuhauser, Annette Michalski, Richard A Scheltema, Jesper V Olsen, and Matthias Mann. "Andromeda: a peptide search engine integrated into the MaxQuant environment". In: *Journal of proteome research* 10.4 (2011), pp. 1794–1805.

[166] Peter Schuck. "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling". In: *Biophysical journal* 78.3 (2000), pp. 1606–1619.

[167] Chad A Brautigam. "Calculations and publication-quality illustrations for analytical ultracentrifugation data". In: *Methods in enzymology*. Vol. 562. Elsevier, 2015, pp. 109–133.

[168] Maxim V Petoukhov, Daniel Franke, Alexander V Shkumatov, Giancarlo Tria, Alexey G Kikhney, Michal Gajda, Christian Gorba, Haydyn DT Mertens, Petr V Konarev, and Dmitri I Svergun. "New developments in the ATSAS program package for small-angle scattering data analysis". In: *Journal of applied crystallography* 45.2 (2012), pp. 342–350.

[169] Petr V Konarev, Vladimir V Volkov, Anna V Sokolova, Michel HJ Koch, and Dmitri I Svergun. "PRIMUS: a Windows PC-based system for small-angle scattering data analysis". In: *Journal of applied crystallography* 36.5 (2003), pp. 1277–1282.

[170] DI Svergun. "Determination of the regularization parameter in indirect-transform methods using perceptual criteria". In: *Journal of applied crystallography* 25.4 (1992), pp. 495–503.

[171] Daniel Franke and Dmitri I Svergun. "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering". In: *Journal of applied crystallography* 42.2 (2009), pp. 342–346.

[172] Thomas S Walter, Christoph Meier, Rene Assenberg, Kin-Fai Au, Jingshan Ren, Anil Verma, Joanne E Nettleship, Raymond J Owens, David I Stuart, and Jonathan M Grimes. "Lysine methylation as a routine rescue strategy for protein crystallization". In: *Structure* 14.11 (2006), pp. 1617–1622.

[173] Julie Chaumeil, Sandrine Augui, Jennifer C Chow, and Edith Heard. "Combined immunofluorescence, RNA fluorescent in situ hybridization, and DNA fluorescent in situ hybridization to study chromatin changes, transcriptional activity, nuclear organization, and X-chromosome inactivation". In: *The nucleus*. Springer, 2008, pp. 297–308.

[174] Florence Corpet. "Multiple sequence alignment with hierarchical clustering". In: *Nucleic acids research* 16.22 (1988), pp. 10881–10890.

[175] Xavier Robert and Patrice Gouet. "Deciphering key features in protein structures with the new ENDscript server". In: *Nucleic acids research* 42.W1 (2014), W320–W324.

[176] Burkhard Rost, Guy Yachdav, and Jinfeng Liu. "The predictprotein server". In: *Nucleic acids research* 32.suppl_2 (2004), W321–W326.

[177] Norma J Greenfield. "Using circular dichroism spectra to estimate protein secondary structure". In: *Nature protocols* 1.6 (2006), p. 2876.

[178] Shane A Seabrook and Janet Newman. "High-throughput thermal scanning for protein stability: making a good technique more robust". In: *ACS combinatorial science* 15.8 (2013), pp. 387–392.

[179] Frank H Niesen, Helena Berglund, and Masoud Vedadi. "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability". In: *Nature protocols* 2.9 (2007), p. 2212.

[180] Burak H Alver, Kimberly H Kim, Ping Lu, Xiaofeng Wang, Haley E Manchester, Weishan Wang, Jeffrey R Haswell, Peter J Park, and Charles WM Roberts. "The SWI/SNF chromatin remodelling complex is required for maintenance of lineage specific enhancers". In: *Nature communications* 8.1 (2017), pp. 1–10.

[181] Lıgia Tavares, Emilia Dimitrova, David Oxley, Judith Webster, Raymond Poot, Jeroen Demmers, Karel Bezstarosti, Stephen Taylor, Hiroki Ura, Hiroshi Koide, et al. "RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3". In: *Cell* 148.4 (2012), pp. 664–678.

[182] Renjing Wang, Alexander B Taylor, Belinda Z Leal, Linda V Chadwell, Udayar Ilangovan, Angela K Robinson, Virgil Schirf, P John Hart, Eileen M Lafer, Borries Demeler, et al. "Polycomb group targeting through different binding partners of RING1B C-terminal domain". In: *Structure* 18.8 (2010), pp. 966–975.

