4.3.2.3.2 SWISS-MODEL analysis

Superimposition of MORC4 models from the MORC2 and MORC3 templates revealed that the ATPase domains of both models were similar (Figure 4.21a) except for a small number of residue substitutions (Figure 4.21b). However, it was evident that the MORC3 model predicted more of the MORC4 model as it had a 3' extension to 465 residues compared to the MORC2 template which only predicted up to 309 residues and did not include the Zf-CW domain (Figure 4.21a and b). In addition, the MORC3 template MORC4 model predicted the N-terminal of MORC4 to begin from residue 30, unlike MORC2 which was predicted from residue 39 (Figure 4.21b). These model predictions of MORC4 suggest that the C-terminus of MORC4 is more difficult to predict, which could be because neither MORC2 nor MORC3 have a C-terminus similar to that of MORC4.

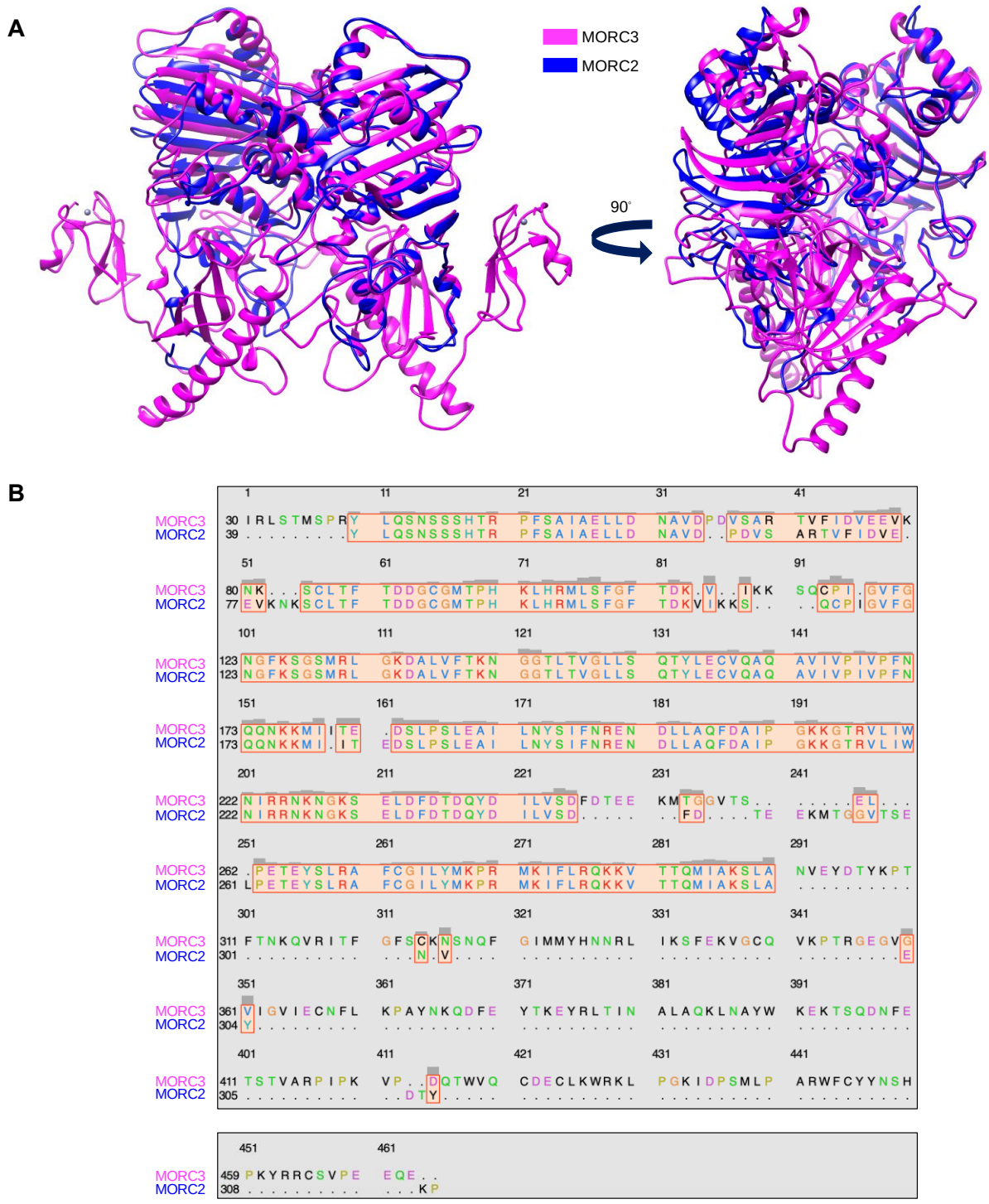


Figure 4.21: **Superimposed homology model of MORC4 using both MORC2 and MORC3 as templates.** SWISS-MODEL structural prediction software was used to create this superimposition of the structure of MORC4 using MORC2 and MORC3 as templates. **(A)** Superimposed model of MORC4 also rotated at 90° and created using UCSF Chimera software. **(B)** Aligned protein sequences of both MORC4 models. MORC2 PDB: 5OF9 MORC3 PDB: 6O1E.

4.3.2.4 Comparison of Phyre2 and SWISS-MODEL MORC4 homology models

In an attempt to assess the accuracy of each MORC4 homology model, the models created with the same template but with different homology model engines were superimposed using UCSF Chimera. The MORC4 homology models generated from the MORC3 template showed the ATPase domain from both the SWISS-MODEL and Phyre2 models were extremely similar (Figure 4.22a and b). However, there were some substitutions of residues and some deletions, specifically, a large deletion was found in the Phyre2 model from 395-439 residues in this MORC3 template model (Figure 4.22a and b), which is evident in both the homology model and protein sequence alignment. However, this deletion was not seen in the SWISS-MODEL MORC4 homology model using the MORC3 template, providing further evidence that the SWISS-MODEL results in a more accurate prediction of MORC4. In addition, the SWISS-MODEL and Phyre2 models had a QMEAN value of -2.13 (Table 4.1), respectively. This suggests that the Phyre2 model was of higher quality than the SWISS-MODEL model, indicating that it is more accurate. However, the Phyre2 model was produced using a MORC3 model which was bound to the H3K4me3 peptide, suggesting that it may not be as accurate as the SWISS-MODEL model. This binding to the H3K4me3 peptide could also explain why both of these models have different C-termini (Figure 4.22) as the SWISS-MODEL model was not bound to a peptide.

The MORC4 homology model resulting from the MORC2 template using both SWISS-MODEL and Phyre2 showed that both models had a similar ATPase domain arrangement (Figure 4.23a and b). However, as with the MORC3 template model, these models also contained multiple residue substitutions (Figure 4.23b). In addition, the SWISS-MODEL MORC4 model ended at 299 residues whereas, the Phyre2 model continued to 468 residues. The Phyre2 model also had a longer N-terminus than the SWISS-MODEL model, where it began with a glycine at residue 1 instead of the SWISS-MODEL model which began with a

tyrosine at residue 11 (Figure 4.23b). Furthermore, the QMEAN of the SWISS-MODEL and Phyre2 models were -4.51 (Table 4.1), respectively. This suggests that both models were low quality, indicating that the MORC3 template MORC4 model is a more accurate prediction of MORC4 as it is higher quality. Overall, homology modelling and statistical analysis of each model determined that the MORC4 model which used MORC3 as a template was higher quality than the MORC2 template model (Table 4.1).

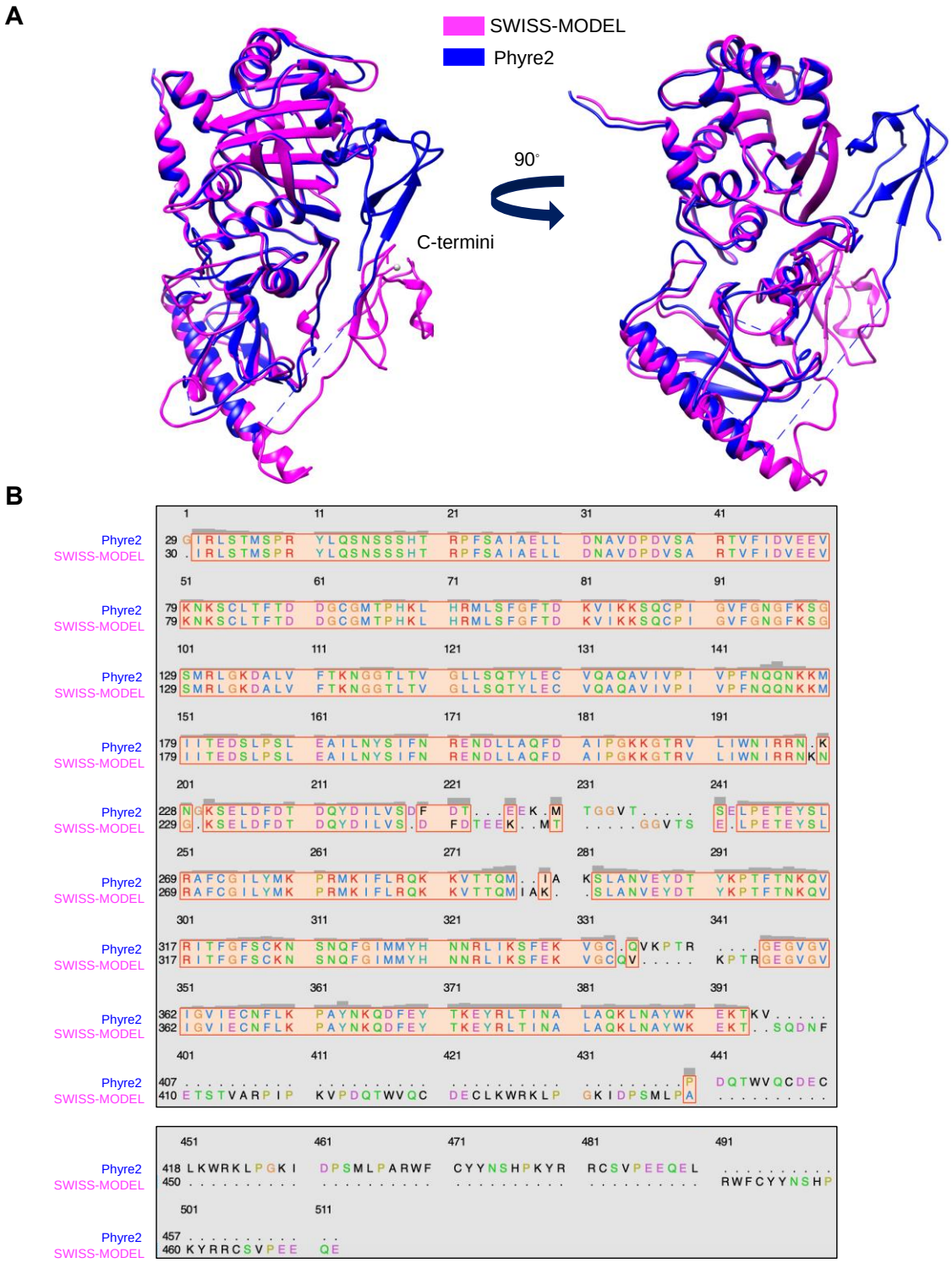


Figure 4.22: **Superimposed homology model of MORC4 using MORC3 as a template in both Phyre2 and SWISS-MODEL.** Phyre2 and SWISS-MODEL structural prediction software was used to create this superimposition of the structure of MORC4 using MORC3 as a template. **(A)** Superimposed model of MORC4 also rotated at 90° and created using UCSF Chimera software. The Phyre2 model is shown in blue and the SWISS-MODEL model is shown in magenta pink. **(B)** Aligned protein sequences of both MORC4 models. SWISS-MODEL PDB: 6O1E. Phyre2 PDB: 5IX1.

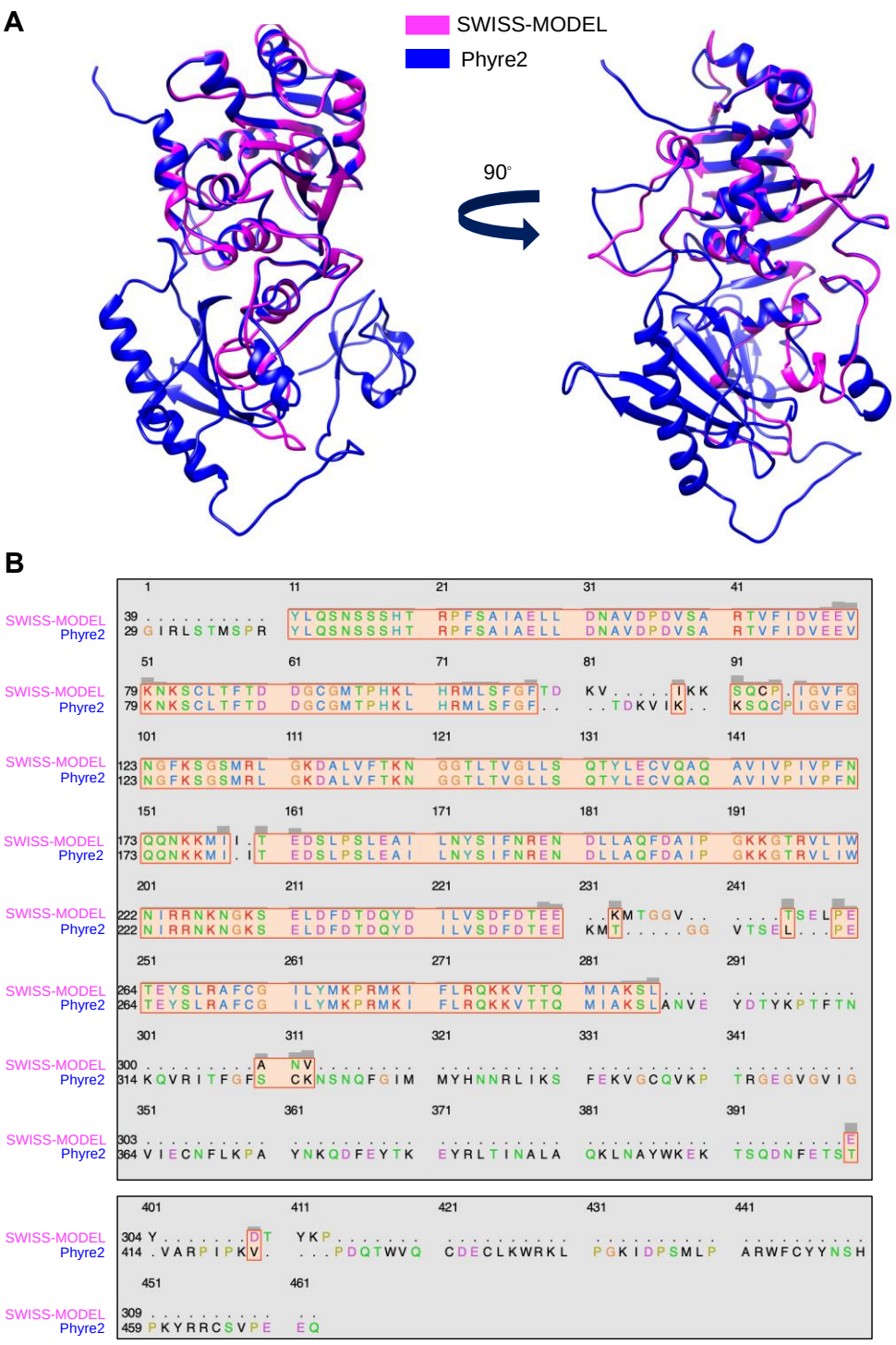


Figure 4.23: **Superimposed homology model of MORC4 using MORC2 as a template in both Phyre2 and SWISS-MODEL.** Phyre2 and SWISS-MODEL structural prediction software was used to create this superimposition of the structure of MORC4 using MORC2 as a template. **(A)** Superimposed model of MORC4 also rotated at 90° and created using UCSF Chimera software. The Phyre2 model is shown in blue and the SWISS-MODEL model is shown in magenta pink. **(B)** Aligned protein sequences of both MORC4 models. SWISS-MODEL PDB: 5OF9. Phyre2 PDB: 5OFB.

4.3.2.5 Robetta analysis

Robetta homology modelling is a tool used to create models from either a model with a known protein structure template and/or using the *de novo* Rosetta fragment insertion method (Kim *et al.*, 2004; Song *et al.*, 2013) The *de novo* Rosetta fragment insertion method is used to predict structures of unresolved proteins, using the basic local alignment search tool (BLAST) (Kim *et al.*, 2004; Raman *et al.*, 2009). This method of structural prediction involves identifying the lowest free-energy structure of each predicted residue in the protein (Philip Bradley *et al.*, 2005).

Following Robetta analysis of full length MORC4 (1-900 residues), several models were predicted. Model A (Figure 4.24a) was the top hit for this MORC4 protein model. It had a global distance test (GDT) score of 0.42, with 1.0 being good quality and 0 being bad quality. Each model of MORC4 also had an angstrom error estimate. Model A (Figure 4.24a) had an angstrom error estimate between 0-10 Å for the ATPase domain, suggesting that this domain is similar to the template sequence. However, from residue 500, the angstrom error increased to 50 Å. This suggests that the C-terminus of this MORC4 model is dissimilar to the template sequence, as neither the MORC2 or MORC3 full C-terminal structures have yet been solved. These models also indicate that MORC4 exists as a dimer except for model E (Figure 4.24e), which is monomeric. Models B-E (Figure 4.24b-e) had a higher angstrom error estimate, suggesting that they were less likely to be an accurate representation of the structure of MORC4. In addition, the C-termini appear to be different to the SWISS-MODEL and Phyre2 models of MORC4 as some of them have long extended C-terminal tails (Figure 4.24). In particular, model E appeared to be an unlikely structure of MORC4, which is also supported by the high angstrom error estimate. MORC4 is likely to be a globular protein without an extended C-terminal tail, similar to other MORC proteins. Furthermore, model E does not show

dimerisation (Figure 4.24), dissimilar to MORC2 (Douse *et al.*, 2018) and MORC3 (Zhang *et al.*, 2019).

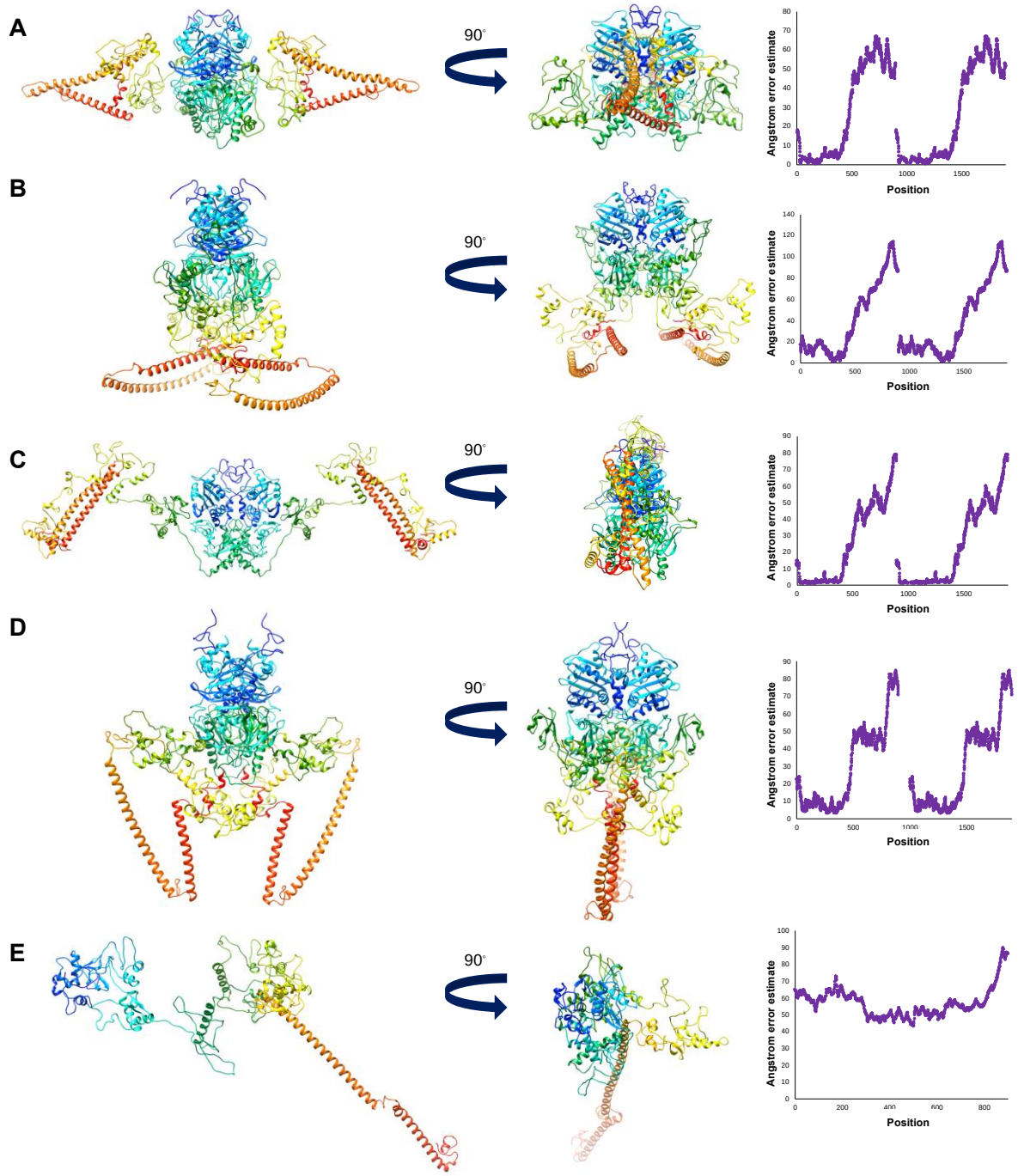


Figure 4.24: **Homology model of MORC4 using Robetta.** Robetta structural prediction software was used to create these predictions of the structure of MORC4. Each model of MORC4 has also been rotated at 90° and created using UCSF Chimera software. The corresponding plot shows the angstrom error estimate of each residue in the model. **(A-E)** MORC4 models in order of quality of the model (A = good-quality and E = poor-quality).

4.3.2.6 Mutations in MORC proteins prevent binding of AMP-PNP and peptides

N39A and D68A MORC2 mutations were previously found to abrogate the binding of AMP-PNP and Mg^{2+} (Douse *et al.*, 2018; D Q Li *et al.*, 2012). This was also evident following structural analysis of WT MORC2, which was used in USCF Chimera to show that specific mutations prevented AMP-PNP binding (Figure 4.26). Following USCF Chimera mutagenesis, the N39A MORC2 mutation confirmed that binding of Mg^{2+} was no longer possible (Figure 4.25a), therefore, MORC2 could no longer bind to AMP-PNP (Douse *et al.*, 2018). This was due to a substitution mutation from an asparagine (N) to an alanine (A), where the A residue could no longer bind to Mg^{2+} (Figure 4.25a) through the formation of hydrogen bonds, ultimately preventing binding to AMP-PNP. In addition, the MORC2 D68A mutation analysis confirmed that AMP-PNP could no longer bind to MORC2 (Figure 4.25b) (D Q Li *et al.*, 2012). The W419K mutation in MORC3 was previously found to interrupt binding to the H3K4me3 peptide (H Andrews *et al.*, 2016). USCF Chimera analysis also revealed that the MORC4 W419K mutation, which changed from a residue with a hydrophobic side chain (W) to a residue with a positive side chain (K), also interfered with binding to the H3K4me3 peptide (Figure 4.26).