[183] Sanchita Bhatnagar, Claude Gazin, Lynn Chamberlain, Jianhong Ou, Xiaochun Zhu, Jogender S Tushir, Ching-Man Virbasius, Ling Lin, Lihua J Zhu, Narendra Wajapeyee, et al. "TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein". In: *Nature* 516.7529 (2014), pp. 116–120.

[184] Ulrich Rothbauer, Kourosh Zolghadr, Serge Muyldermans, Aloys Schepers, M Cristina Cardoso, and Heinrich Leonhardt. "A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins". In: *Molecular & Cellular Proteomics* 7.2 (2008), pp. 282–289.

[185] Tetsuro Hirose, Giorgio Virnicchi, Akie Tanigawa, Takao Naganuma, Ruohan Li, Hiroshi Kimura, Takahide Yokoi, Shinichi Nakagawa, Marianne Bénard, Archa H Fox, et al. "NEAT1 long noncoding RNA regulates transcription via protein sequestration within subnuclear bodies". In: *Molecular biology of the cell* 25.1 (2014), pp. 169–183.

[186] Katsutoshi Imamura, Naoto Imamachi, Gen Akizuki, Michiko Kumakura, Atsushi Kawaguchi, Kyosuke Nagata, Akihisa Kato, Yasushi Kawaguchi, Hiroki Sato, Misako Yoneda, et al. "Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli". In: *Molecular cell* 53.3 (2014), pp. 393–406.

[187] Charles S Bond and Archa H Fox. "Paraspeckles: nuclear bodies built on long noncoding RNA". In: *Journal of Cell Biology* 186.5 (2009), pp. 637–644.

[188] E Colombo, M Alcalay, and PG Pelicci. "Nucleophosmin and its complex network: a possible therapeutic target in hematological diseases". In: *Oncogene* 30.23 (2011), pp. 2595–2609.

[189] Ju Yeon Lee, Robert J Lake, Jaewon Kirk, Vilhelm A Bohr, Hua-Ying Fan, and Sungchul Hohng. "NAP1L1 accelerates activation and decreases pausing to enhance nucleosome remodeling by CSB". In: *Nucleic acids research* 45.8 (2017), pp. 4696–4707.

[190] Iltaeg Cho, Pei-Fang Tsai, Robert J Lake, Asjad Basheer, and Hua-Ying Fan. "ATP-dependent chromatin remodeling by Cockayne syndrome protein B and NAP1-like histone chaperones is required for efficient transcription-coupled DNA repair". In: *PLoS Genet* 9.4 (2013), e1003407.

[191] JW Shih, TY Tsai, Chi-Hong Chao, and YH Wu Lee. "Candidate tumor suppressor DDX3 RNA helicase specifically represses cap-dependent translation by acting as an eIF4E inhibitory protein". In: *Oncogene* 27.5 (2008), pp. 700–714.

[192] Ming-Chih Lai, Wen-Cheng Chang, Sheau-Yann Shieh, and Woan-Yuh Tarn. "DDX3 regulates cell growth through translational control of cyclin E1". In: *Molecular and cellular biology* 30.22 (2010), pp. 5444–5453.

[193] Jonathan G Lees, Andrew J Miles, Robert W Janes, and BA Wallace. "Novel methods for secondary structure determination using low wavelength (VUV) circular dichroism spectroscopic data". In: *BMC bioinformatics* 7.1 (2006), p. 507.

[194] Simon Alberti, Amy Gladfelter, and Tanja Mittag. "Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates". In: *Cell* 176.3 (2019), pp. 419–434.

[195] Adam G Larson, Daniel Elnatan, Madeline M Keenen, Michael J Trnka, Jonathan B Johnston, Alma L Burlingame, David A Agard, Sy Redding, and Geeta J Narlikar. "Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin". In: *Nature* 547.7662 (2017), pp. 236–240.

[196] Amy R Strom, Alexander V Emelyanov, Mustafa Mir, Dmitry V Fyodorov, Xavier Darzacq, and Gary H Karpen. "Phase separation drives heterochromatin domain formation". In: *Nature* 547.7662 (2017), pp. 241–245.