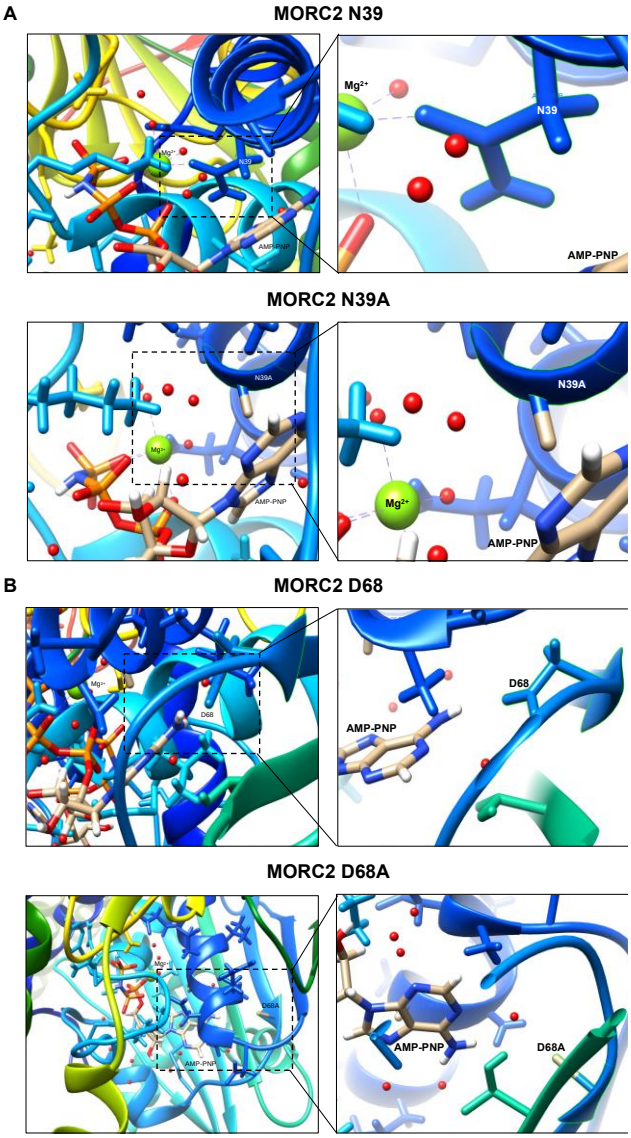


Figure 4.25: **MORC2 mutations.** UCSF chimera software was used to create known mutations in MORC2 (PDB: 5OF9), N39A and D68A. Images were taken before and after the mutations. (**A)** MORC2 N39A mutation. (**B)** MORC2 D68A mutation. Mg^{2+} (green sphere), AMP-PNP and mutated residues are labelled.

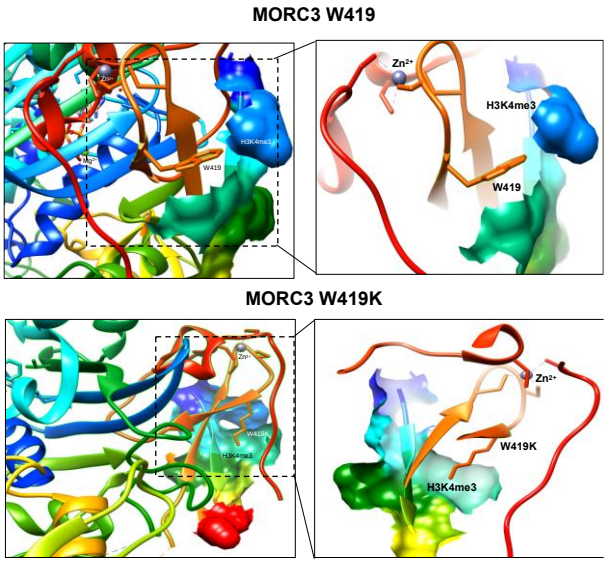


Figure 4.26: **MORC3 W419K mutation.** UCSF chimera software was used to create a known mutation in MORC3 (PDB: 5IX1), W419K. Images were taken before and after the mutations. Mg^{2+} (green sphere), Zn^{2+} (grey sphere), H3K4me3 peptide (solid surface) and residues of interest residues are labelled.

USCF Chimera analysis of the MORC4 homology model created using Phyre2 (using the 5IX1 MORC3 template) revealed that MORC4 contains the residues required to bind both AMP-PNP/Mg^{2+} and the H3K4me3 peptide. USCF Chimera mutagenesis also supported prevention of Mg^{2+}, AMP-PNP and peptide binding, previously seen in MORC2 (Douse *et al.*, 2018; D Q Li *et al.*, 2012) and MORC3 (H Andrews *et al.*, 2016), which is likely caused by abrogation of the normal hydrogen bond patterns. This suggests that MORC4 could also bind to AMP-PNP and the H3K4me3 peptide, leading to dimerisation and chromatin remodelling, respectively. This also indicates that introducing mutations to these regions of MORC4 will prevent AMP-PNP binding and peptide binding, similar to MORC2 and MORC3. MORC4 mutations which were shown to interfere with AMP-PNP binding and peptide binding includes N60A (Figure 4.27a), D88A (Figure 4.27b) and W435K (Figure 4.27d). USCF Chimera analysis also suggested that the K126R mutation in MORC4 interferes with the ATPase domain (Figure 4.27c).

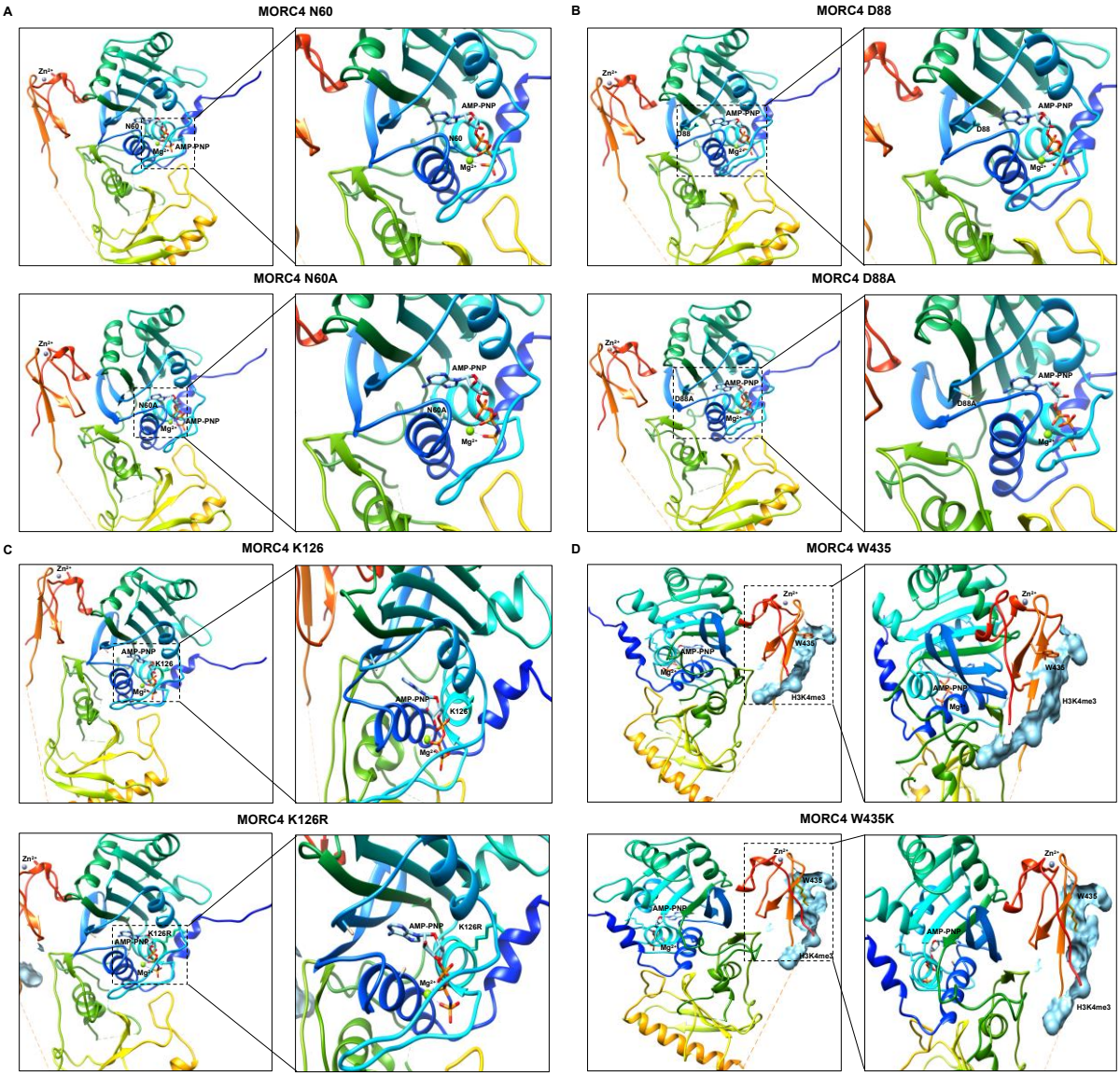


Figure 4.27: **MORC4 mutations using a MORC4 homology model.** UCSF chimera software was used to create mutations, N60A, D88A, K126E and W435K, in a MORC4 homology model created using Phyre2 (PDB: 5IX1). Images were taken before and after the mutations. (**A)** MORC4 N60A mutation. (**B)** MORC2 D88A mutation. (**C)** MORC4 K124R mutation. (**D)** MORC4 W435K mutation. Mutated residues are labelled.

4.4 Biochemical characterisation of MORC4

Biochemical characterisation is important for confirming any predictions from structural/modelling analysis, which could help to connect potential hypotheses about biochemical mechanisms. Evidence suggests that both MORC2 and MORC3 dimerise on binding ATP, allowing catalytical activity and DNA repair mechanisms such as chromatin remodelling (Zhang *et al.*, 2019; Xie *et al.*, 2019). This characterisation of MORC2 and

MORC3 was carried out with a number of biochemical and biophysical studies. This project utilised similar studies to characterise MORC4 using chemical crosslinking and size exclusion chromatography (SEC) oligomerisation analysis approaches to assess if similar dimerisation occurs. In addition, WT and site-directed mutant forms of MORC4 were utilised to investigate this potential biochemical mechanism.

4.4.1 Site-directed mutagenesis (SDM) of MORC4 fragments

4.4.1.1 Design of site-directed mutagenesis (SDM) approach

Previous studies on MORC proteins have suggested that MORC2 and MORC3 have ATPase activity (D Q Li *et al.*, 2012; H Andrews *et al.*, 2016) and that MORC3 binds histone peptides (Li *et al.*, 2016). To investigate MORC4 biochemistry as with MORC2 and MORC3, several substitution/point mutations were identified which were used in previous studies for MORC2 (D Q Li *et al.*, 2012; Douse *et al.*, 2018) and MORC3 (H Andrews *et al.*, 2016) (Table 4.2).

Table 4.2: **MORC4 mutants and their potential function.**

MORC4 mutation	MORC2	MORC3	Mutant Phenotype	Reference
N60A	N39A	-	Prevents binding of the ligand AMP-PNP and Mg^{2+}	(Douse *et al.*, 2018)
D88A	D68A	-	Prevents binding and hydrolysis of ATP	(D Q Li *et al.*, 2012)
K126R	-	-	Prevents binding of ATP	(Wiese *et al.*, 2006)
W435K	-	W419K	Prevents binding to H3K4me3	(H Andrews *et al.*, 2016)

There are two known motifs of the Walker ATPase domain, A and B (Romero Romero *et al.*, 2018). The Walker A motif is important for ATP binding and the Walker B motif is responsible for the formation of hydrogen bonds and ATP hydrolysis (Wiese *et al.*, 2006). The MORC2 D88A mutation prevents binding and hydrolysis of ATP, which was confirmed using ATPase activity assays. This D88A mutation changes the Walker B motif from aspartic acid (D), which binds to Mg^{2+} in order to stimulate ATP hydrolysis (Chiraniya *et al.*, 2013), to alanine (A), which loses the negatively-charged carboxylate side chain. This MORC2 study also found that following DNA damage via ionising radiation (IR), chromatin relaxation was inhibited in cells with the D88A mutant (D Q Li *et al.*, 2012). This indicates that ATPase activity is required for chromatin remodelling. MORC2 N39A (MORC4 N60A) is restricted from dimerising by preventing the binding of AMP-PNP/Mg^{2+} (Douse *et al.*, 2018)**.** The MORC3 W419K (MORC4 W345K) mutation completely abrogated binding to H3K4me3 (H Andrews *et al.*, 2016), suggesting that the CW domain is required for chromatin remodelling. In addition, the Walker A motif, which is important for ATP binding (Wiese *et al.*, 2006), was identified in MORC4 mutated from a lysine (K) to a arginine (R) in an attempt to abrogate ATP binding. Following identification of MORC4 mutations using such previous MORC studies (Figure 4.28), MORC4 mutant primers containing the substitution mutations were designed with the intention of testing if the equivalent mutants affect ATPase activity and/or dimerisation (Table 4.3).

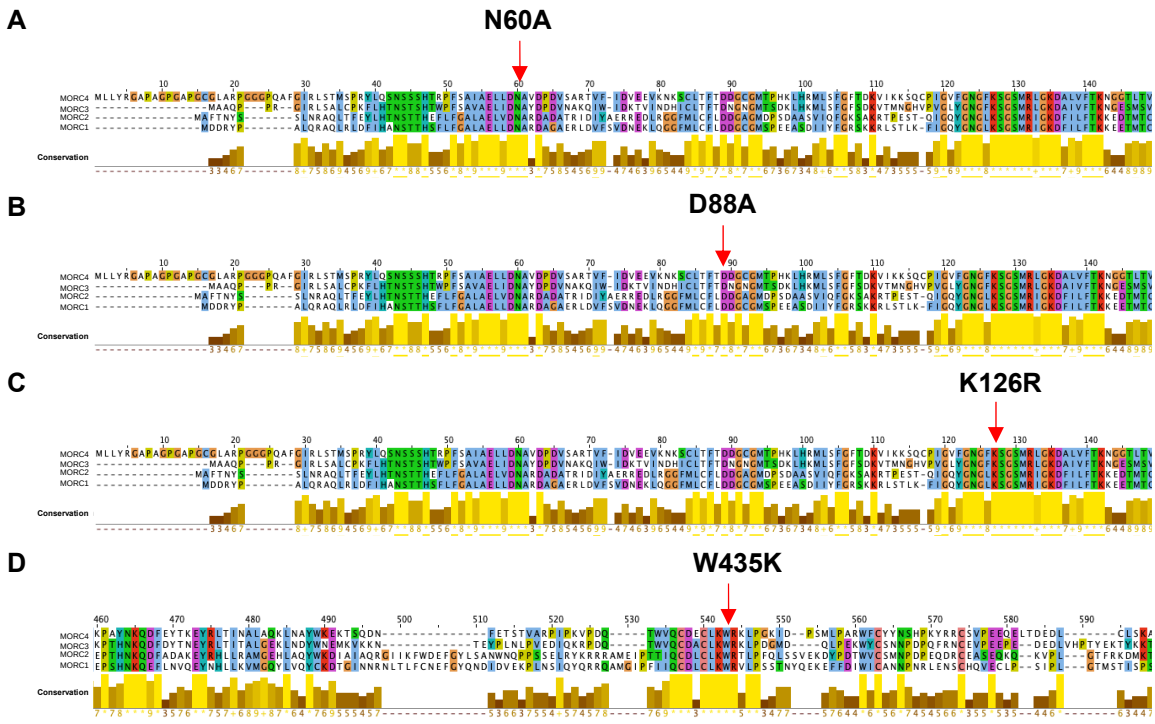


Figure 4.28: **Multiple sequence alignment of protein sequences of all human MORC proteins (MORC1-4) showing mutations.** Human MORC protein sequences were used for a MUSCLE multiple sequence alignment using Jalview software and MORC4 residues used for mutagenesis are highlighted with a red arrow. **(A)** N60A **(B)** D88A **(C)** K126R **(D)** W435K. The conservation score across all MORCs was shown below the alignment (Yellow is the most conserved and brown or no colour is the least conserved). The multiple sequence alignment residues were coloured using ClustalX colours.

Table 4.3: MORC4 **primers containing the substitution mutation.** Mutations are shown in underlined.

Primer	Mutation	Forward Primer	Reverse Primer
FLMORC4_N60A_F	N60A	GATGCTGCTGTAGATCCAGAT	-
FLMORC4_N60A_R	N60A	-	TAGCAGCTCCGCGATGGC
FLMORC4_D88A_F	D88A	ACCGCTGATGGATGTGGGATG	-
FLMORC4_D88A_R	D88A	-	AAAGGTCAAACAAGATTT
FLMORC4_K126R_F	K126R	TTCCGGTCAGGCTCCATGCGG	-
FLMORC4_K126R_R	K126R	-	ACCATTACCAAAGACCCC
FLMORC4_W435K_F	W435K	AAAAAGAGAAAGCTTCCTGGG	-
FLMORC4_W435K_R	W435K	-	AAGACACTCATCACACTG

4.4.1.1.1 PCR amplification of *MORC4* with substitution mutations

Initially, the full-length wild type gene encoding for FLAG-tagged MORC4$^{1\text{-}900}$ (Chapter 5, section 5.3.1) was used as the DNA template for site-directed mutagenesis. However, following screening of the PCR products it was revealed that amplification from this template-DNA was unsuccessful as there was no band visible at approximately 10.0 Kb (Figure 4.29), except for the D88A mutant, where there was some amplification seen with the sample using MyTaq buffer with a mixture of Phusion/MyTaq polymerase (Figure 4.29).

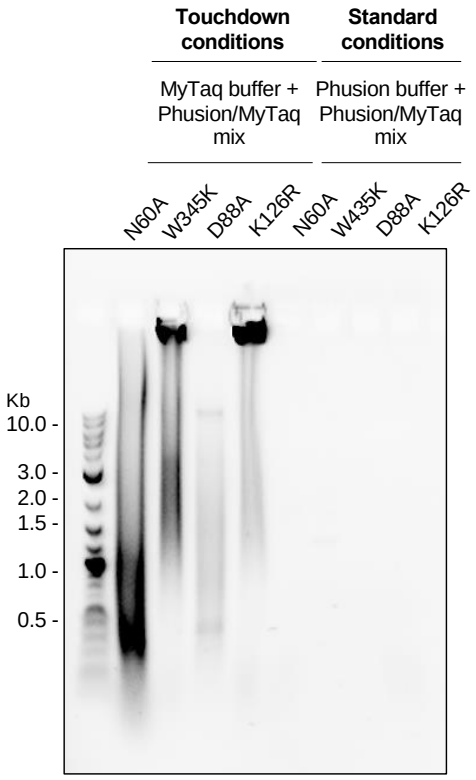


Figure 4.29: **Amplification of the full-length wild type gene encoding for FLAG-tagged MORC4$^{1\text{-}900}$ using PCR primers with substitution mutations.** Following amplification of the wild type gene encoding for MORC4$^{1\text{-}900}$ with mutagenic primers, PCR samples were resolved on a 0.8% agarose gel. Each mutant was amplified using MyTaq buffer with a mix of Phusion/MyTaq polymerase with touchdown conditions and Phusion buffer with a mix of Phusion/MyTaq polymerase with standard conditions.

To overcome the amplification issue with the full-length MORC4 gene, the gene encoding for an ATPase-Zf-CW truncated fragment of MORC4, MORC4$^{29\text{-}480}$ was instead used for PCR amplification to produce ATPase mutants including N60A, W345K, D88A and K126R,under

both standard and touchdown conditions. The gene encoding a MORC4 CW/C-terminal truncated fragment, MORC4$^{420-757}$, was also amplified to produce a C-terminal mutant, W345K, using mutagenic primers under both standard and touchdown conditions (Figure 4.30) using Phusion buffer and Phusion polymerase. Samples of *MORC4* amplified with mutagenic primers under touchdown conditions were pooled for subsequent cloning. It was important to create both ATPase and C-terminal mutants in order to investigate the CW/C-terminal in absence of the ATPase domain.

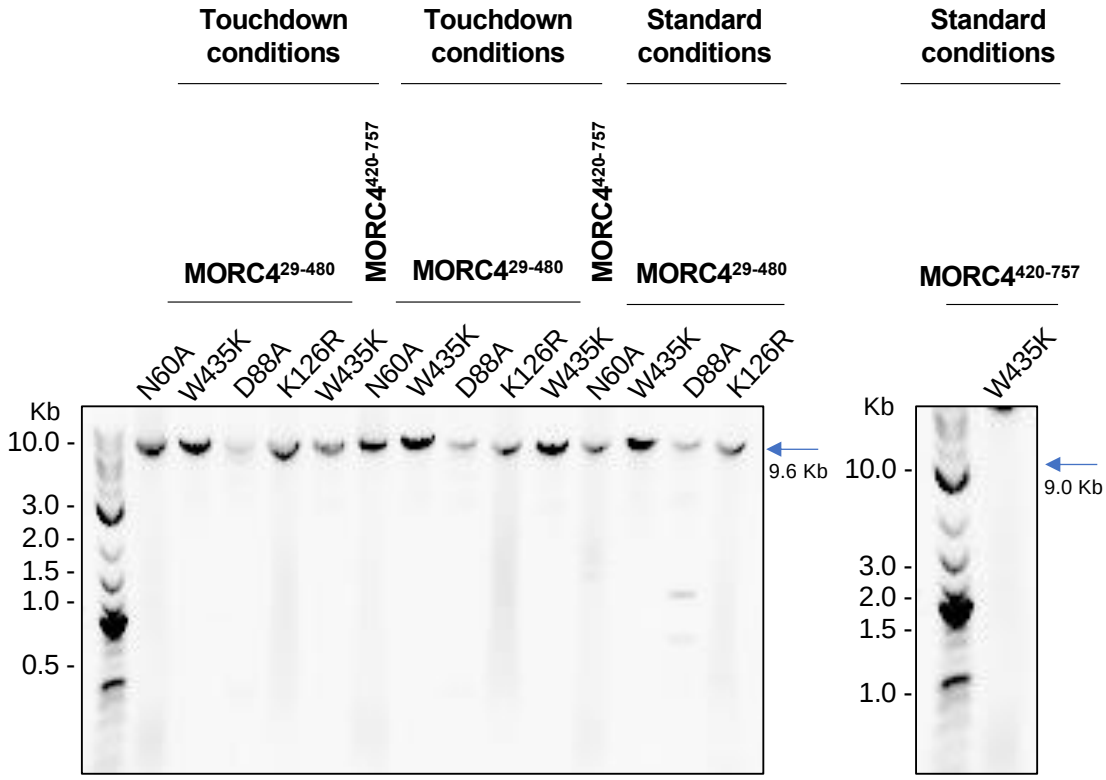


Figure 4.30: **An example of amplification of truncated fragments of *MORC4* using PCR primers with substitution mutations.** Following amplification of truncated fragments of MORC4 DNA with Phusion buffer and Phusion polymerase, samples were resolved on a 0.8% agarose gel. N60A, W345K, D88A and K126R mutagenic primers were used for amplification of MORC4$^{29-480}$ and W345K mutagenic primer was also used for amplification of MORC4$^{420-757}$ under both touchdown and standard conditions. Blue arrows indicate expected sizes of PCR products. The blue arrow indicates expected sizes of amplicons.

4.4.1.1.2 PCR screening of mutated *MORC4* fragments

Following successful amplification of truncated fragments of *MORC4* with mutagenic primers, PCR screens were used to confirm successful transformations of mutant DNA. The PCR screen revealed that at least one sample containing K126R, D88A, W435K (Appendix 8.16) and N60A (Figure 4.30) MORC4$^{29\text{-}480}$ mutants were of the correct size at 2.0 kilobases (Appendix 8.17 and 8.18). Further Zf-CW domain W435K mutants to the MORC4$^{420\text{-}757}$, MORC4$^{420\text{-}677}$ and MORC4$^{420\text{-}480}$ Zf-CW, were also cloned and were of the correct size (1.3 kb and 0.8 kb, respectively) (Appendix 8.18). All the *MORC4* mutant genes were sent for DNA sequencing to confirm the presence of the substitution mutation in the cloned constructs.

4.4.2 Heterologous expression and purification of MORC4 wild type and mutants

Mutated cloned MORC4 fragments in an *E. coli* expression vector were initially used for 50 mL test protein expression tests in *E. coli*. Mutated MORC4 with a high yield of protein following a 50 mL test expression were subsequently used for large-scale protein expression and purification. Mutant MORC4 proteins that expressed a high yield of protein and were fully cleavable were used for large-scale expression and purification, towards biochemical assays.

4.4.2.1 50 mL test expressions determined high yielding mutants of MORC4

Following mutagenesis, 50 mL protein expression tests were required to ensure that all mutants could express a significant amount of protein similar to the wild type (WT) seen in Chapter 3, to facilitate purification of proteins and to have sufficient material for analysis. The ATPase-Zf-CW MORC4$^{29\text{-}480}$ K126R, D88A and N60A mutants appeared to express the same level of protein as the WT in both LB and TB (Figure 4.31a and b). However, MORC4$^{29\text{-}480}$ N60A appeared to also express a smaller contaminant in LB (Figure 4.31a). The MORC4$^{29\text{-}480}$ W434K mutant expressed at a lot lower level than all mutant proteins in both LB and TB, indicating that this mutant protein may not be suitable for use in subsequent biochemical assays, if there is not a sufficient yield of protein (Figure 4.31a and b). However, the Zf-CW MORC4$^{420\text{-}757}$,

MORC4$^{420-677}$ and MORC4$^{420-480}$ W435K mutants expressed protein at a similar level to the WT in both LB and TB media (Figure 4.31a, b, c and d). Overall, mutant MORC4 expressed to higher levels in TB media compared to LB, which was also seen for WT 50 mL test expressions in Chapter 3 (Section 3.2.1). This suggests that in order to achieve a high expression of protein, TB will be used for subsequent large-scale protein expression and purification experiments, similar to Chapter 3.

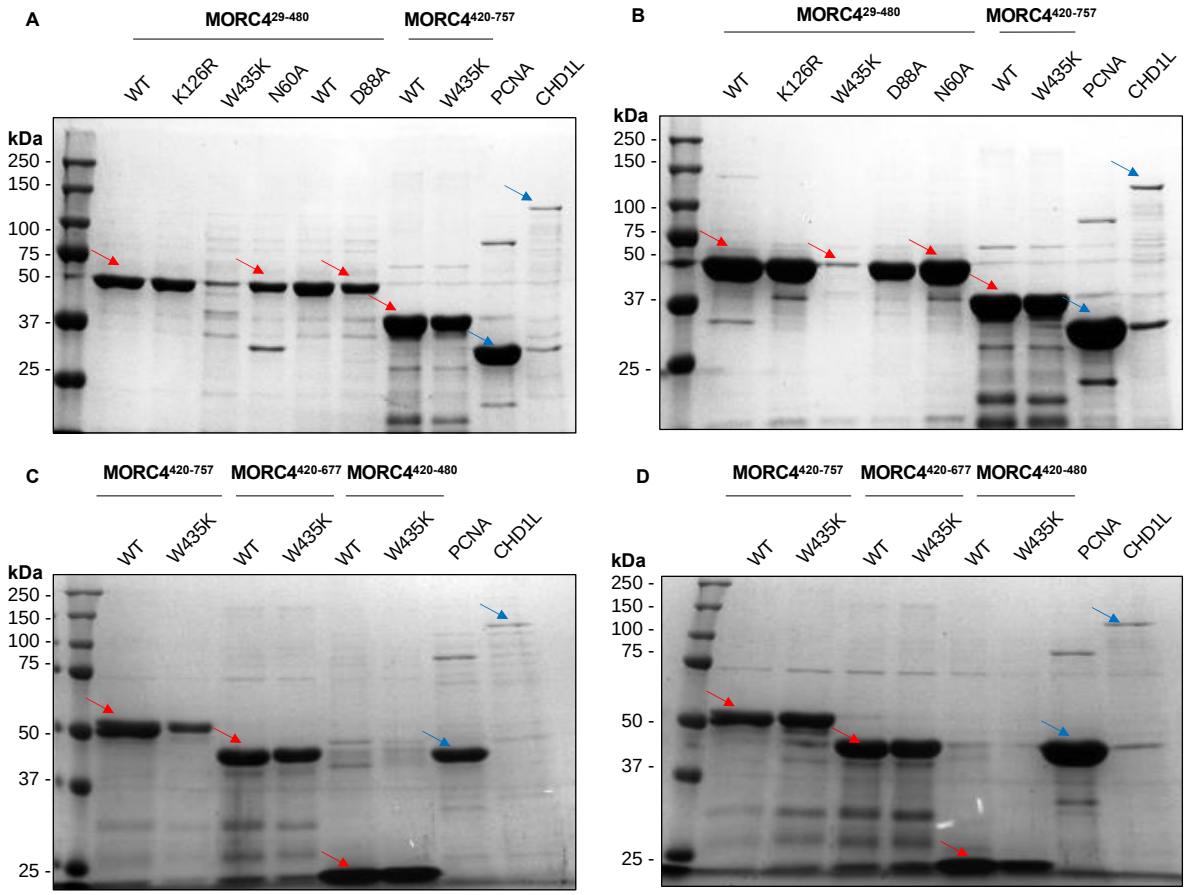


Figure 4.31: **50 mL test expressions of mutated truncated fragments of MORC4 in LB and TB.** Mutated proteins were expressed in 50 mL LB or TB, purified using a Ni-IMAC column and resolved on a 10% tris-glycine SDS (TGS)-PAGE gel. ATPase-Zf-CW MORC4$^{29-480}$ WT, K126R, W435K, N60A and D88A were expressed and purified as well as Zf-CW MORC4$^{420-757}$, MORC4$^{420-677}$ and MORC4$^{420-480}$ WT and W435K, in (**A and C)** LB and (**B and D)** TB. Red arrows indicate examples of expressed proteins and blue arrows indicate the positive controls (PCNA and CHD1L).

4.4.2.2 Large-scale protein expression and purification

Following 50 mL test expressions of WT and mutant MORC4 proteins, large-scale protein expression and purification of WT MORC4 and mutants was required to create protein suitable for use in subsequent biochemical experiments and assays. Unlike, structural protein analysis methods, biochemical assays do not require such a high quantity of protein. A MORC3 study only required micromolar concentrations of protein sample to investigate ATPase activity compared to a high concentration of protein sample for X-ray crystallography studies (Zhang *et al.*, 2019) (Chapter 3, section 3.3.2).

4.4.2.2.1 Purification of ATPase-Zf-CW MORC4 wild type and mutants

4.4.2.2.1.1 Ni-IMAC and TEV rebind of MORC4$^{29-480}$ WT and mutants

MORC4$^{29-480}$ WT, N60A, D88A and K126R were expressed in 2 L of TB. The initial Ni-IMAC purification gel shows that all of the proteins were expressed at a similar level to the WT and of the same size (63 kDa) (Figure 4.32a, b, c and d). Samples E300 1-5 were pooled and incubated overnight at 4°C with TEV protease then an additional Ni-IMAC purification was performed. Following incubation with TEV protease and Ni-IMAC, it was evident that all proteins (WT and mutants) had been cleaved as they reduced in size to 50 kDa (Figure 4.33a, b, c and d).

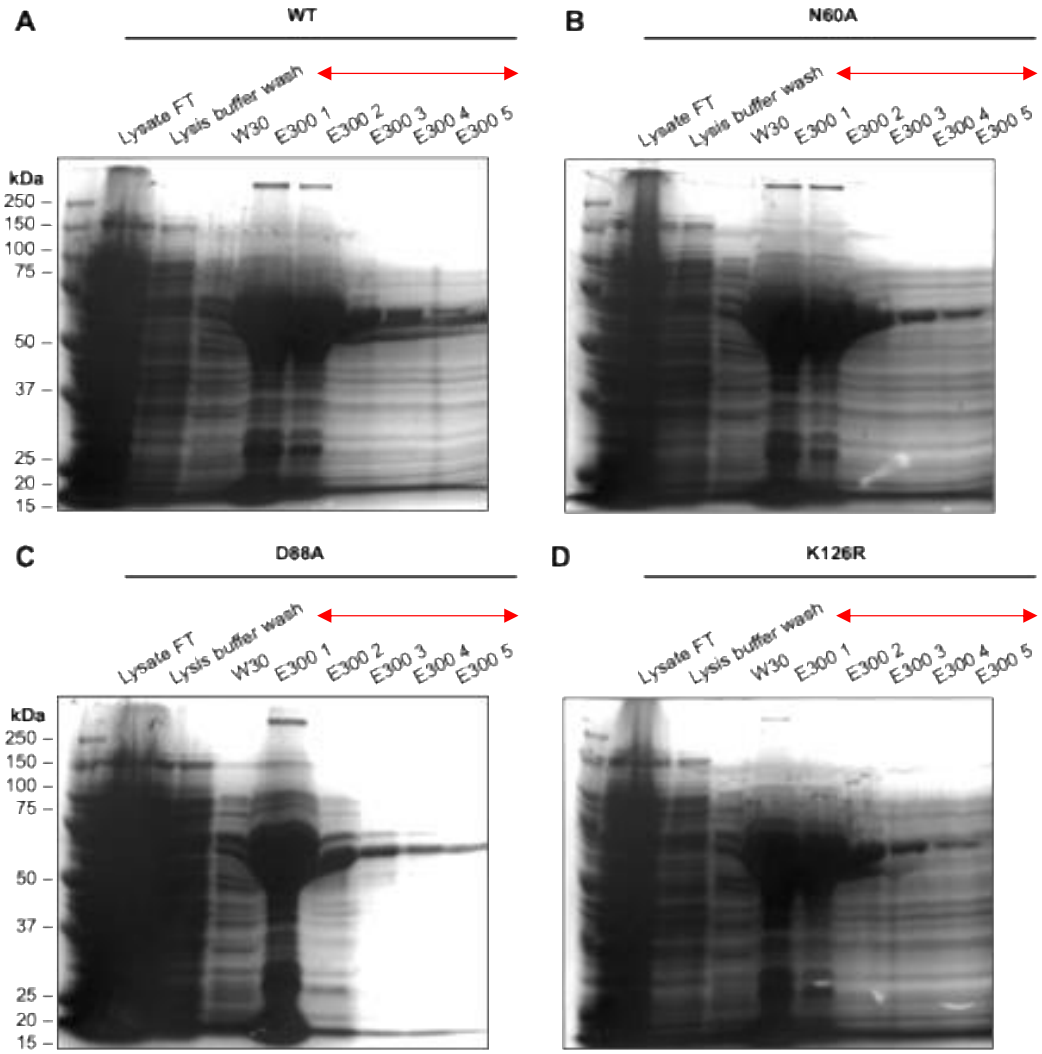


Figure 4.32: **2 L expression and initial purification step of wild type (WT) MORC4$^{29\text{-}480}$ and mutants using Ni-IMAC.** Each MORC4 protein was expressed in 2 L TB and purified using a Ni-IMAC gravity flow column. Samples were resolved on a 10% tris-glycine SDS (TGS)-PAGE gels. Each gel shows lysate flow through (FT), lysis buffer wash, W30 (wash 30 mM imidazole) and 5 E300 (elution 300 mM imidazole) washes. **(A)** WT MORC4$^{29\text{-}480}$ **(B)** N60A MORC4$^{29\text{-}480}$ **(C)** D88A MORC4$^{29\text{-}480}$ **(D)** K126R MORC4$^{29\text{-}480}$. Red arrows indicate protein fractions which were pooled.

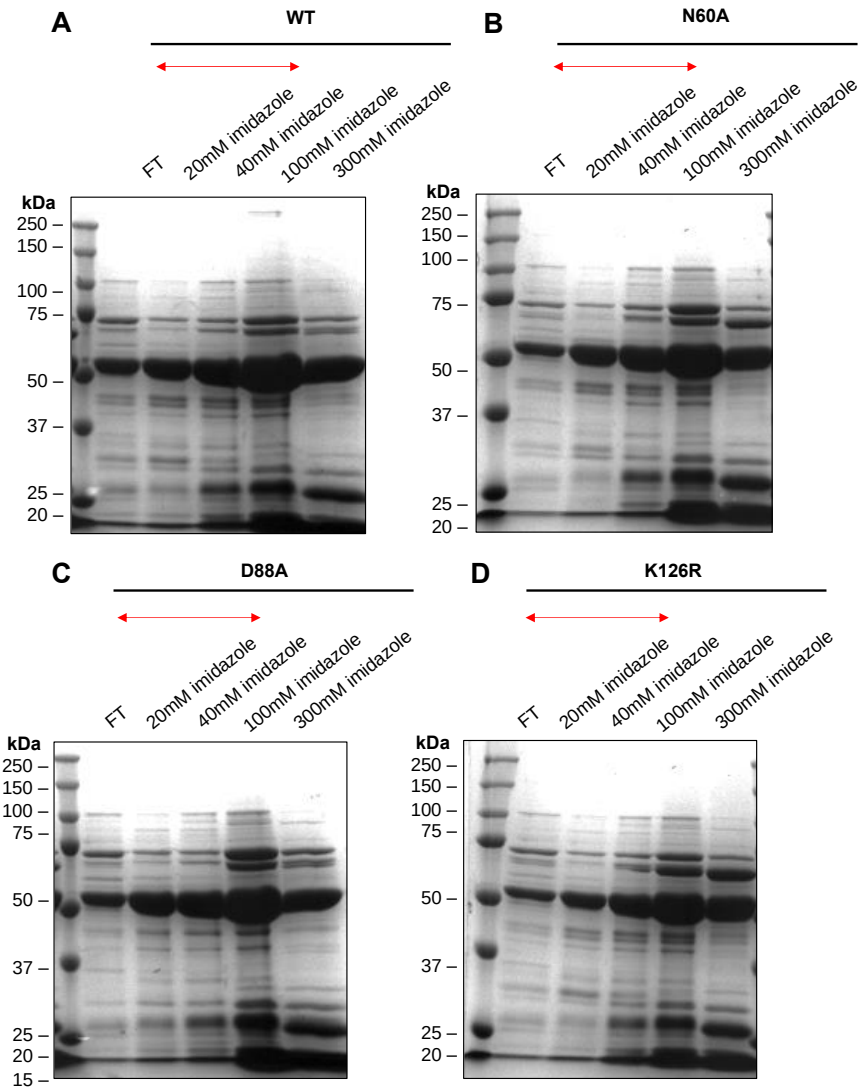


Figure 4.33: **2 L expression and purification step of wild type (WT) MORC4$^{29\text{-}480}$ and mutants following Ni-IMAC TEV rebind.** Each MORC4 protein was expressed in 2 L TB and purified using a Ni-IMAC gravity flow column. Samples were incubated overnight with TEV protease at 4°C. Samples were resolved on a 10% tris-glycine SDS (TGS)-PAGE gels. Each gel shows lysate flow through (FT), and 4 imidazole (20 mM, 40 mM, 100 mM and 300 mM imidazole) washes. (**A)** WT MORC4$^{29\text{-}480}$ **(B)** N60A MORC4$^{29\text{-}480}$ **(C)** D88A MORC4$^{29\text{-}480}$ **(D)** K126R MORC4$^{29\text{-}480}$. Red arrows indicated protein fractions which were pooled.

4.4.2.2.1.2 Cation exchange chromatography (IEX) of MORC4$^{29-480}$ WT and mutants

Following the TEV rebind Ni-IMAC stage of the purification of MORC4$^{29-480}$ WT and mutants, proteins were applied to a 1 ml HiTrap SP-IEX column on an ÄKTA pure purification system for ion exchange (IEX) chromatography, which purified the proteins according to charge (similar to Chapter 3, section 3.3.2.3.1.2). All of the IEX chromatographs for MORC4$^{29-480}$ WT and mutant proteins appeared to elute at a similar volume (~30 mL). This suggests that both the WT and mutants all have similar characteristics and that the mutations have not had an effect on the structure of MORC4$^{29-480}$ too much (Figure 4.34a, b, c and d). However, the chromatogram and the gel generated from IEX suggest that MORC4$^{29-480}$ D88A has a higher yield than the WT (Figure 4.34a and c). There were fewer smaller contaminants present for D88A. However, this could have been due to variation between purifications (Figure 4.35c). The N60A mutant also had a good protein yield at approximately 110 mAU and also had fewer contaminants than the WT (Figure 4.34b). K126R was the MORC4$^{29-480}$ mutant with the lowest yield of protein at approximately 65 mAU, half the yield of the WT (Figure 4.34d).

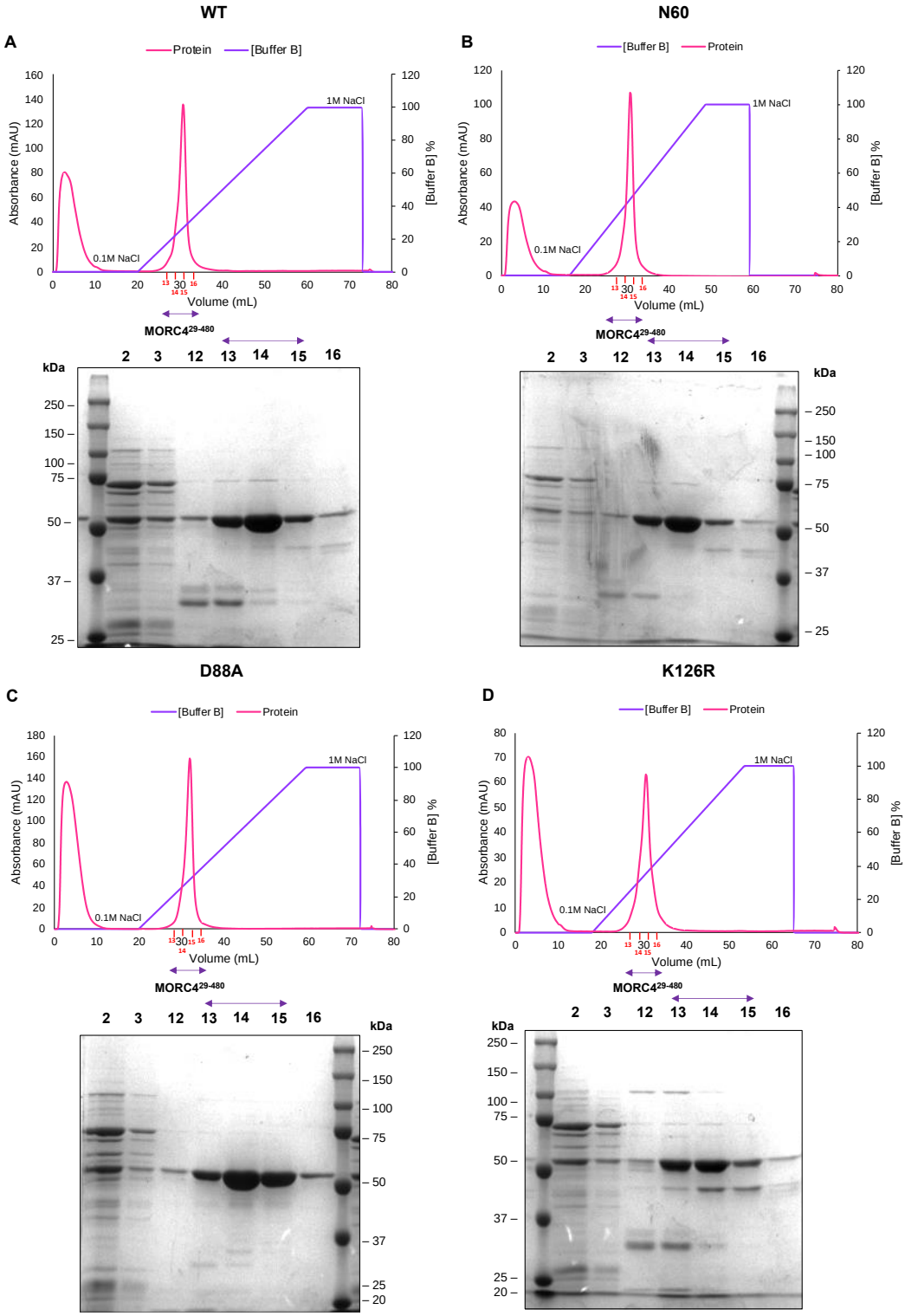


Figure 4.34: **Purification of wild type (WT) MORC4$^{29-480}$ and mutants, using SP-IEX.** Following SP-IEX, selected samples were resolved on a 10% tris-glycine SDS(TGS)-PAGE gel. All proteins were purified using a 1 ml HiTrap SP-IEX column using an ÄKTA pure. All gels are shown with the corresponding chromatograph for each sample. **(A)** WT **(B)** N60A **(C)** D88A (**D)** K126R. For each protein, samples 13, 14 and 15 were pooled for further purification using SEC. The purple arrows indicate protein fractions which were pooled.

4.4.2.2.1.3 Size exclusion chromatography (SEC) of MORC4$^{29-480}$ WT and mutants

Following IEX chromatography and pooling of samples 13-15 for MORC4$^{29-480}$ WT and mutants, proteins were applied to a size exclusion chromatography (SEC) column to remove aggregates and to exchange into a common buffer. All of the proteins purified had similar chromatogram profiles with an initial larger peak, which was the target protein, an intermediate contaminant and finally a small medium contaminant (Figure 4.35a, b, c and d). All proteins were eluted from the SEC column at a volume of approximately 84 mL (Figure 4.35a, b, c and d). This suggests that both the MORC4$^{29-480}$ WT and mutants were the same molecular weight at 50 kDa (Chapter 3, section 3.3.2.3.1.3) and were monomeric as there was no ligand (AMP-PNP) present. WT, N60A and D88A mutants were all of equivalent purity (Figure 4.35a, b and c). However, MORC4$^{29-480}$ WT had the highest yield of protein with 1.6 mg compared to N60A and D88A mutants, which had 0.8 mg and 1.2 mg, respectively, following 2 L *E. coli* culture for protein expression. In contrast, the K126R mutant appeared to have a lower target protein yield of 0.6 mg and a number of smaller contaminants (Figure 4.35d). Following purification of MORC4$^{29-480}$ WT and mutants, samples were pooled, concentrated and stored at -80°C.