[197] Andrea Cerase, Alexandros Armaos, Christoph Neumayer, Philip Avner, Mitchell Guttman, and Gian Gaetano Tartaglia. "Phase separation drives X-chromosome inactivation: a hypothesis". In: *Nature structural & molecular biology* 26.5 (2019), pp. 331–334.

[198] S Sanulli, MJ Trnka, V Dharmarajan, RW Tibble, BD Pascal, AL Burlingame, PR Griffin, JD Gross, and GJ Narlikar. "HP1 reshapes nucleosome core to promote phase separation of heterochromatin". In: *Nature* 575.7782 (2019), pp. 390–394.

[199] Marlinde L van den Boogaard, Richard JLF Lemmers, Judit Balog, Mariëlle Wohlgemuth, Mari Auranen, Satomi Mitsuhashi, Patrick J van der Vliet, Kirsten R Straasheijm, Rob FP van den Akker, Marjolein Kriek, et al. "Mutations in DNMT3B modify epigenetic repression of the D4Z4 repeat and the penetrance of

facioscapulohumeral dystrophy". In: *The American Journal of Human Genetics* 98.5 (2016), pp. 1020–1029.

[200] Michiko Hirano and Tatsuya Hirano. "Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA". In: *The EMBO journal* 21.21 (2002), pp. 5733–5744.

[201] Shige H Yoshimura, Kohji Hizume, Akiko Murakami, Takashi Sutani, Kunio Takeyasu, and Mitsuhiro Yanagida. "Condensin architecture and interaction with DNA: regulatory non-SMC subunits bind to the head of SMC heterodimer". In: *Current biology* 12.6 (2002), pp. 508–513.

[202] Madhusudhan Srinivasan, Johanna C Scheinost, Naomi J Petela, Thomas G Gligoris, Maria Wissler, Sugako Ogushi, James E Collier, Menelaos Voulgaris, Alexander Kurze, Kok-Lung Chan, et al. "The cohesin ring uses its hinge to organize DNA using non-topological as well as topological mechanisms". In: *Cell* 173.6 (2018), pp. 1508–1519.

[203] Young-Min Soh, Frank Bürmann, Ho-Chul Shin, Takashi Oda, Kyeong Sik Jin, Christopher P Toseland, Cheolhee Kim, Hansol Lee, Soo Jin Kim, Min-Seok Kong, et al. "Molecular basis for SMC rod formation and its dissolution upon DNA binding". In: *Molecular cell* 57.2 (2015), pp. 290–303.

[204] Christian H Haering, Jan Löwe, Andreas Hochwagen, and Kim Nasmyth. "Molecular architecture of SMC proteins and the yeast cohesin complex". In: *Molecular cell* 9.4 (2002), pp. 773–788.

[205] Alexander Kurze, Katharine A Michie, Sarah E Dixon, Ajay Mishra, Takehiko Itoh, Syma Khalid, Lana Strmecki, Katsuhiko Shirahige, Christian H Haering, Jan Löwe, et al. "A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion". In: *The EMBO journal* 30.2 (2011), pp. 364–378.

[206] Michiko Hirano, David E Anderson, Harold P Erickson, and Tatsuya Hirano. "Bimodal activation of SMC ATPase by intra-and inter-molecular interactions". In: *The EMBO journal* 20.12 (2001), pp. 3238–3250.

[207] Thomas E Melby, Charles N Ciampaglio, Gina Briscoe, and Harold P Erickson. "The symmetrical structure of structural maintenance of chromosomes (SMC)

and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge". In: *The Journal of cell biology* 142.6 (1998), pp. 1595–1604.

[208] Yosuke Hiramuki and Stephen J Tapscott. "Identification of SMCHD1 domains for nuclear localization, homo-dimerization, and protein cleavage". In: *Skeletal muscle* 8.1 (2018), p. 24.

[209] Richard Lemmers. *Leiden Open Variation Database*. Accessed on: 14.06.2020. URL: `http://databases.lovd.nl/shared/variants/SMCHD1/unique`.

[210] Zhubing Shi, Haishan Gao, Xiao-chen Bai, and Hongtao Yu. "Cryo-EM structure of the human cohesin-NIPBL-DNA complex". In: *Science* (2020).

[211] Jeongyoon Choi, Andreas Linus Bachmann, Katharina Tauscher, Christian Benda, Beat Fierz, and Jürg Müller. "DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation". In: *Nature structural & molecular biology* 24.12 (2017), p. 1039.