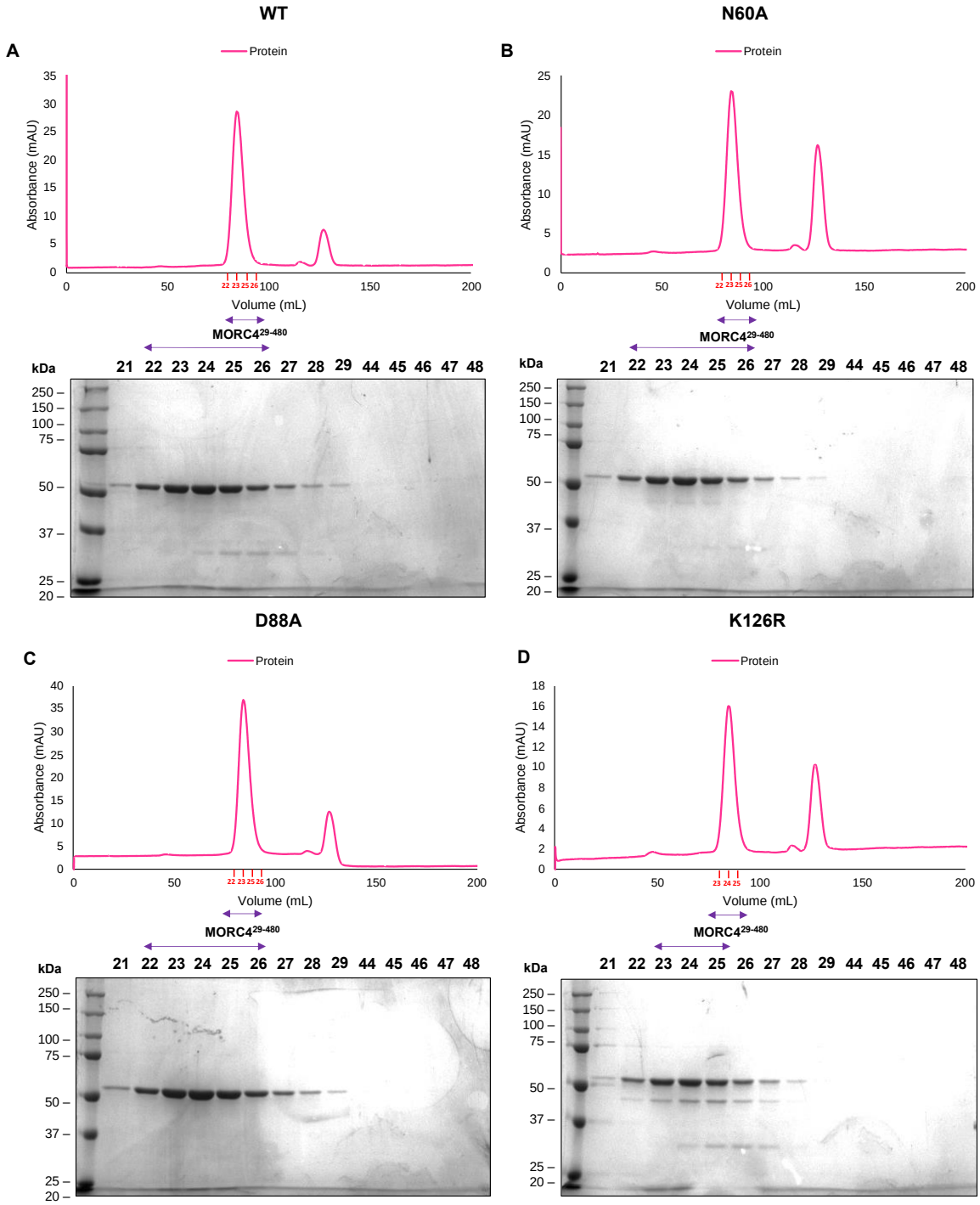


Figure 4.35: **Purification of wild type (WT) MORC4$^{29-480}$ and mutants, using SEC.** Following SEC, selected samples were resolved on a 10% tris-glycine SDS (TGS)-PAGE gel. All proteins were purified using a HiLoad 16/600 Superdex 200 pg column using an ÄKTA pure. All gels are shown with the corresponding chromatograph for each sample. **(A)** WT **(B)** N60A **(C)** D88A **(D)** K126R. For each protein, samples 22-26 (except for K126R, where samples 23-25 were pooled) were pooled, concentrated and flash-frozen in LN_2 and stored at -80° prior to use for biochemical assays. Purple arrows indicate protein fractions which were pooled.

4.4.2.2.2 Purification of Zf-CW MORC4 wild type and mutants

Purification of a Zf-CW MORC4 only fragment WT and mutants were required to investigate the C-terminal end of MORC4 without the ATPase domain. Only the W435K mutant could be made for this truncated MORC4 fragment as the other mutants were in the ATPase region.

4.4.2.2.2.1 Ni-IMAC and TEV rebind of MORC4$^{420-757}$ and W435K mutant

Zf-CW MORC4$^{420-757}$ WT and the W435K mutant proteins were expressed in 4 L of TB. Following sonication and clarification of cell lysate, proteins were purified using an initial Ni-IMAC step. The gels show that the W435K mutant protein was expressed at the same size (50 kDa) and at a similar level to the WT (Figure 4.36a and b). For both the WT and the mutant, samples E300 1-5 were pooled and incubated overnight at 4°C with TEV protease. Following incubation of proteins with TEV protease and an additional Ni-IMAC, it was apparent that both the WT and W435K mutant proteins were successfully cleaved by TEV protease as they reduced from 50 kDa to 38 kDa (Figure 4.36c and d). For each sample, the flow through (FT) and 20 mM imidazole washes were pooled for further purification steps.

4.4.2.2.2.2 Anion exchange chromatography (IEX) of MORC4$^{420-757}$ WT and mutant

Following the TEV rebind Ni-IMAC stage of the purification of MORC4$^{420-757}$ WT and W435K mutant, proteins were applied to a 1 ml HiTrap Q-IEX column on an ÄKTA pure purification system for ion exchange (IEX) chromatography. Both the MORC4$^{420-757}$ WT and W435K mutant chromatograms showed that the proteins were eluted at a volume of approximately 28 mL (Figure 4.37a and b). The absorbance value was also approximately 1000 mAU for both proteins, suggesting that the WT and W435K mutant produce a similar yield of protein. However, the W435K flowthrough peak at approximately 3 mL had a higher absorbance reading than the WT (Figure 4.37a and b). This suggests that the W435K mutant has a higher number of molecules which were not able to bind to the Q-IEX column. Following

IEX, samples 11-13 fractions were pooled for both the WT and W435K mutant for subsequent purification procedures.

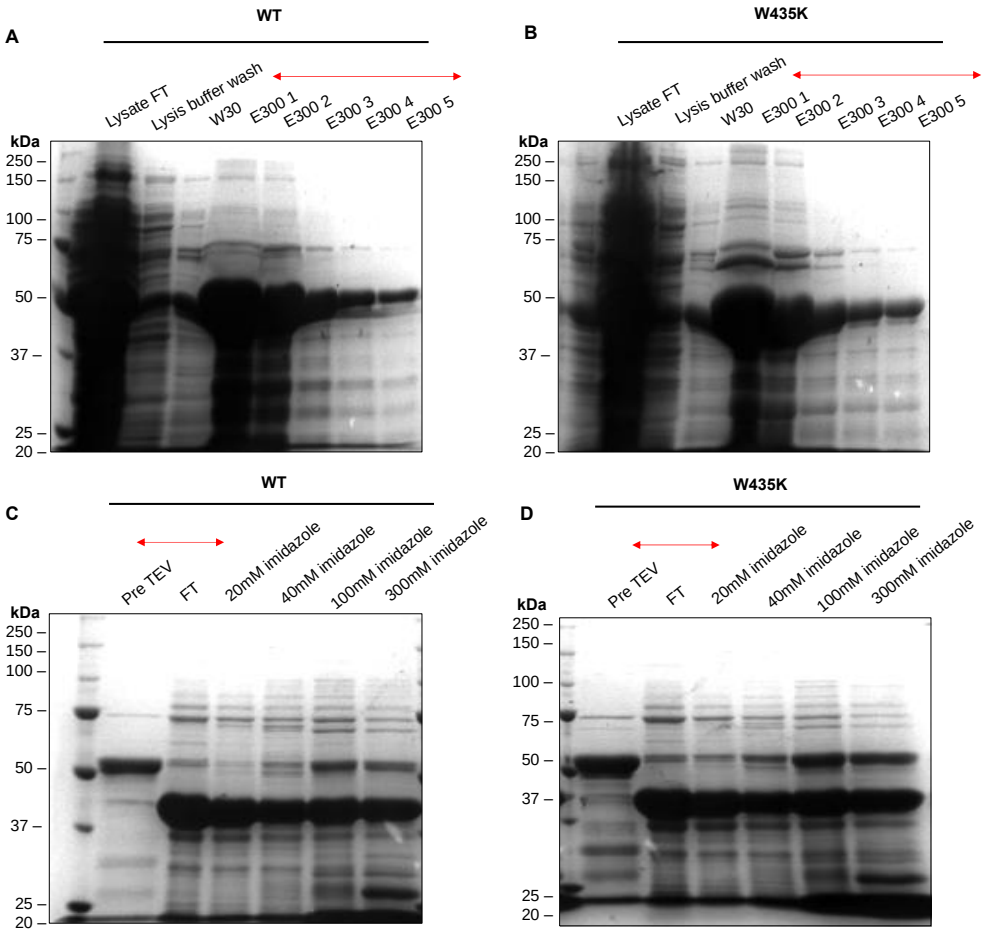


Figure 4.36: **4 L expression and purification of wild type Zf-CW (WT) MORC4$^{420\text{-}757}$ and the W435K mutant following Ni-IMAC and TEV rebind.** Each MORC4 protein was expressed in 4 L of TB and purified using a Ni-IMAC gravity flow column. Samples were resolved on a 10% tris-glycine SDS(TGS)-PAGE gel. Samples were then incubated overnight with TEV protease at 4°C and resolved on a 10% tris-glycine SDS(TGS)-PAGE gel. (**A)** Ni-IMAC MORC4$^{420\text{-}757}$ WT **(B)** Ni-IMAC MORC4$^{420\text{-}757}$ W435K **(C)** TEV rebind Ni-IMAC MORC4$^{420\text{-}757}$ WT and (**D)** TEV rebind Ni-IMAC MORC4$^{420\text{-}757}$ W435K. Red arrows indicate protein fractions which were pooled.

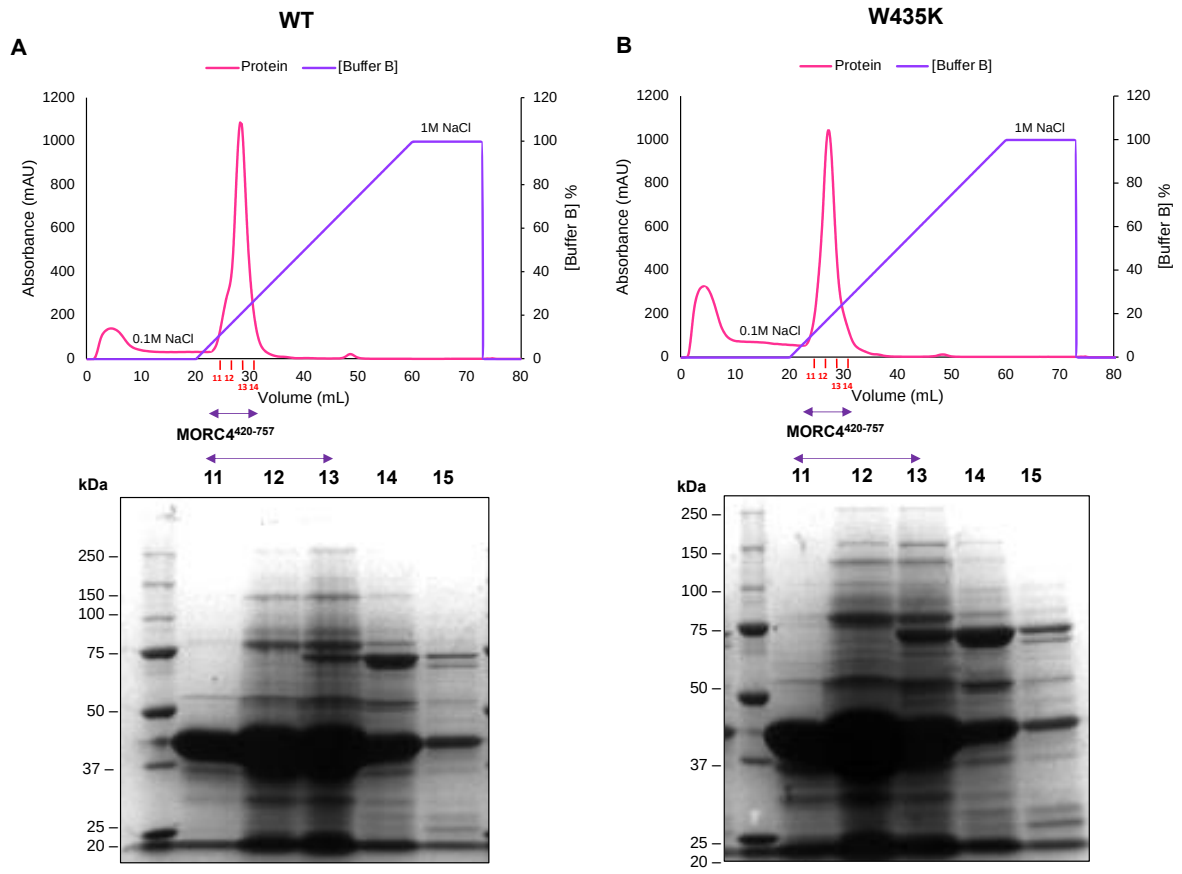


Figure 4.37: **Purification of Zf-CW wild type (WT) MORC4$^{420-757}$ and W435K mutant, using Q-IEX.** Following Q-IEX, selected samples were resolved on a 10% tris-glycine SDS(TGS)-PAGE gel. All proteins were purified using a 1 ml HiTrap SP-IEX column using an ÄKTA pure. Both gels are shown with the corresponding chromatograph for each sample. **(A)** WT **(B)** W435K. For each protein, samples 11-13 were pooled for further purification using SEC. Purple arrows indicate protein fractions which were pooled.

4.4.2.2.2.3 Size exclusion chromatography (SEC) of MORC4$^{420-757}$ WT and mutant

Following IEX chromatography and pooling of samples 11-13 for MORC4$^{420-757}$ WT and W435K mutant, the proteins were applied to a HiLoad 16/600 Superdex 200 pg column on an ÄKTA pure purification system for the final stage of the purification procedure. From analysis of the chromatographs, it was evident that the W435K mutant produced a higher yield of protein as the protein peak was approximately 170 mAU compared to the WT, which had an absorbance reading of approximately 150 mAU (Figure 4.38a and b). The W435K mutant also appeared to have fewer contaminants co-purifying with the mutant as the chromatography showed less peaks compared to the WT (Figure 4.38a and b). However, the gels indicate that

both the WT and W435K were of similar purity and confirm that there was a higher yield of the W435K mutant (Figure 4.38a and b).

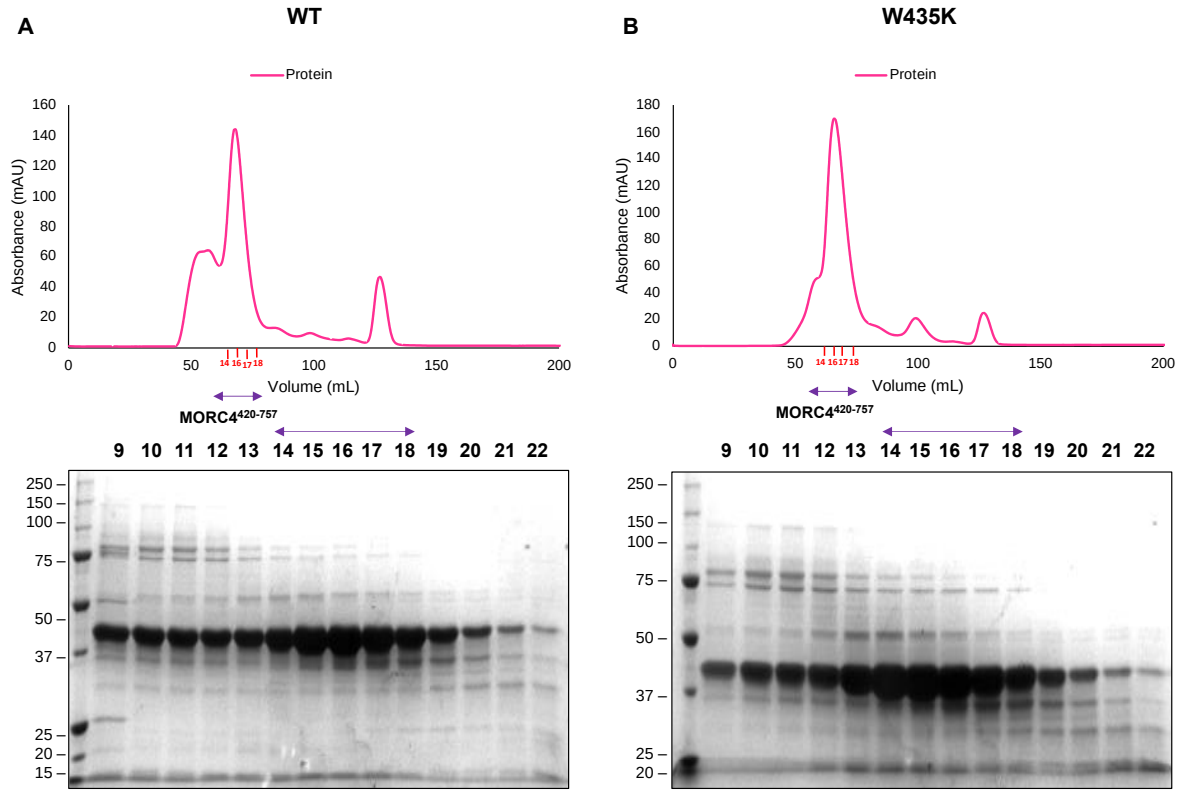
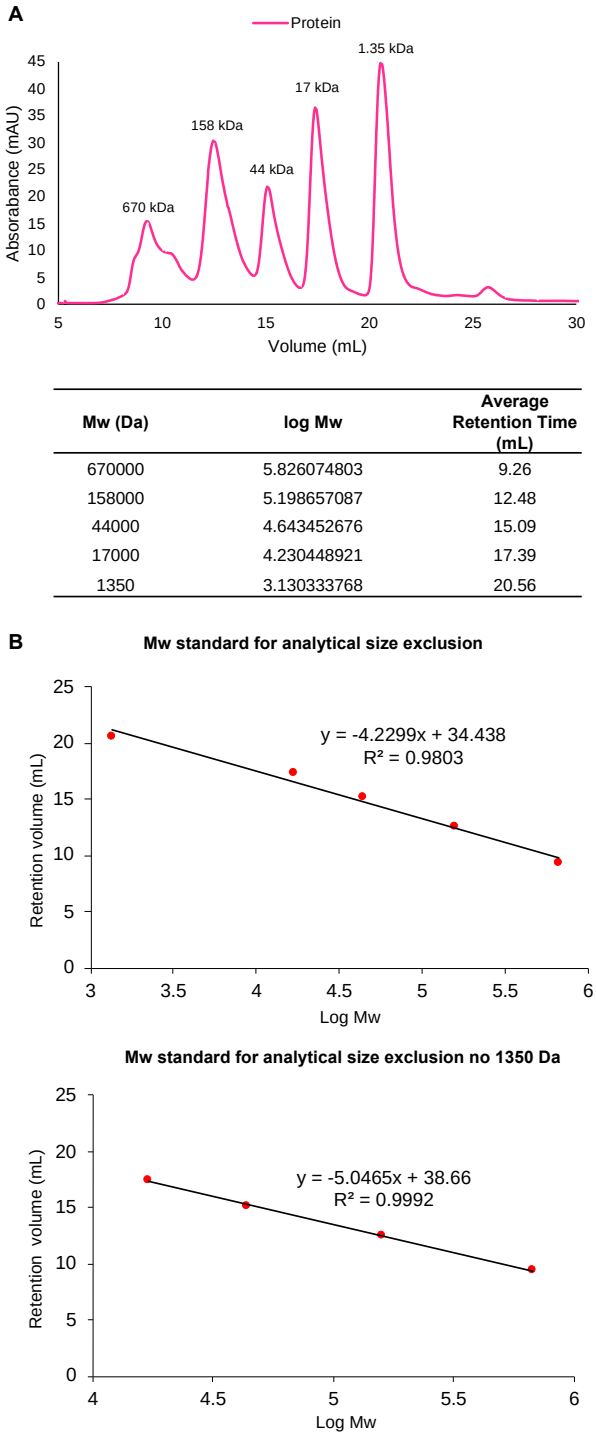


Figure 4.38: **Purification of Zf-CW wild type (WT) MORC4$^{420-757}$ and W435K mutant, using SEC.** Following SEC, selected samples were resolved on a 10% tris-glycine SDS (TGS)-PAGE gel. All proteins were purified using a HiLoad 16/600 Superdex 200 pg column using an ÄKTA pure. Both gels are shown with the corresponding chromatograph for each sample. **(A)** WT **(B)** W435K. For each protein, samples 14-18 were pooled, concentrated, flash-frozen in LN_2 and stored at -80°C prior to use for biochemical assays. Purple arrows indication protein fractions which were pooled.

4.4.3 Analytical size exclusion chromatography (SEC)

Analytical size exclusion chromatography (SEC) was a technique used to investigate both the ATPase-Zf-CW and Zf-CW domains of MORC4 as a tangential method to assess oligomerisation by cross-linking, as SEC-multiangle laser light scattering (MALLS) was beneficial in oligomeric analysis of MORC2 (Douse *et al.*, 2018). SEC was utilised to investigate the ATPase-Zf-CW tandem domains in relation to dimerisation and to link to chemical crosslinking studies. In addition, SEC was also used to investigate the C-terminal coiled coils MORC4 fragments and how it can influence homodimerisation, as active MORC3 has previously been shown to bind to H3K4me3 peptide, allowing homodimerisation and subsequently ATP hydrolysis (Zhang *et al.*, 2019). In addition, a recent study identified that the MORC2 C-terminal coiled-coil domain is responsible for homodimerisation (Xie *et al.*, 2019). Overall, SEC will help to assess if MORC4 acts in a similar manner to MORC2 and MORC3.

Prior to injecting the target proteins into the analytical SEC column, a sample containing known standards was injected into the ÄKTA pure system and the retention times and known molecular weights (1.3 kDa – 670 kDa) were utilised (Figure 4.39a) to create 2 standard curves, with and without the smallest known molecular weight standard, 1.3 kDa (Figure 4.39b). The standard curve equation of a straight line was then used for calculating the molecular weight of target proteins and subsequently, a predicted oligomeric state.



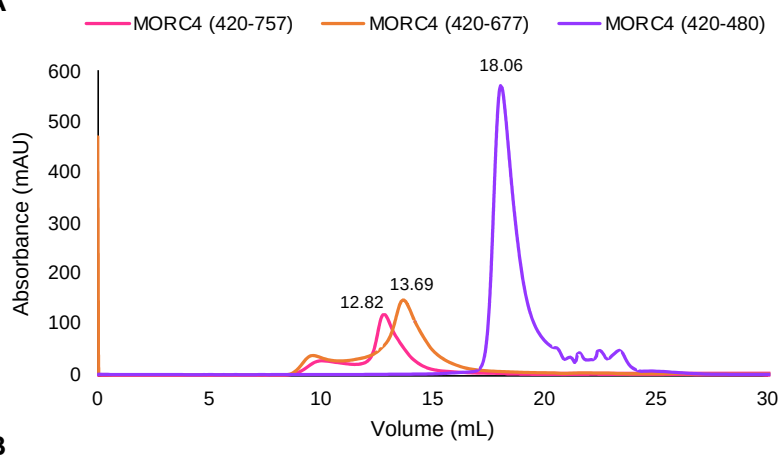
Mw (Da)	log Mw	Average Retention Time (mL)
670000	5.826074803	9.26
158000	5.198657087	12.48
44000	4.643452676	15.09
17000	4.230448921	17.39
1350	3.130333768	20.56

Figure 4.39: **Standards required for analytical size exclusion chromatography (SEC).** A control marker size exclusion standard of known molecular weight was diluted 1 in 5 in low salt SEC buffer. samples were injected into the HiLoad 100/300 Superdex 200 pg column for analytical size exclusion chromatography and protein was determined using the ultraviolet (UV) trace at 280 nm on an ÄKTA pure system. **(A)** SEC chromatograph of known standards with corresponding table showing the molecular weight of each standard and average retention time. **(B)** Standard curves (with and without the smallest standard sample, 1350 Da) required to calculate molecular weight of target proteins with the equation of a straight line shown above the curve.

4.4.3.1 Oligomerisation analysis of Zf-CW MORC4 proteins

Following analysis of known standard proteins, analytical SEC was used to assess the oligomeric status of C-terminal Zf-CW MORC4 proteins. MORC4$^{420-757}$, MORC4$^{420-677}$, and MORC4$^{420-480}$ C-terminal proteins were loaded onto the HiLoad 100/300 Superdex 200 pg column, which separates proteins according to molecular weight. The results revealed that the largest C-terminal fragment, MORC4$^{420-757}$, was eluted first at 12.82 mL followed by MORC4$^{420-677}$ at 13.69 mL. The smallest C-terminal fragment, MORC4$^{420-480}$, was eluted last at 18.06 mL (Figure 4.40a). The equation of straight line from known protein standards (Figure 4.40b) was used to calculate the observed molecular weight of each C-terminal MORC4 protein. The observed mass of both MORC4$^{420-757}$ and MORC4$^{420-677}$ was three times the mass of the actual mass of both C-terminal proteins (Figure 4.40b). This suggests that both MORC4$^{420-757}$ and MORC4$^{420-677}$ have a trimeric oligomeric state. The MORC4$^{420-480}$ observed mass was calculated at 7.5 kDa, the same as the actual mass of MORC4$^{420-480}$ (Figure 4.40b), suggesting that this Zf-CW only MORC4$^{420-480}$ protein is monomeric. These data collectively suggest that the addition of appended C-terminal coiled-coils causes trimerisation of the Zf-CW region, although this oligomerisation could be a consequence of shape effects of the Zf-CW domain. Furthermore, MORC2 acts in a similar manner as recent evidence suggests that MORC2 dimerises through the C-terminal coiled-coils (Xie *et al.*, 2019), supporting this data of oligomerisation of MORC4$^{420-480}$ Zf-CW only domain protein.

A



B

Target	Retention volume (mL)	Log Mw (Da)	Observed Mw (kDa)	Actual Mw (kDa)	Fraction
MORC4$^{420\text{-}757}$	12.82	5.1	129	38	3.3
MORC4$^{420\text{-}677}$	13.69	4.9	80	30	2.7
MORC4$^{420\text{-}480}$	18.06	3.9	7.5	7.5	1.0

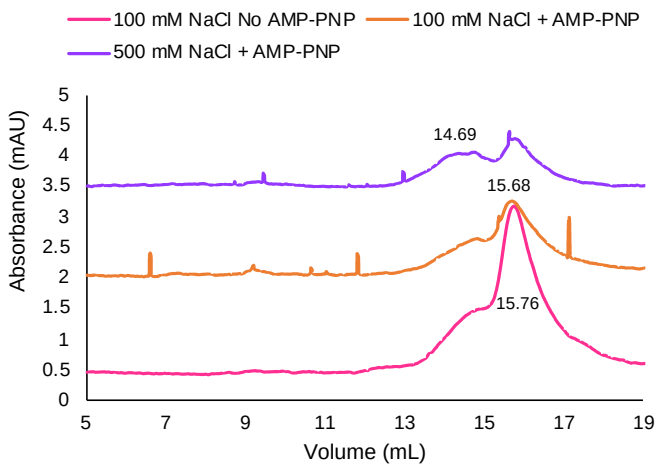
Figure 4.40: **Analytical size exclusion chromatography (SEC) of C-terminal fragments of MORC4.** A HiLoad 100/300 Superdex 200 pg column (GE Healthcare life sciences) was washed with ddH_2O at 1 mL/min and equilibrated in low salt SEC buffer (20 mM HEPES, 100 mM NaCl, 5 % glycerol and 1 mM DTT) on the ÄKTA Pure system. All pure protein samples were injected into the HiLoad 100/300 Superdex 200 pg column for analytical size exclusion chromatography (SEC) and protein was determined using the ultraviolet (UV) trace at 280 nm. (**A)** Chromatograph showing the retention volume of each C-terminal MORC4 protein. (**B)** Table showing each C-terminal MORC4 protein with retention volume, log molecular weight, actual molecular weight, observed molecular weight and oligomeric state.

4.4.3.2 Oligomeric status of ATPase-Zf-CW MORC4$^{29\text{-}480}$

In an attempt to investigate the oligomeric state of the ATPase-Zf-CW MORC4$^{29\text{-}480}$ protein, multiple samples of MORC4$^{29\text{-}480}$ were analysed by analytical SEC. Two samples were diluted in SEC buffer without NaCl, to reduce the NaCl concentration of the protein as high salt can abrogate complex formation. Protein was then incubated with 1.0 mM AMP-PNP/Mg^{2+} and utilised for analytical SEC. Only the protein containing high salt (500 mM NaCl) and AMP-PNP was found to be the correct size (Figure 4.41a and b) of approximately 50 kDa, which was the monomeric protein size. However, this could suggest that MORC4$^{29\text{-}480}$ does not dimerise in the presence of high salt concentrations and AMP-PNP. On the other hand, both protein samples which had been diluted to achieve low salt (100 mM NaCl), with one protein sample

incubated in the presence of AMP-PNP, were both found to be 26/27 kDa (Figure 4.41a and b). This suggests that these protein samples were half the expected size of monomer MORC4$^{29-480}$, indicating degradation of the protein samples. Overall, this experiment did not reveal that AMP-PNP induces dimerisation of ATPase-Zf-CW MORC4$^{29-480}$ protein, suggesting alternative methods may be required for this analysis.

A



B

Target	Retention volume (mL)	Log Mw (Da)	Observed Mw (kDa)	Actual Mw (kDa)	Fraction
MORC4$^{29-480}$ 500 mM NaCl + AMP-PNP	14.69	4.6	47	51.6	0.9
MORC4$^{29-480}$ 100 mM NaCl +AMP-PNP	15.68	4.4	27	51.6	0.53
MORC4$^{29-480}$ 100 mM NaCl No AMP-PNP	15.76	4.4	26	51.6	0.50

Figure 4.41: **Analytical size exclusion chromatography (SEC) of the ATPase-Zf-CW MORC4$^{29-480}$.** A HiLoad 100/300 Superdex 200 pg column (GE Healthcare life sciences) was washed with ddH_2O at 1 mL/min and equilibrated in low salt SEC buffer (20 mM HEPES, 100 mM NaCl, 5 % glycerol and 1 mM DTT) on the ÄKTA pure system. Protein samples were diluted 1 in 5 with low salt SEC buffer and incubated with 1.0 mM AMP-PNP/Mg^{2+} for 1 hour at 4°C. All protein samples were injected into the HiLoad 100/300 Superdex 200 pg column for analytical size exclusion chromatography (SEC) and protein was determined using the ultraviolet (UV) trace at 280 nm. **(A)** Chromatograph showing the retention volume of each MORC4$^{29-480}$ protein sample. **(B)** Table showing each MORC4$^{29-480}$ protein sample with retention volume, log molecular weight, actual molecular weight, observed molecular weight and oligomeric state.

4.4.4 Size exclusion chromatography-Multiangle laser light scattering (SEC-MALLS)

Size exclusion chromatography-Multiangle laser light scattering (SEC-MALLS) is used to characterise proteins by separation of molecular weight using SEC and light scattering to detect the absolute molecular weight using MALLS (Some *et al.*, 2019). MORC4$^{29\text{-}480}$ ATPase-Zf-CW protein was used for SEC-MALLS analysis in a further attempt to identify the oligomeric status and molecular weight in the presence and absence of the ligand, AMP-PNP. Following SEC-MALLS, there was no detection of MORC4 which had been diluted in low salt buffer. Therefore, undiluted protein at 8 mg/ml was used and a UV peak was present on the chromatograph in red (Figure 4.42). However, the ASTRA 6 software analysis predicted a molecular weight of approximately between 600-900 kDa with a 5 %± error in calculation, suggesting MORC4 protein aggregation. This is an unusual finding as the control protein alcohol dehydrogenase (ADH), was 150 kDa with a retention time of 8.5 mL (M. Thomsen, University of Leeds, *pers. comm*.). MORC4$^{29\text{-}480}$ had a retention time of 11 mL (600-900 kDa) (Figure 4.42), suggesting that the observed molecular weight was smaller than the calculated molecular weight. In addition to this, MORC4$^{29\text{-}480}$ in presence of the ligand, AMP-PNP, was also used for SEC-MALLS analysis. However, the analysis of MORC4$^{29\text{-}480}$ showed that there was no peak shift in the presence of AMP-PNP, suggesting that perhaps an alternative method is required to detect MORC4$^{29\text{-}480}$ dimerisation in the presence of AMP-PNP.

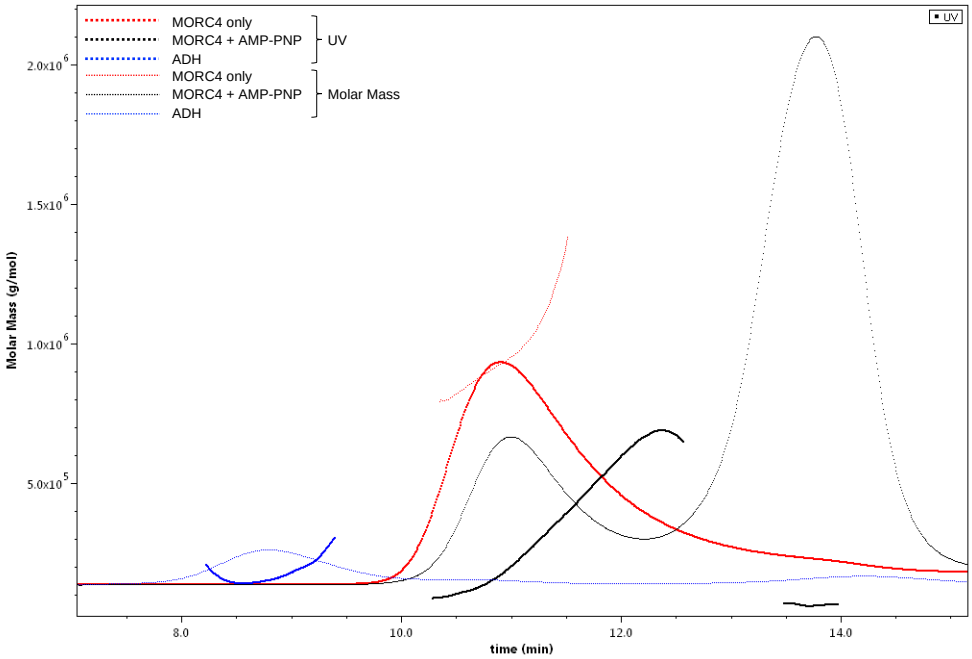


Figure 4.42: **Analytical size exclusion chromatography Multiangle laser light scattering (SEC-MALLS) of the ATPase-Zf-CW MORC4$^{29\text{-}480}$.** A Superdex 200 5/150 increase column was washed with ddH_2O at 0.2 mL/min and was equilibrated in low salt SEC-MALLS buffer (20 mM HEPES, 150 mM NaCl and 5 % glycerol). Some protein samples were diluted 1 in 3 with dilution buffer (20 mM HEPES pH 7.5 and 5 % glycerol, 1 mM DTT) and incubated with 1.0 mM AMP-PNP/Mg^{2+} for 1 hour at 4°C. 25 µL was injected for SEC. The MALLS detector DAWN 8+ and a RF-20A spectrofluorometric detector were used for detection of proteins by UV. ASTRA 6 software was used to calculate the molar mass of protein. Performed with the assistance of Dr. Maren Thomsen, The University of Leeds. Red is the sample containing MORC4 only, black is MORC4 with the addition of AMP-PNP and blue is the control protein, alcohol dehydrogenase (ADH).

4.4.5 Chemical crosslinking

Chemical crosslinking is often used to investigate protein interactions such as MORC3 and the interaction with the ligand, AMP-PNP, which has been shown to lead to dimerisation (Li *et al.*, 2016). Chemical crosslinkers are able to join two molecules (e.g. protein and AMP-PNP ligand) and by a covalent bond (Arora *et al.*, 2017) that are in close proximity. The length of the crosslinker determines how close atoms are to link the type of functional group (Rappsilber, 2011). In addition, the concentration of crosslinker used is also important as if the concentration is too high multiple non-specific atoms may be crosslinked. Therefore, both short (glutaraldehyde) and long crosslinkers (bis(sulfosuccinimidyl)suberate (BS^3)) were utilised for testing. MORC4$^{29\text{-}480}$ WT and mutant proteins (Section 4.4.2) were used throughout

crosslinking experiments, as this fragment of MORC4 is equivalent the fragments of MORC2 and MORC3 observed to dimerise, in addition to containing the ATPase domain.

4.4.5.1 Optimisation of chemical crosslinking

4.4.5.1.1 Determining a suitable concentration of protein

Prior to chemical crosslinking experiments, it was important to determine a suitable concentration of MORC4$^{29\text{-}480}$ to be loaded onto an SDS-PAGE gel following chemical crosslinking as these assays results are visualised by SDS-PAGE. 1 µg, 0.5 µg, 0.2 µg and 0.1 µg of MORC4$^{29\text{-}480}$ were resolved a gel and the result suggested that a least 1 µg of protein is required for protein bands to be visible (Appendix 8.19).

4.4.5.1.2 Determination of a suitable chemical crosslinker

Varying concentrations of glutaraldehyde and BS^3 crosslinkers in the presence and absence of the-AMP-PNP ligand were used to investigate the oligomeric state of MORC4$^{29\text{-}480}$. AMP-PNP was utilised throughout these experiments as it is a non-hydrolysable ATP analogue. If ATP was used MORC4 may have been able to hydrolyse it, reversing any dimerisation, as expected with MORC2 and MORC3. In addition, AMP-PNP was also used in biochemical characterisation of MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016).

High BS^3 concentrations resulted in high molecular weight smearing of MORC4$^{29\text{-}480}$ WT from approximately 100-250 kDa (Figure 4.43a), suggesting that high concentrations of BS^3 cause non-specific crosslinking. However, protein samples treated with glutaraldehyde had more smearing than BS^3 (Figure 4.43b) and although the MORC4 protein used for the glutaraldehyde was slightly less pure, this could suggest that glutaraldehyde is responsible for more non-specific crosslinking of monomers in solution than BS^3. Overall, this data suggests that BS^3 is more suitable for chemical crosslinking experiments as it results in more defined complexes rather than non-specific heterogeneous (smeared) proteins as observed with glutaraldehyde.

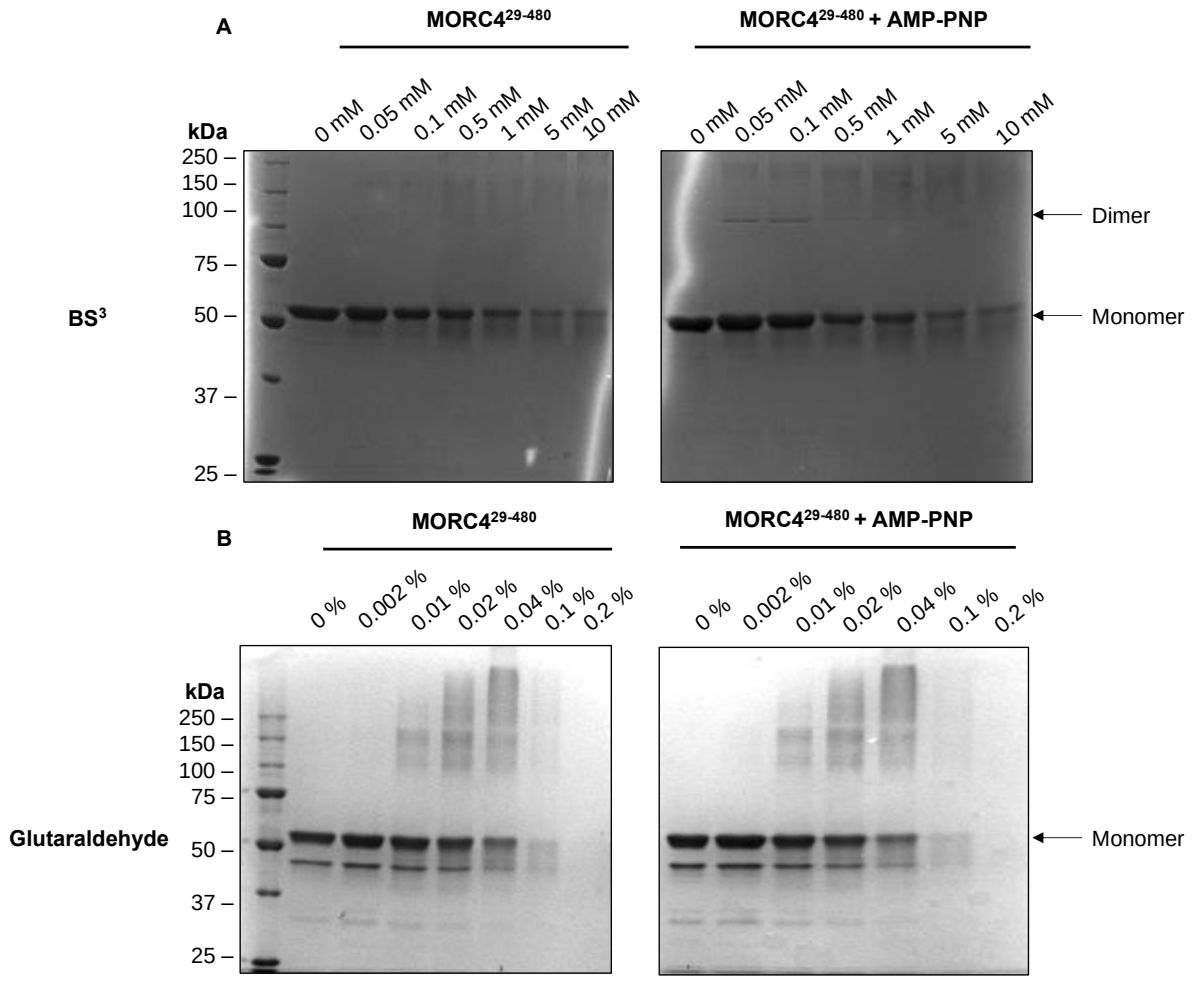


Figure 4.43: **Chemical crosslinking of MORC4$^{29\text{-}480}$ in the presence and absence of the ligand, AMP-PNP.** 2 mg/ml of protein samples, which had been chemically crosslinked with either **(A)** BS3 or **(B)** glutaraldehyde in the presence or absence of AMP-PNP, were loaded onto a 10% tris-glycine SDS (TGS)-PAGE gel.

4.4.5.2 MORC4$^{29\text{-}480}$ WT dimerises in the presence of AMP-PNP

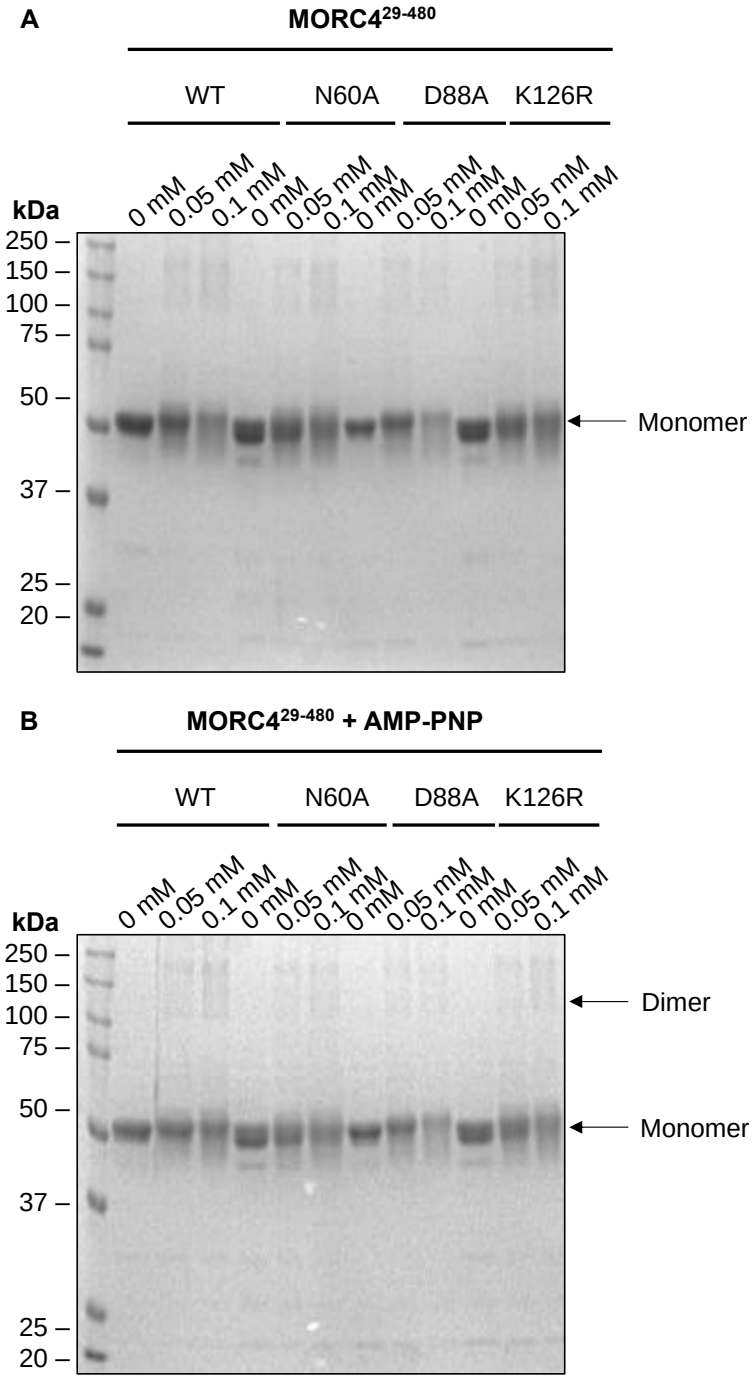
Following chemical crosslinking of MORC4$^{29\text{-}480}$ in the presence and absence of AMP-PNP using BS^3 and glutaraldehyde, it was evident that only a fraction of MORC4$^{29\text{-}480}$ exists as a defined dimer in the presence of AMP-PNP, when BS^3 was used at a concentration of 0.5-0.05 mM (Figure 4.43a) (MORC4$^{29\text{-}480}$ is a dimer at ~100 kDa). This could mean that AMP-PNP/ATP binding to MORC4 has a low equilibrium dissociation constant (K_D), which could be tested in future using isothermal titration calorimetry (ITC) as it can measure the affinity of a binding interaction (Duff Jr *et al.*, 2011). In addition, as dimerisation only took place in the

presence of AMP-PNP; there was no dimerisation in the absence of AMP-PNP. This suggests that, like MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016), MORC4 also requires the binding of ATP to dimerise. Furthermore, when using high concentrations of BS^3 non-specific heterogeneous proteins (smears) were visible (Figure 4.43a), hence, no dimerisation. This could also account for the lack of dimerisation using glutaraldehyde (Figure 4.43b). In addition, there was a band visible at approximately 150 kDa following the glutaraldehyde experiment (Figure 4.43b). However, it was amongst a broader smear, suggesting it was non-specific.