[212] Jean-François Trempe, Véronique Sauvé, Karl Grenier, Marjan Seirafi, Matthew Y Tang, Marie Ménade, Sameer Al-Abdul-Wahid, Jonathan Krett, Kathy Wong, Guennadi Kozlov, et al. "Structure of parkin reveals mechanisms for ubiquitin ligase activation". In: *Science* 340.6139 (2013), pp. 1451–1455.

[213] Felicia Gray, Hyo Je Cho, Shirish Shukla, Shihan He, Ashley Harris, Bohdan Boytsov, Łukasz Jaremko, Mariusz Jaremko, Borries Demeler, Elizabeth R Lawlor, et al. "BMI1 regulates PRC1 architecture and activity through homo-and hetero-oligomerization". In: *Nature communications* 7.1 (2016), pp. 1–12.

[214] David Komander and Michael Rape. "The ubiquitin code". In: *Annual review of biochemistry* 81 (2012), pp. 203–229.

[215] Charlotte A Scarff, Martin JG Fuller, Rebecca F Thompson, and Matthew G Iadanza. "Variations on negative stain electron microscopy methods: tools for tackling challenging systems". In: *JoVE (Journal of Visualized Experiments)* 132 (2018), e57199.

[216] Pawel Sledz, Heping Zheng, Krzysztof Murzyn, Maksymilian Chruszcz, Matthew D Zimmerman, Mahendra D Chordia, Andrzej Joachimiak, and Wladek Minor. "New surface contacts formed upon reductive lysine

methylation: improving the probability of protein crystallization". In: *Protein Science* 19.7 (2010), pp. 1395–1404.

[217] Glenn R Masson, John E Burke, Natalie G Ahn, Ganesh S Anand, Christoph Borchers, Sébastien Brier, George M Bou-Assaf, John R Engen, S Walter Englander, Johan Faber, et al. "Recommendations for performing, interpreting and reporting hydrogen deuterium exchange mass spectrometry (HDX-MS) experiments". In: *Nature methods* 16.7 (2019), pp. 595–602.

[218] Changill Ban, Murray Junop, and Wei Yang. "Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair". In: *Cell* 97.1 (1999), pp. 85–97.

[219] James L Cole, Jeffrey W Lary, Thomas P Moody, and Thomas M Laue. "Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium". In: *Methods in cell biology* 84 (2008), pp. 143–179.

[220] Harald Wegele, Paul Muschler, Melanie Bunck, Jochen Reinstein, and Johannes Buchner. "Dissection of the contribution of individual domains to the ATPase mechanism of Hsp90". In: *Journal of Biological Chemistry* 278.41 (2003), pp. 39303–39310.

[221] Galen A Collins and Alfred L Goldberg. "Proteins containing ubiquitin-like (Ubl) domains not only bind to 26S proteasomes but also induce their activation". In: *Proceedings of the National Academy of Sciences* 117.9 (2020), pp. 4664–4674.

[222] Hyoung Tae Kim and Alfred L Goldberg. "UBL domain of Usp14 and other proteins stimulates proteasome activities and protein degradation in cells". In: *Proceedings of the National Academy of Sciences* 115.50 (2018), E11642–E11650.

[223] Véronique Sauvé, Asparouh Lilov, Marjan Seirafi, Marta Vranas, Shafqat Rasool, Guennadi Kozlov, Tara Sprules, Jimin Wang, Jean-François Trempe, and Kalle Gehring. "A Ubl/ubiquitin switch in the activation of Parkin". In: *The EMBO journal* 34.20 (2015), pp. 2492–2505.

[224] Kaspar Hollenstein, Roger JP Dawson, and Kaspar P Locher. "Structure and mechanism of ABC transporter proteins". In: *Current opinion in structural biology* 17.4 (2007), pp. 412–418.

[225] Karl-Peter Hopfner and John A Tainer. "Rad50/SMC proteins and ABC transporters: unifying concepts from high-resolution structures". In: *Current opinion in structural biology* 13.2 (2003), pp. 249–255.

[226] Da-Qiang Li, Sujit S Nair, and Rakesh Kumar. "The MORC family: new epigenetic regulators of transcription and DNA damage response". In: *Epigenetics* 8.7 (2013), pp. 685–693.