4.4.5.3 MORC4$^{29\text{-}480}$ mutants do not dimerise in the presence of AMP-PNP

Following chemical crosslinking of MORC4$^{29\text{-}480}$ WT and mutants in the presence of AMP-PNP using BS^3, it was apparent that the WT MORC4$^{29\text{-}480}$ dimerised, unlike the N60A, D88A and K126R mutants, which have been previously shown to abrogate ATP binding and hydrolysis from structural modelling in this study and equivalent mutations in MORC2 (Douse *et al.*, 2018; D Q Li *et al.*, 2012) and MORC3 (H Andrews *et al.*, 2016) (Figure 4.44a, b and c). As with MORC4$^{29\text{-}480}$ WT and mutants, in the absence of AMP-PNP, there was no apparent dimerisation as MORC4 protein resolved at 50 kDa and there was no protein band at 100 kDa (Figure 4.44b). However, there was some non-specific crosslinking at approximately 200 kDa for samples containing 0.05 mM and 0.1 mM BS^3 (Figure 4.44a). In the presence of AMP-PNP, there was some dimerisation of MORC4$^{29\text{-}480}$ WT containing 0.05 mM and 0.1 mM BS^3 (Figure 4.44b), supporting the original results from Fig. 4.43, that ATP is required for MORC4 dimerisation. However, there was no dimerisation seen for any of the MORC4$^{29\text{-}480}$ mutants in the presence of AMP-PNP (Figure 4.44b), suggesting that these mutants were responsible for abrogation of AMP-PNP binding. Although, there was some faint bands at 100 kDa for N60A/D88A mutants in the presence of AMP-PNP (Figure 4.44b). However, they were visible as smears, unlike the WT, which showed a clearer dimer band, although faint (Figure 4.44).

There was also some evidence that MORC4$^{29-480}$ WT did not dimerise in the presence of ADP (Figure 4.44c), as distinct dimer protein bands were not visible at 100 kDa for WT or mutants. This suggests that ADP (the product of ATP hydrolysis) does not induce dimerisation of MORC4$^{29-480}$, and that dimerisation is specific to AMP-PNP (and by proxy, ATP), as observed for MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016) elsewhere.



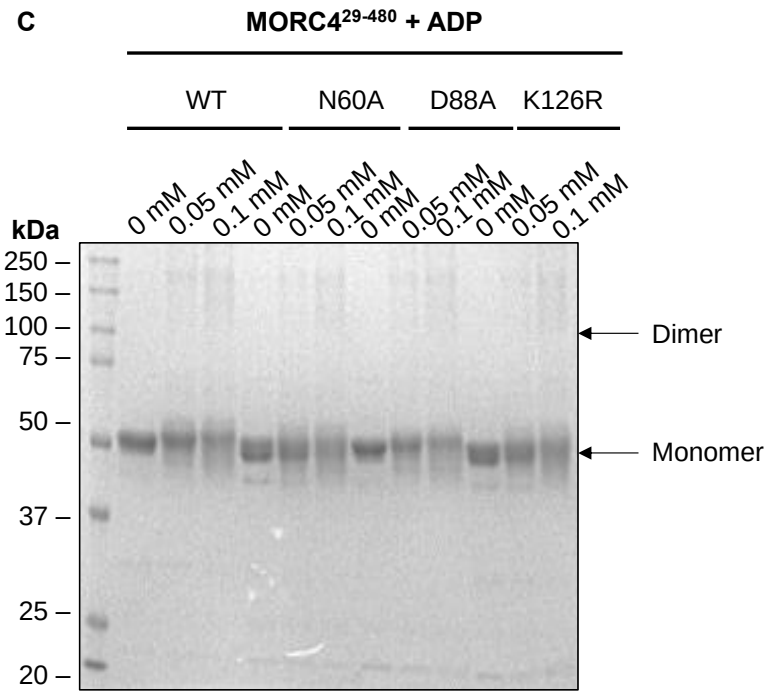


Figure 4.44: **Chemical crosslinking of MORC4$^{29-480}$ WT and mutants in the presence and absence of ligands, AMP-PNP and ADP.** 2 mg/ml of protein samples, which had been chemically crosslinked with BS^3 in the (**A)** absence of ligands (**B)** presence of AMP-PNP or (**C)** presence of ADP. WT, N60A, D88A or K126R mutants were loaded onto a 10% tris-glycine SDS(TGS)-PAGE gel for analysis of MORC4$^{29-480}$ WT and mutants oligomeric state.

4.4.6 Characterisation of MORC4 ATPase activity

4.4.6.1 Establishment of ATPase assays

ATPase assays were important for determination of MORC4 as an active ATPase, similar to MORC2 (D Q Li *et al.*, 2012) and MORC3 (Zhang *et al.*, 2019). Furthermore, ATPase assays measure the release of inorganic phosphate (Pi) by quantitation using colorimetry (Rule *et al.*, 2016). This assay could help to contribute to the characterisation of novel ATPases such as MORC4. Both MORC2 (Douse *et al.*, 2018) and MORC3 (Zhang *et al.*, 2019) have been identified as ATPases from the conserved domain/residues and biochemical assays where both thin layer chromatography (TLC) (D Q Li *et al.*, 2012) and colorimetric ATPase assays (H Andrews *et al.*, 2016) were utilised. This initial work presented in this project was required to develop the ATPase assay with a relevant control.

4.4.6.2 Assessment of ATPase assays using RecA as an ATPase control

Prior to assessing MORC4 ATPase activity in ATPase assays, it was important to establish a reliable positive ATPase control. This was required to determine if the assays were accurate, by investigating the release of inorganic phosphate (Pi) from ATP using a known recombinant ATPase. RecA is a protein found in *Escherichia coli*, which has been identified as a DNA-dependent ATPase (Bell, 2005; Gataulin *et al.*, 2018).

Initially, several phosphate standards were made ranging from 0-50 µM to create a standard curve, in order to calculate released Pi from ATP. The standard curve was subsequently used to calculate the concentration of Pi released following the addition of RecA (Figure 4.45a). Following the standard curve, an ATPase assay was performed using RecA with the addition of either ssDNA or dsDNA. The equation of a straight line was used to calculate the amount of Pi released every 1 minute (Figure 4.45b). The results showed that with an increasing concentration of RecA, the concentration of Pi also increased per minute, specifically, with the sample containing ssDNA. ssDNA was shown to release 0.46 Pi/minute compared to dsDNA, which released 0.17 Pi/minute (Figure 4.45b). This was also evident in the colour seen in the 96-well plate. The samples containing 10000 nM of RecA and ssDNA had the most intense green colour (Figure 4.45c), suggesting that a higher concentration of Pi was released in these samples. Lower concentrations of RecA were tested in other experiments, however, no ATPase activity was identified, suggesting that RecA is a weak ATPase. Some release of Pi was also seen for RecA samples containing dsDNA. No Pi release was seen for samples of RecA containing no DNA. This suggests that ssDNA was required to stimulate RecA-mediated ATP

hydrolysis, and confirming establishment of the assay (Gataulin *et al.*, 2018). However, ATPase assays could not be established for MORC4 due to time constraints.

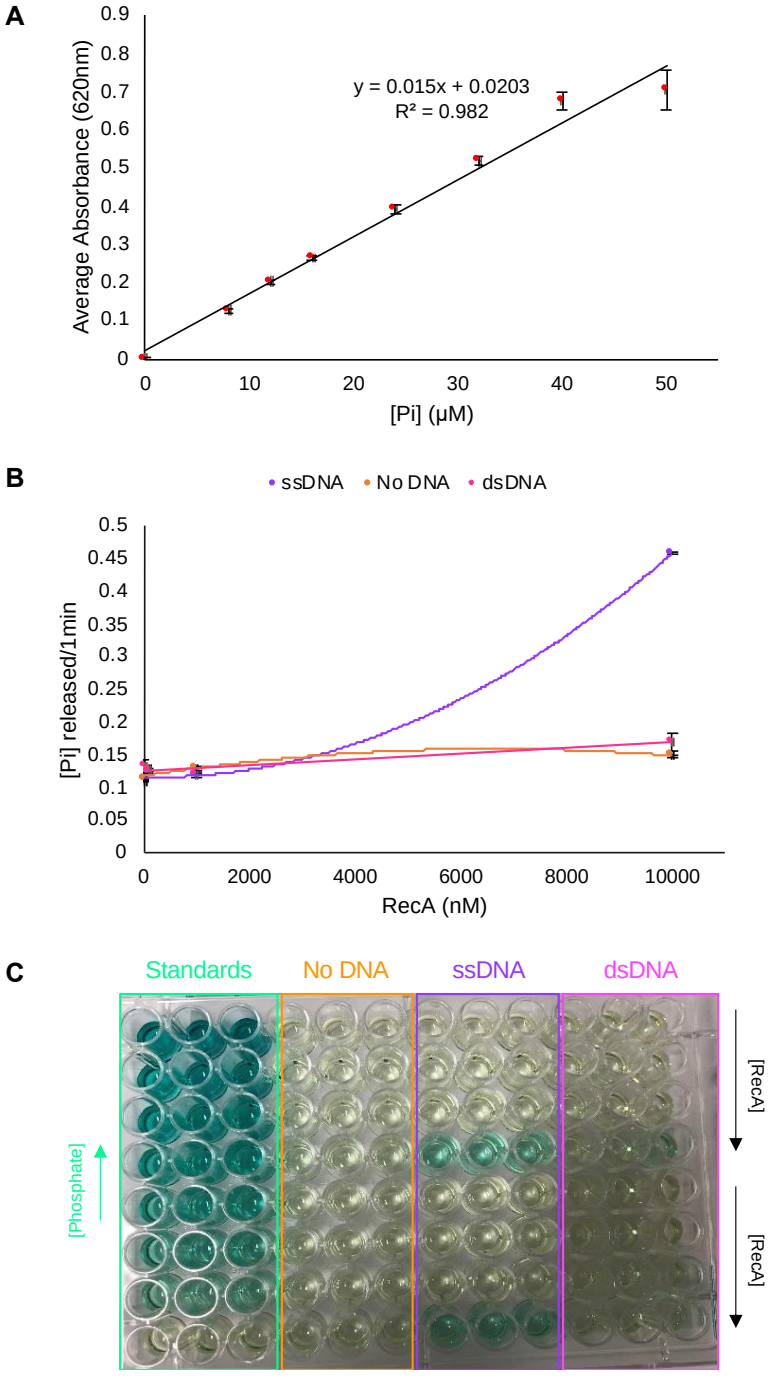


Figure 4.45: **Establishment of ATPase assay.** Phosphate release was measured using a malachite green phosphate assay kit colorimetric test. Phosphate release was measured at 620nm using a plate reader. (**A)** Phosphate standard curve. (**B)** RecA was treated with dsDNA or ssDNA and used to perform the ATPase assay. [Pi] released/min was calculated and plotted. (**C)** 96-well plate from RecA ATPase assay showing the change in colour on release of [Pi]. Each sample was prepared in triplicate.

4.5 Discussion

To aid the determination of MORC4 function, structural and biochemical analysis was performed. Several mutants were identified in previous MORC2 and MORC3 studies, which led to disruption of the ATPase domain and resulting abrogation of AMP-PNP binding in MORC2 (Douse *et al.*, 2018; D Q Li *et al.*, 2012) and the H3K4me3 peptide in MORC3 (H Andrews *et al.*, 2016). Structural analysis using UCSF Chimera confirmed that the mutations present in MORC2 and MORC3 led to the interference of the ligand and the peptide as they led to a loss of vital non-covalent interactions. Sequence analysis was used to identify the MORC2/MORC3 mutations in MORC4 and UCSF Chimera was also used to confirm the structural location of the mutations. Following site directed mutagenesis, the WT and mutant MORC4 proteins were expressed in *E. coli* and purified. Purification revealed that all mutant proteins were cleaved by TEV protease, suggesting that the mutations do not cause any steric hindrance around the tag cleavage site. However, the K126R mutant, which disrupts the ATPase domain, had a lower yield and more smaller contaminants compared to other mutants, suggesting that this mutation disrupts the yield of protein, potentially by less optimal folding. Although, this difference in yield could also be due to variation between protein purification and expression of the mutant/wild type proteins.

Following successful site-directed mutagenesis and protein purification, chemical crosslinking was utilised to investigate the oligomerisation propensity of MORC4. As other MORC proteins dimerise with ATP or its analogue AMP-PNP, oligomeric status analysis was important to assess if MORC4 also functioned in a similar manner. SEC-MALLS and analytical SEC failed to show any evidence of dimerisation in the presence of AMP-PNP. Chemical crosslinking, however, indicated that WT MORC4$^{29\text{-}480}$ dimerises in the presence of AMP-PNP, suggesting that ATP is required for dimerisation, similar to MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016; Zhang *et al.*, 2019). However, this experiment result was weak, but generally

reproducible. This could suggest that perhaps ATP/AMP-PNP binds weakly to MORC4, leading to less dimer formation or that the protein was not fully functional. Chemical crosslinking also revealed that MORC4$^{29\text{-}480}$ mutants, N60A, D88A and K126R, did not dimerise in the presence of AMP-PNP, confirming that the mutations successfully abrogated AMP-PNP binding and subsequently prevented dimerisation of MORC4. The similar chemical ligand, ADP, did not induce dimerisation either, suggesting that dimerisation was specific to AMP-PNP/ATP. In previous studies, the MORC2 N39A mutation (equivalent to N60A in MORC4) was also found to cause inactivity in the ATPase assay, suggesting that the mutation causes MORC2 to be inactive (Douse *et al.*, 2018). The D68A (equivalent to D88A in MORC4) mutation in MORC3 was also found to inhibit ATP binding and ATPase activity (D Q Li *et al.*, 2012).

In addition to chemical crosslinking, alternative biophysical methods (Vedadi *et al.*, 2010) were required in an attempt to identify the oligomeric state of MORC4$^{29\text{-}480}$ and to provide further evidence that MORC4$^{29\text{-}480}$ could dimerise in the presence of AMP-PNP/ Mg^{2+}. Initially, analytical size exclusion chromatography (SEC) was utilised. However, there was no evidence to suggest dimerisation of MORC4$^{29\text{-}480}$ in the presence of AMP-PNP/Mg^{2+}. In addition, size exclusion chromatography-multiangle laser light scattering (SEC-MALLS) was also used to investigate MORC4$^{29\text{-}480}$ oligomeric state. However, this also provided no evidence to suggest that MORC4$^{29\text{-}480}$ dimerises in the presence of AMP-PNP/Mg^{2+}, unlike MORC2. MORC2 SEC-MALLS identified a peak shift with the addition of AMP-PNP/Mg^{2+}, suggesting that MORC2 forms a dimer in the presence of AMP-PNP, which was later confirmed using X-ray crystallography (Douse *et al.*, 2018).

Protein crystallography trials also yielded some evidence supporting MORC4 dimerisation with the addition of the ligand, AMP-PNP. The truncated MORC4 ATPase-Zf-CW fragment, MORC4$^{29\text{-}480}$, did not crystallise in the absence of AMP-PNP/Mg^{2+}. However, MORC4$^{29\text{-}480}$

crystallised in the presence of presence of AMP-PNP/Mg^{2+}, suggesting that $MORC4^{29-480}$ forms a dimer in the presence of AMP-PNP/Mg^{2+}, similar to MORC2 and MORC3. In an attempt to rule out the possibility that the crystals formed were AMP-PNP, some samples were prepared including AMP-PNP/Mg^{2+} in size exclusion buffer only. However, crystals were identified, suggesting that $MORC4^{29-480}$ forms protein crystals in the presence of AMP-PNP/Mg^{2+}. Conversely, the truncated C-terminal fragment $MORC4^{420-757}$, did not crystallise, suggesting that this region of MORC4 could be flexible or disordered. Flexible proteins do not often crystallise as this can interfere with nucleation, preventing the formation of protein crystals (Holcomb *et al.*, 2017).

Following several crystallisation screens, it was evident that $MORC4^{29-480}$ was more likely to crystallise when conditions were under low salt, low PEG and high a pH. However, $MORC4^{29-480}$ protein. Crystals of moderate resolution (approximately 2.7 Å) can be solved, however, crystals diffracting at 6 Å or more are very rarely solved (Wlodawer *et al.*, 2008), indicating that the $MORC4^{29-480}$ protein structure could not be solved as the screened crystal diffracted at 20 Å. Future work may include optimisation of crystal trials including incubation at 4°C in addition to using a more powerful source of X-rays such as a synchrotron. Furthermore, entropy reduction, which requires site-directed mutagenesis (SDM) to manipulate high energy residues such as lysine (K) or glutamic acid (E) to a lower energy residue such as alanine (A), reduces the surface entropy of the protein (Derewenda and Vekilov, 2006) and may be used to enhance protein crystallisation (Loll *et al.*, 2014). In addition, chemical methylation can also be utilised to methylate lysine residues, leading to a change in protein properties, improving the crystal packing and leading to a higher-quality crystal (Kim *et al.*, 2008).

In an attempt to develop an understanding of a potential structure of MORC4, protein homology modelling was required using several homology modelling servers to predict the structure of MORC4. The MORC4 model with MORC3 as a template created in SWISS-

MODEL was the most accurate representation of MORC3 as it did not include the H2K4me3 peptide, unlike the Phyre2 model. This MORC4 model was also presented as a homodimer, similar to that of MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016). However, the binding site of AMP-PNP was not conserved in this model along with the Mg^{2+} ion. This could suggest that the binding site of AMP-PNP is different to the binding site found in MORC3.

Robetta homology modelling was used to predict the *de novo* (Kim *et al.*, 2004) C-terminal domain structure of MORC4, and it was evident that the predicted models of MORC4 had a long extended C-terminal tail. Conversely, other MORC proteins are globular proteins (Douse *et al.*, 2018; Li *et al.*, 2016), suggesting that further analysis for the C-terminus of MORC4 is required. For example, small-angle X-ray scattering (SAXS) (Allec *et al.*, 2015) or full-length expression of recombinant MORC4 in insect cells using bacmids previously made in this study, in an attempt to increase the yield of protein.

To investigate MORC4 as an ATPase and to determine activity of recombinant MORC4 proteins, ATPase assays were established. However, time constraints resulted in the establishment of ATPase assays of a known ATPase protein, RecA. This malachite green ATPase assay did not appear robust as there were often variable results, although ssDNA stimulation of RecA was observed as expected (Gataulin *et al.*, 2018). Future work could include establishing the ATPase assay with recombinant WT and mutant MORC4 proteins. In addition, a more robust ATPase assay such as a NADH coupled assay (Lee *et al.*, 2007; Sebesta *et al.*, 2017; Kiianitsa *et al.*, 2003) or a radioactive ATP degradation assay resolved by thin layer chromatography (TLC) (Janas *et al.*, 2003) could be utilised for future experiments. These robust ATPase assays could be used if recombinant MORC4 mutant proteins are inactive as ATPases, and if the (methylated) histone peptides modulate ATP, as with MORC3 (Zhang *et al.*, 2019).

Analysis of the MORC4 C-terminus was important as the region is dissimilar in each MORC, suggesting that it plays a role in function. It may also suggest that each MORC is regulated by the Zf-CW domain as it is close to the coiled-coil domain. Some evidence also suggests that the CW coiled-coil domain is involved in oligomerisation, specifically in MORC2 (Xie *et al.*, 2019). Furthermore, analytical SEC of the C-terminal fragments only in this study found that a trimer was formed, supporting the theory that perhaps the MORC4 coiled-coil domain is also involved in dimerisation or other oligomerisation, as with MORC2. Future work may include analysis of the MORC4 coiled-coil domain with isothermal titration colorimetry (ITC) to assess histone peptide binding with all MORC4 recombinant proteins to investigate if the C-terminal coiled-coil region is involved in other functions such as regulating histone peptide binding by the Zf-CW domain.

Multiple sequence alignments (MSAs) of the protein sequences of the four human MORC protein paralogues revealed that the ATPase domains of all MORCs were highly conserved. It also revealed that MORC3 and MORC4 were similar in protein sequence in comparison to other MORCs, suggesting that MORC3 and MORC4 may have similar functions. This could give an indication as to why the homology modelling prediction of MORC4 using MORC3 as a template, led to a better-quality prediction of the MORC4 ATPase domain using SWISS-MODEL than with the MORC2 structural template. In addition, an MSA of the MORC4 protein sequences in vertebrates was also conducted in an attempt to investigate similarities and differences between species. In mammals, the MORC4 sequence of the ATPase and C-terminal regions was highly conserved, suggesting that MORC4 plays a similar role in all mammals.

Overall, this chapter has helped to give some biochemical and structural insight into MORC4 function, indicating it may be an active ATPase on the basis of protein sequence homology and its AMP-PNP induced dimerisation, as with MORC2 and MORC3. Although the MORC4

protein crystals could not be further optimised and the structure solved, this study demonstrated that bioinformatics approaches such as homology modelling and MSAs could be utilised in an attempt to predict the structure of MORC4 and to investigate protein sequences of MORC proteins in mammals.

CHAPTER 5

Results

Investigations into MORC4 expression in human cancers

Chapter 5 Investigations into MORC4 expression in human cancers

5.1 Introduction

Biomarkers are substances, structures or pathways which can be measured to predict disease outcomes (Strimbu and Tavel, 2010) and they are important for the determination of diagnosis, prediction and treatment of cancers (Henry and Hayes, 2012) and other diseases (Mayeux, 2004). Biomarkers can be up or downregulated in different types of diseases, suggesting that they play a role in the disease such as the ADP-ribosylation factor-like 3, which is downregulated in malignant brain tumours (Yulin Wang *et al.*, 2019). Whereas, the carcinoembryonic antigen is upregulated in patients with asthma (Nie *et al.*, 2019). This indicates that these proteins play a role in this specific type of disease and cancer.

Lymphoma is a type of cancer which originates from lymphocytes (Dalla-Favera, 2012) causing fever, night sweats and enlarged lymph nodes in patients (Mugnaini and Ghosh, 2016). Non-Hodgkin lymphoma (NHL) accounts for ~90% of all lymphoma diseases with only ~10% accounting for Hodgkin lymphoma (Shankland *et al.*, 2012). Diffuse large B-cell lymphoma (DLBCL) is the most common type of NHL (Friedberg and Fisher, 2008; Liu and Barta, 2019), which is rapidly progressing and aggressive (Liu and Barta, 2019). Some evidence suggests that MORC4 is highly expressed in NHL, in particular, DLBCL in comparison to normal healthy B-cells. Evidence also suggests that activated B-cell (ABC)-DLBCL MORC4 mRNA expression was significant higher compared to germinal center B-cell (GCB)-DLBCL (Liggins *et al.*, 2007), suggesting that MORC4 is a potential ABC-DLBCL biomarker.

To aid the determination of MORC4 function and its characterisation, it was important to investigate the role of MORC4 and any association with disease, in particular, human cancers. In this chapter, the original study of MORC4 mRNA expression in lymphoma (Liggins *et al.*, 2007) was expanded into different diseases and additional DLBCL cell lines, with the addition

of further bioinformatic analysis using publicly deposited datasets. Furthermore, protein expression of MORC4 was also investigated as protein expression is important as a clinical biomarker i.e. for immunohistochemical analysis in clinical patient samples (Sullivan and Chung, 2008). In an attempt to investigate protein expression of several diseases, full-length coding sequences of all four MORC paralogues with epitope protein tags (FLAG and HA) were transfected into mammalian cells lines, and identified using western blotting to help verify and characterise novel MORC4 monoclonal antibodies, which were previously optimised using a MORC paralogue for specificity. Commercial MORC polyclonal antibodies were also used to investigate protein expression patterns in several cancer cell lines. In addition, MORC4 RNA expression was analysed in several cancerous cell lines using publicly deposited datasets, reverse transcription (RT)-PCR and quantitative (q)RT-PCR to assess both mRNA transcript splicing variation and levels. Overall, this chapter gives further insight into the function of MORC4 in relation to its expression in human cancers.

5.2 MORC4 RNA analysis

5.2.1 Reverse transcription (RT)-PCR of MORC4

Reverse transcription (RT)-PCR is an end point PCR technique which was used to investigate potential splice variants of MORC4 found in various cancer cell lines by amplification of RNA. Quantitative (q)RT-PCR was required for subsequent experiments to quantitate mRNA expression levels. It was important to explore mRNA transcripts (and hence, protein isoforms) found in each type of cancer as this could help to distinguish similarities and differences across cell lines and diseases, which could help to elucidate MORC4 function in disease. Several RT-PCR primers were designed, and a combination of these primers were used for subsequent RT-PCR experiments (Table 5.1). Primers were designed (by Dr. Christopher Cooper, The University of Huddersfield) at the termini of 5’/3’ exons, ensuring that contamination of genomic DNA in the cDNA preparation would not be an issue, as if introns were amplified

from introns (or retained introns from splicing), then this could be screened out on agarose gels from the PCR product size. Primers were also not designed to span exon boundaries (which would completely remove the genomic DNA/retained potential), in case alternative splicing and exon switching meant that exon combinations could exist that may not have predicted and may have otherwise been missed.

Table 5.1: **Primers used for RT-PCR to identify MORC4 splice variants in several types of cancerous cell lines.** The table shows each primer name and primer sequence (Designed by Dr. Christopher Cooper, The University of Huddersfield).

Primer Name	Primer Sequence
MORC4_F1	TTCGGGATCCGCCTGAGC
MORC4_F2	TAGAGATTCAGGAGCTCC
MORC4_R1	GTCCAACTCAGATTTTCC
MORC4_R2	ATCCAGTATGTGATTTGC
MORC4_R3	CGAAGCTCCAAGTGAGGG
MORC4_R4	AGGTGTGTCTCTGTTTC

5.2.1.1 Sequence analysis of MORC4 mRNA splice variants

Prior to RT-PCR experiments, sequence analysis of MORC4 was conducted to identify known individual splice variants. This analysis helped to identify resulting potential protein isoforms for subsequent western blotting of MORC4 in cancer cell lines and also known splice variants which could be amplified during RT-PCR (Table 5.2). There were several known splice variants of MORC4 in public databases, two of which consisted of 17 exons (Ensembl and NCBI databases). One of the known isoforms of MORC4 encoded an extended C-terminus to previously used full-length MORC4, which ended in ‘HILD’ residues (Figure 5.1a) resulting in a protein of 106 kDa. The transcript of full-length MORC4 used throughout this study (Chapter 3) encoded an alternative C-terminus ending in ‘GAS’ residues (Figure 5.1b), resulting in a protein of 101 kDa.

***Table 5.2*: Bioinformatic analysis identified known and predicted several protein coding human splice variants of MORC4.** Data was collected from Ensembl, NCBI and UniProt public databases.

Accession Number	Known or Predicted?	Exons	Length (bp)	Length (aa)	Molecular Weight (kDa)
ENST00000355610.9	Known	17	3798	937	106
ENST00000255495.7	Known	17	2940	900	101
ENST00000604604.1	Known	2	2027	38	4
Q8TE76-2	Known	-	1947	648	74
NM_001085354.3	Known	17	3760	900	101
XM_011531027.2	Predicted	-	3556	760	87
XM_017029844.1	Predicted	-	4120	808	92
XM_006724691.2	Predicted	-	3902	815	97
XM_005262190.3	Predicted	-	4158	845	97

Sequence analysis was also used to map primers, facilitating their design for RT-PCR. Primer mapping revealed that not all of the known MORC4 splice variants contained the annealing positions of the RT-PCR designed primers (Appendix 8.20) (e.g. the ENST00000355610.9 splice variant had all of the annealing positions for all primers. In contrast, the small ENST00000604604.1 splice variant, which did not include much of the ATPase domain, also did not include any of the primer annealing positions, so would not be amplified by any of the primers during RT-PCR amplification (Table 5.3)). primer mapping was also used to calculate the sizes of the amplicons which were expected for each known splice variant, to aid identification of amplified transcripts (Table 5.4) (Appendix 8.20).

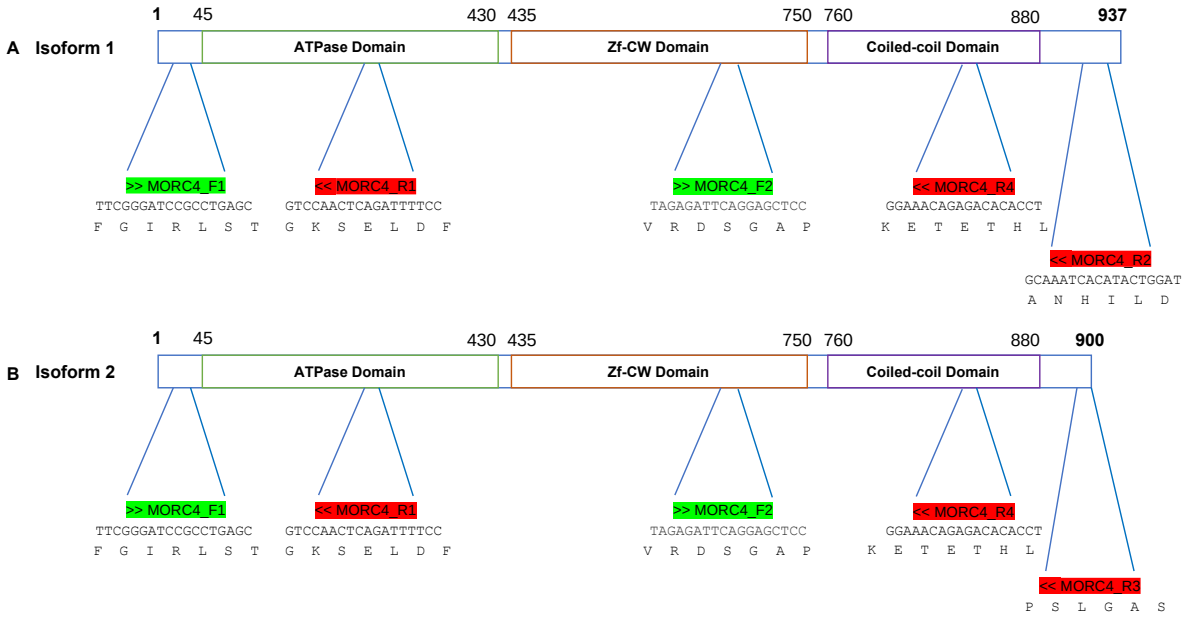


Figure 5.1: **Schematic diagram of two known splice variants of MORC4 including RT-PCR primer locations.** Simple modular architecture research tool (SMART) was used to identify protein domains and sequence analysis was used to determine locations of RT-PCR primers. (A) Less common MORC4 isoform with extended C-terminal ('HILD' residues). (B) commonly used MORC4 protein isoform used throughout this study (C-terminal ending in 'GAS' residues. Protein domains highlighted in green represents the ATPase domain, orange represents the Zf-CW domain and purple represents the coiled-coil domain. MORC4 forward and reverse primers are highlighted in green and red, respectively.

Table 5.3: **Primer combinations for each MORC4 splice variant.** The table shows each MORC4 splice variant and corresponding primer combinations. Primer combinations with a '+' means that both of the annealing positions for each primer listed can be found within the splice variant. '-' means that one or both of the annealing positions for each primer is not present in the splice variant.

	Primer Combinations				
Domain(s)	**ATPase**	**Zf-CW**	**Zf-CW**	**Zf-CW**	**ATPase-Zf-CW**
Splice variants	**F1/R1**	**F2/R2**	**F2/R3**	**F2/R4**	**F1/R4**
ENST00000355610.9	+	+	+	+	+
ENST00000255495.7	+	-	+	+	+
ENST00000604604.1	-	-	-	-	-
Q8TE76-2	-	-	+	+	-
NM_001085354.3	+	-	+	+	+
XM_011531027.2	-	+	+	+	-
XM_017029844.1	-	-	+	+	-
XM_006724691.2	+	+	+	+	+
XM_005262190.3	-	+	+	+	-

Table 5.4: **Primer combinations and corresponding sizes of MORC4 amplicons following RT-PCR**. This table shows primer combinations and expected PCR product sizes. Sizes that are underlined are splice variants which are different to the most common sizes observed for other transcripts.

	Primer combination expected sizes (kb)				
Domain(s)	**ATPase**	**Zf-CW**	**Zf-CW**	**Zf-CW**	**ATPase-Zf-CW**
Splice variants	**F1/R1**	**F2/R2**	**F2/R3**	**F2/R4**	**F1/R4**
ENST00000355610.9	0.6	0.7	0.6	0.5	2.5
ENST00000255495.7	0.6	-	0.6	0.5	2.5
ENST00000604604.1	-	-	-	-	-
Q8TE76-2	-	-	0.6	0.5	-
NM_001085354.3	0.6	-	0.6	0.5	2.5
XM_011531027.2	-	0.7	0.6	0.5	-
XM_017029844.1	-	-	0.6	0.5	-
XM_006724691.2	0.25	0.7	0.64	0.5	2.1
XM_005262190.3	-	0.7	0.64	0.5	-

5.2.1.2 RT-PCR identified MORC4 potential splice variants

Prior to amplification of RNA, total RNA was extracted from cells and purified using a RNeasy Mini Kit. RNA concentration quantitation and quality assessment revealed that all samples were of a sufficient concentration and quality, as all samples had a 260/280 nm value of 2.0 or above, indicating there was no contamination of DNA in any samples (Appendix 8.21) (Koetsier and Cantor, 2019).

5.2.1.2.1 Quality assessment of cDNA using TATA-binding protein (TBP) amplification

Prior to the assessment of MORC4 splice variants, a control RT-PCR was completed to assess the quality of the cDNA preparation from each cell line, to ensure it gave comparably efficient amplification for a 'housekeeping gene'. TATA-binding protein (TBP) is a commonly used housekeeping gene (Kozera and Rapacz, 2013; Liggins *et al.*, 2007; Lossos *et al.*, 2003) and was utilised as a reference gene for this experiment, to assess cDNA integrity. The RT-PCR results showed that TBP was successfully amplified in every cell line except for SK-N-AS (Figure 5.2). This indicates that all cell lines, except for the SK-N-AS neuroblastoma cell line,

were of suitable cDNA integrity and were suitable for use and to assess MORC4 splice variants. The SK-N-AS cDNA preparation was repeated in an attempt to synthesise cDNA for RT-PCR. A small sample of negative controls (comprising cDNA synthesis reactions with no reverse transcriptase added) were also used which did not amplify any TBP, as expected (Figure 5.2). Negative controls of OCI-Ly3, RIVA, SU-DHL-2, NCI H929, JJN.3, RPMI8226, SK-N-BE(2c), IMR32 and Kelly cell lines were also used in cDNA synthesis and did not amplify following RT-PCR. This suggests that there was no genomic DNA contamination, indicating the samples were suitable for RT-PCR and qRT-PCR.

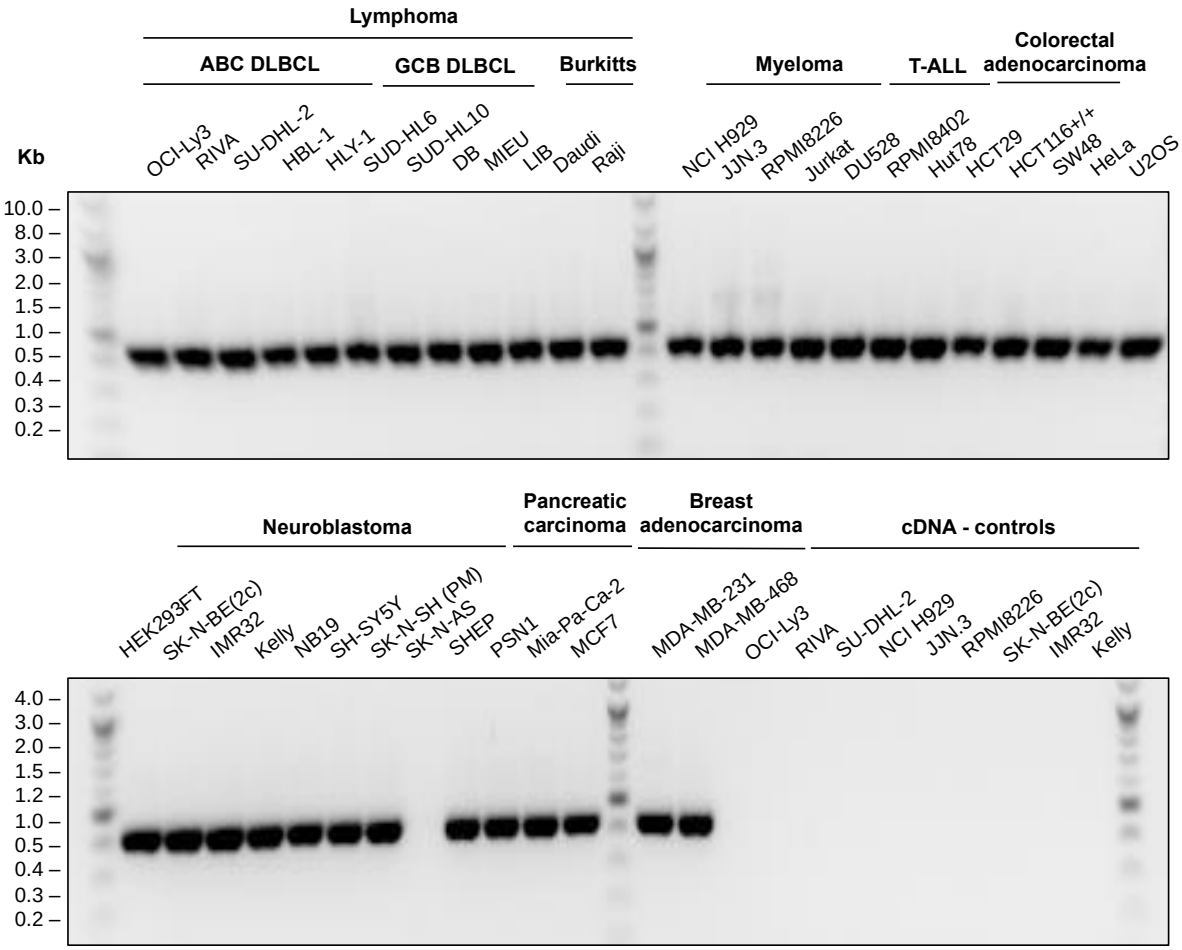


Figure 5.2: **RT-PCR of cancer cell lines using TBP primers to assess cDNA integrity**. Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with TBP primers (35 cycles) and resolved on a 2% agarose gel.

5.2.1.2.2 MORC4 ATPase domain transcript amplification

RT-PCR amplified the DNA region encoding for the ATPase domain of MORC4 in all of the tested cancer cell lines using a combination of F1 and R1 primers. Sequence analysis of several known splice variants of MORC4 revealed that F1 and R1 primers should amplify transcripts of approximately 0.6 kb, with one splice variant amplifying a 0.25 kb transcript (Table 5.4). cDNA for all cell lines was successfully amplified using F1 and R1 primers and all cell lines exhibited a transcript at approximately 0.6 kb (Figure 5.3). However, a small number of cell lines such as T-ALL (RPMI8402), colorectal adenocarcinoma (HCT116+/+ and SW48), pancreatic carcinoma (PSN1) and breast adenocarcinoma (MDA-MB-231) cell lines also showed an additional transcript at approximately 0.25 kb. This suggests that all cell lines could potentially have the splice variants amplified at 0.6 kb such as ENST00000355610.9, ENST00000255495.7 and NM_001085354.3. However, it also suggests that some cell lines also express the XM_006724691.2 splice variant, which was amplified at 0.25 kb. This also indicates that all the isoforms in these cell lines will comprise the ATPase region of MORC4, as some of the ATPase region was lost with the small isoform, potentially causing functional consequences. The positive control RT-PCR was the cloned ENST00000255495.7 MORC4 splice variant, and this sample also successfully amplified using the F1 and R1 primers. The negative control appeared to show some amplification; however, this amplification was likely overspill from the positive control as there was a high amplification of the positive control. In addition to these specific amplicons, several cell lines also showed a transcript at approximately 2.0 kb. However, sequence analysis revealed there were no known splice variants found at this size, indicating that either this 2.0 kb amplicon is either a novel splice variant, or a non-specific PCR product produced from the high number of cycles (40) in this RT-PCR.

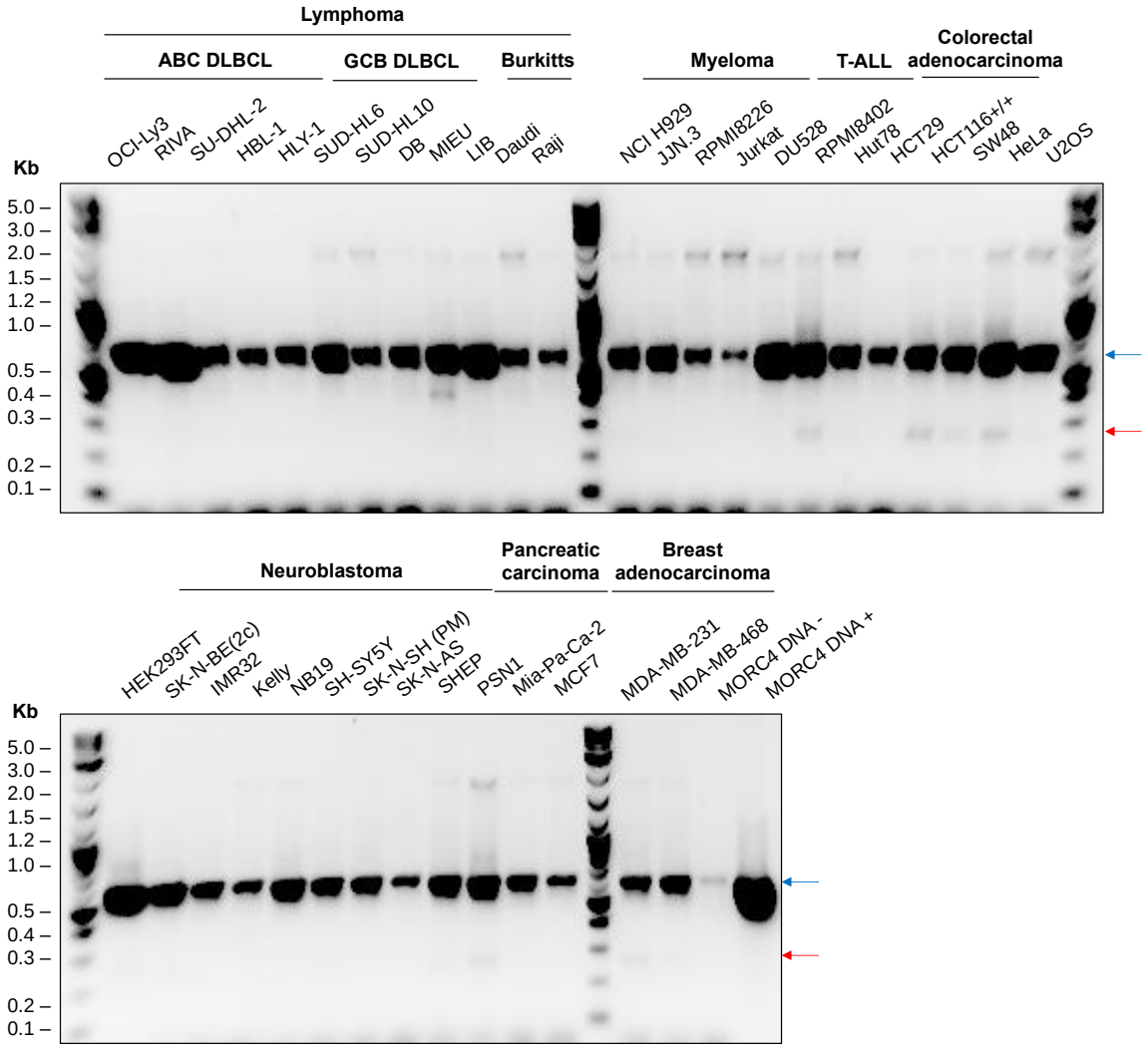


Figure 5.3: **RT-PCR MORC4 mRNA splice variant analysis of cancer cell lines using F1 and R1 primers.** Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with F1 and R1 primers (40 cycles) and resolved on a 2% agarose gel. HA-tagged MORC4 DNA was used as positive and negative controls. Blue arrows and red arrows represent expected size of common mRNA splice variant and expected size of rare mRNA splice variant, respectively.

5.2.1.2.3 MORC4 C-terminal transcript amplification

All cancer cell line MORC4 mRNA was not amplified with a transcript identified from sequence analysis using MORC4 F2 and R2 primers, which amplify the region encoding the C-terminal coiled coil region of MORC4. This suggests that the functional coiled-coil region of MORC4 is compromised in these cell lines. Sequence analysis of known splice variants of MORC4 suggest that a combination of F2 and R2 primers in RT-PCR should create a transcript of approximately 0.7 kb (Table 5.4). This combination of primers should have amplified the MORC4 transcript encoding the commonly found 'HILD' C-terminal end. However, the agarose gel revealed that there was a splice variant at approximately 2.0 kb, suggesting this amplicon was either a novel splice variant or a non-specific amplicon (Figure 5.4). This indicates that there could be an issue with the design of the primers or that the isoforms found in these cell lines do not have the 3' exon. This was unexpected, as public database analysis of potential splice variants suggested that the encoded 'HILD' end was more likely as it was present in the majority of predicted splice variants of MORC4. Following this RT-PCR, the F4 primer was designed, which was an exon just upstream of the alternative 3' ends in an attempt to solve this issue. As expected, the standard positive control was not amplified with these primers, as the control encoded the less common 'GAS' C-terminus end instead of the 'HILD' C-terminus screened for with this primer set.

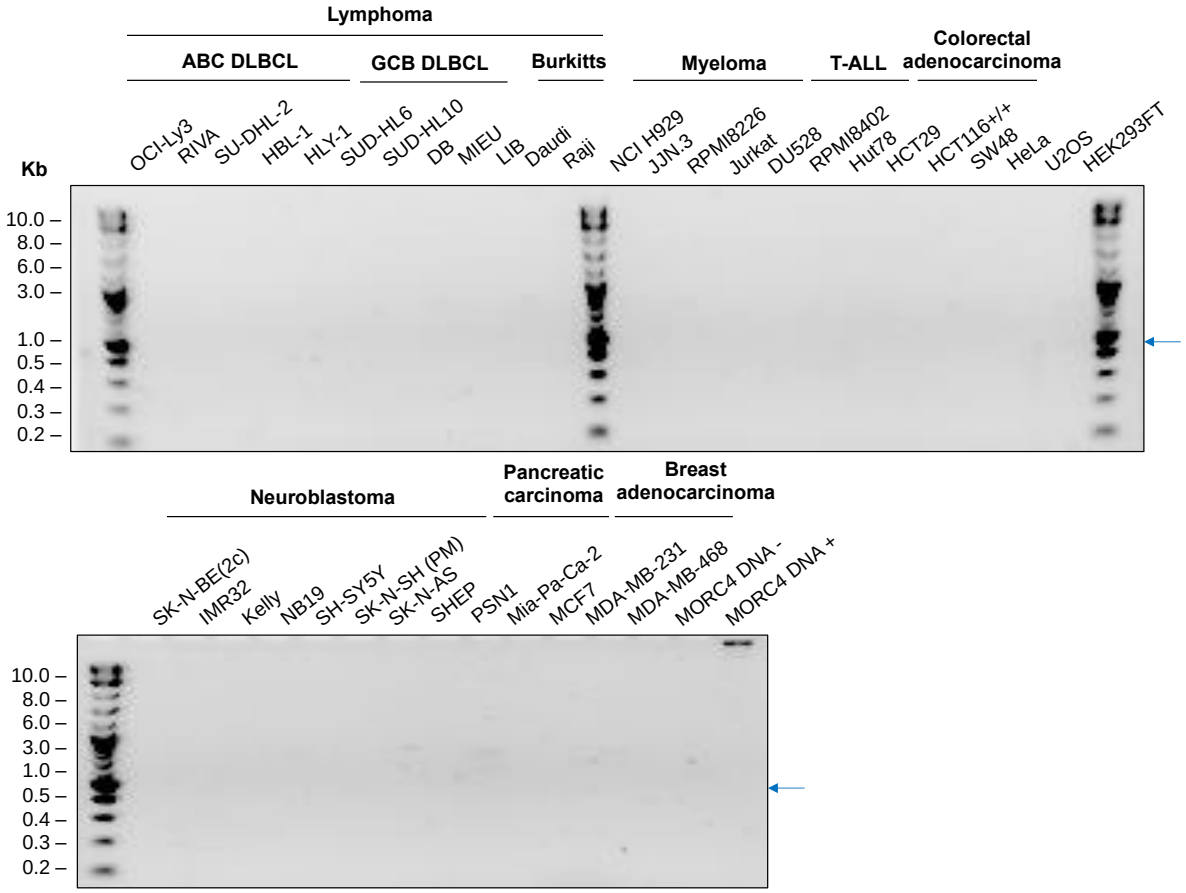


Figure 5.4: **RT-PCR MORC4 mRNA splice variant analysis of cancer cell lines using F2 and R2 primers.** Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with F2 and R2 primers (35 cycles) and resolved on a 2% agarose gel. HA-tagged MORC4 DNA was used as positive and negative controls. Blue arrows represent the expected size of common mRNA splice variant.

The combination of F2 and R3 primers amplify a transcript of MORC4 which encoded the less common 'GAS' C-terminal end, which was only identified in a single transcript in the public databases and was used for all previous experiments in this study. Sequence analysis suggested that these primers would amplify MORC4 transcripts at 0.6 kb, with two splice variants amplifying at 0.64 kb (Table 5.4). This RT-PCR revealed that there was some amplification in colorectal adenocarcinoma, neuroblastoma, pancreatic carcinoma and breast adenocarcinoma cell lines. It also revealed that all of these cell lines had amplification at both 0.6 and 0.64 kb (Figure 5.5). This suggests that each cell line has two splice variants of this region of MORC4.

It also indicates that the cell lines with amplification have the transcript of MORC4 encoding the 'GAS' C-terminal end, commonly used in this study. However, amplification was not seen in every cell line, suggesting that this C-terminal end is not as common in cell lines. As predicted, the MORC4 positive control also amplified at 0.6 kb as it encoded the 'GAS' C-terminal end.

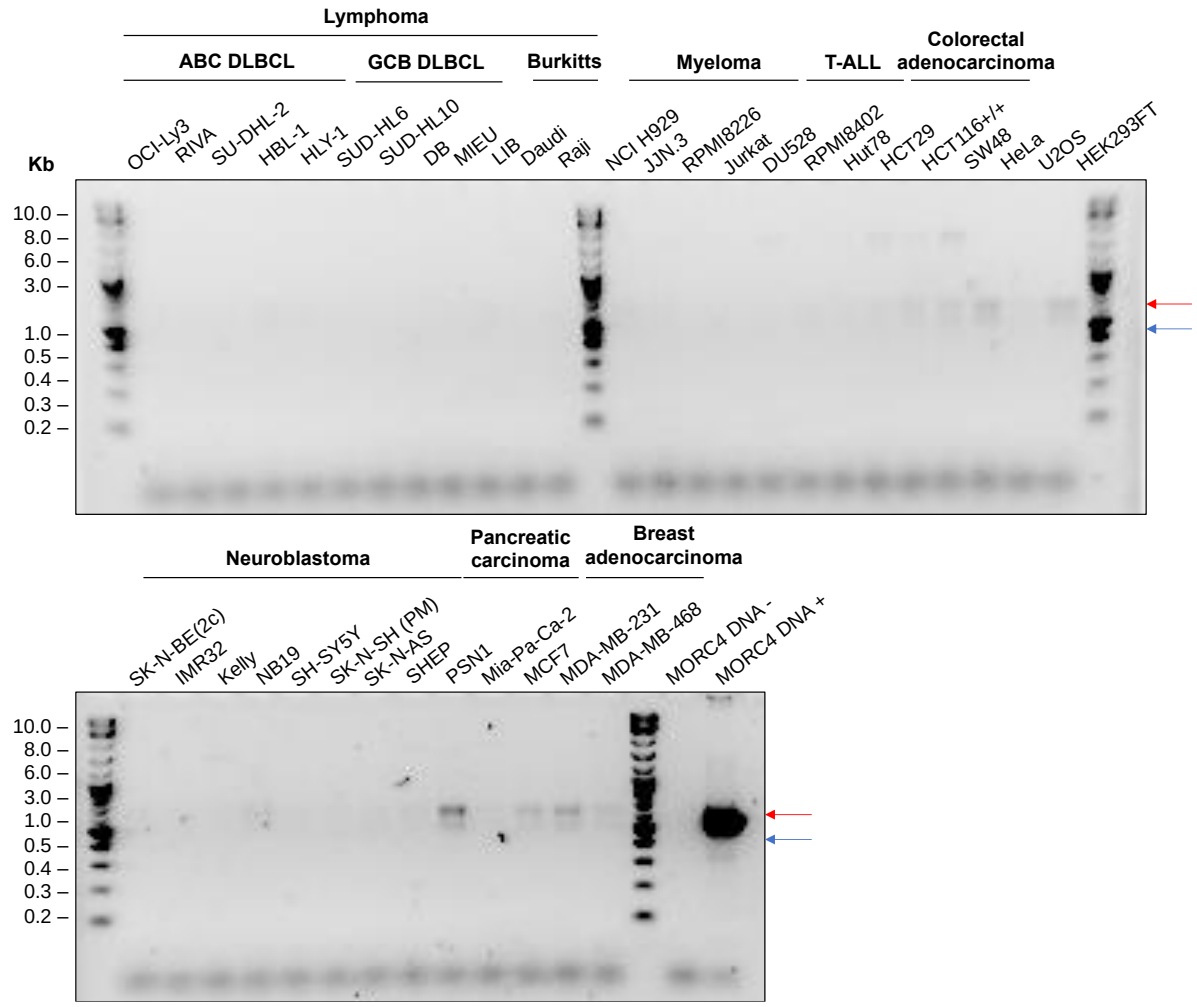


Figure 5.5: **RT-PCR MORC4 mRNA splice variant analysis of cancer cell lines using F2 and R3 primers.** Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with F2 and R3 primers (40 cycles) and resolved on a 2% agarose gel. HA-tagged MORC4 DNA was used as positive and negative controls. Blue arrows and red arrows represent expected size of common mRNA splice variant and expected size of rare mRNA splice variant, respectively.

Sequence analysis also suggested that a combination of F2 and R4 primers in RT-PCR should create a transcript of approximately 0.5 kb for all known splice variants of MORC4 (Figure 5.1 and 5.4). The R4 primer has an annealing position which is an exon that is common with both the common 'HILD' C-terminal end isoform and the less common 'GAS' C-terminal end isoform of MORC4. The RT-PCR result showed that the majority of cell lines had an amplicon of 0.5 kb including the positive control (HA-tagged MORC4 DNA). This suggests that the majority of the cell lines used in this study have a common exon towards the C-terminus. The lack of additional PCR products also suggests that there was no alternative splicing. In addition, this 0.5 kb transcript was not seen in all GCB DLBCL cell lines, which could suggest very low MORC4 RNA levels in these particular cell lines (SUD-HL10, DB and MIEU) (Figure 5.6), as previously found in another qRT-PCR study on MORC4 (Liggins *et al.*, 2007).

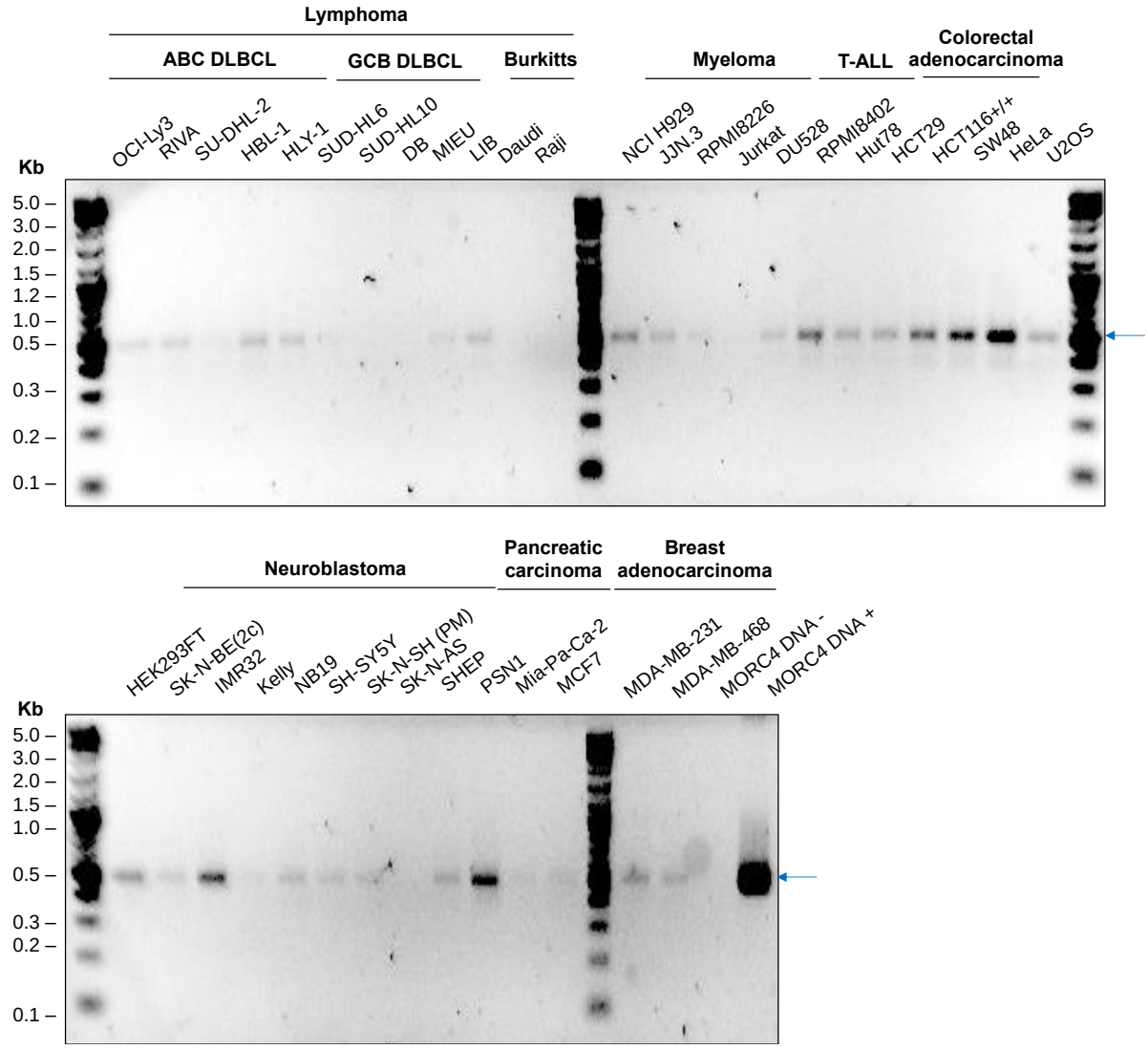


Figure 5.6: **RT-PCR MORC4 mRNA splice variants analysis of cancer cell lines using F2 and R4 primers**. Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with F2 and R4 primers (40 cycles) and resolved on a 2% agarose gel. HA-tagged MORC4 DNA was used as positive and negative controls. Blue arrows represent the expected size of common mRNA splice variant.

5.2.1.2.4 MORC4 ATPase-Zf-CW tandem domain transcript amplification

The final RT-PCR of MORC4 was conducted with a combination of F1 and R4 primers as both the F1 primer and R4 were previously found to amplify potential transcripts. Therefore, using this primer combination will amplify potential splice variants spanning this region of sequence. Sequence analysis suggested that the majority of known splice variants should produce a PCR

product of 2.5 kb (almost the full-length of the MORC4 transcript), with an exception of one splice variant which was expected to produce a PCR product of 2.1 kb (Table 5.4). The RT-PCR results showed that the majority of transcripts were 2.1 kb, with some cell lines also producing an additional transcript at ~2.5 kb such as the SW48, PSN1, NB19 and U2OS cell lines (Figure 5.7). This suggests that the majority of cell lines have the known splice variant, XM_006724691.2, which encodes the MORC4 ATPase region and the C-terminal end in common with both the 'HILD' and 'GAS' C-terminal variants. It also suggests that the majority of cell lines tested expressed full-length MORC4 transcripts. However, full-length MORC4 was not observed in the majority of neuroblastoma cell lines, except for the SHEP cell line, which had the 2.1 kb MORC4 transcript. In addition, the MORC4 DNA positive control was 2.5 kb as expected (Figure 5.7), indicating that future work may include cloning and sequencing of multiple 2.1 kb splice variants in order to confirm the identity.

In addition, the normal transformed embryonic kidney cell line, HEK293FT, had a larger 2.5 kb transcript (Figure 5.7), suggesting that perhaps this transcript could be the 'normal' variant of MORC4. In an attempt to investigate this, future work may involve the amplification of tissue cDNA as HEK293FT is a normal but transformed cell line. Several smaller PCR products were also seen following this RT-PCR, however, there were no known splice variants with a similar size. This suggests that these small PCR products are either novel splice variants that have not yet been identified or they are non-specific PCR products produced from a high number of cycles (40) performed in these RT-PCRs.

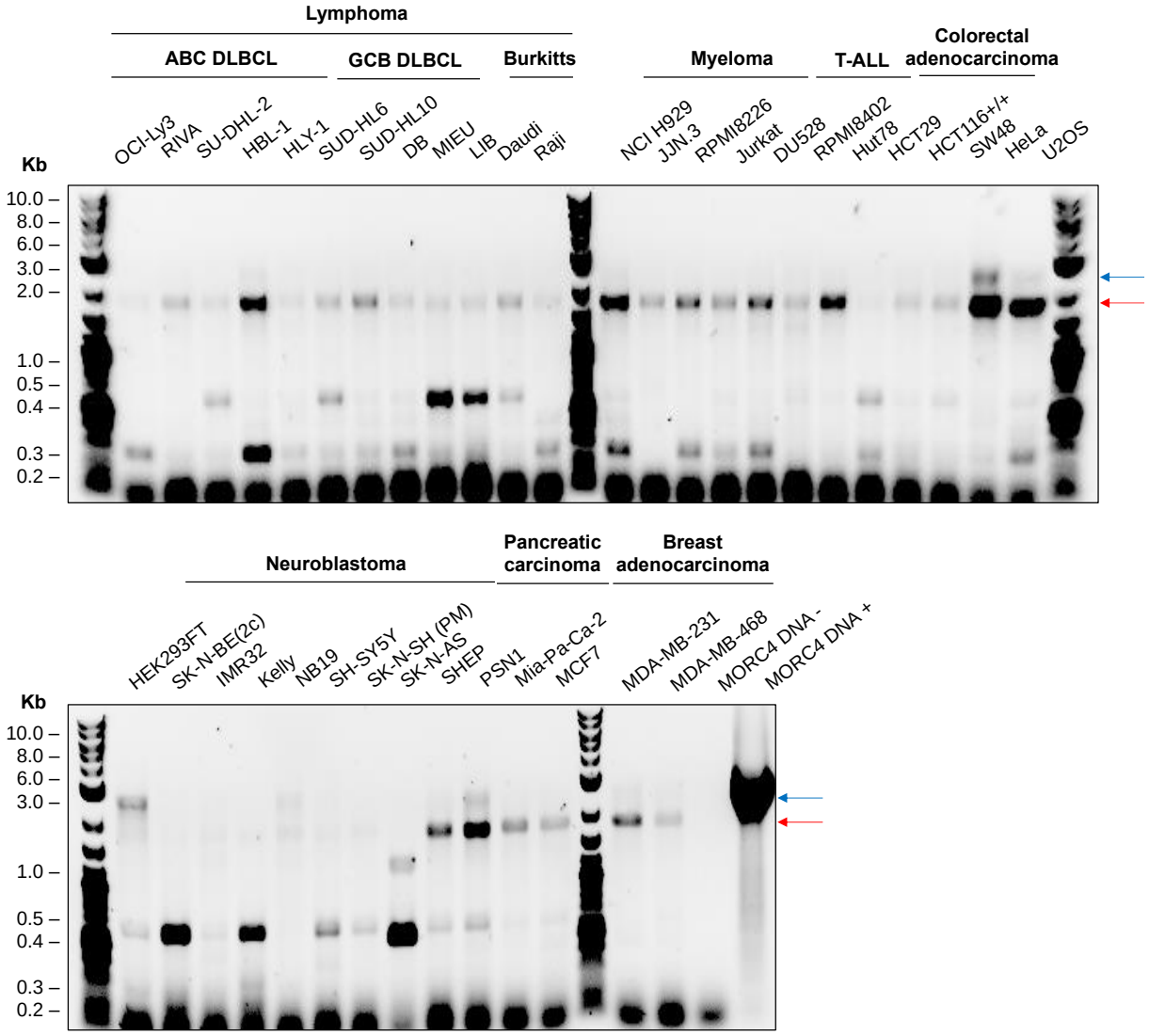


Figure 5.7: **RT-PCR MORC4 mRNA splice variant analysis of cancer cell lines using F1 and R4 primers**. Cells were harvested and RNA was extracted using an RNA extraction kit. RNA was used to synthesise cDNA for RT-PCR MORC4 mRNA splice variant analysis. Cell line cDNA was amplified with F1 and R4 primers (40 cycles) and resolved on a 2% agarose gel. HA-tagged MORC4 DNA was used as positive and negative controls. Blue arrows and red arrows represent expected size of common mRNA splice variant and expected size of rare mRNA splice variant, respectively.

5.2.1.2.5 Summary of MORC4 transcript amplification

In summary, RT-PCR analysis suggested that all of the cell lines used in this study are likely to have the N-terminal ATPase region of MORC4 present in the splice variants. However, there was no evidence to suggest that these cell lines expressed transcripts encoding the extended 'HILD' amino acid C-terminal end. Some evidence also suggests that some of the cell lines

used in this study have the 'GAS' amino acid C-terminal end, which is a less common C-terminal end transcript (identified by Ensembl and NCBI public databases), such as colorectal adenocarcinoma, pancreatic carcinoma and breast adenocarcinoma cell lines. Data presented here also suggested that the majority of cell lines have a full-length splice variant of (encoding the C-terminal end in common with 'HILD' and 'GAS' isoforms) MORC4 except for the majority of neuroblastoma cell lines, which did not reveal any MORC4 transcript with F1 and F4 primers. However, transcripts were visible in neuroblastoma cell lines when F1 and F4 primers were used as another combination. Cell lines which contained the near full-length splice variant of MORC4 would have also contained the Zf-CW domain and coiled-coil regions in addition to the ATPase domain, suggesting that these domains are important for function of MORC4 in the majority of cell lines tested. Future work may involve cloning and sequencing amplified PCR products, in particular the 2.1 kb product near the full-length splice variant and unexpected amplified transcripts, in order to ascertain identity and to eliminate potential primer dimers. If multiple splice variants in cell lines were confirmed by sequencing, this could indicate that there are two potential simultaneous MORC4 roles found in those cell lines.

5.2.2 Quantitative reverse transcription-polymerase chain reaction (qRT-PCR) of MORC4 mRNA expression in cancer cell lines

Following production of cDNA, two-step quantitative reverse transcription polymerase chain reaction (qRT-PCR) was utilised for amplification and quantitation of MORC4 mRNA in several cancer cell lines including lymphoma, breast cancer, pancreatic cancer and neuroblastoma cell lines. Prior to normalisation of qPCR data, Δ cycle threshold (Ct) value was calculated and plotted to investigate the MORC4 total expression of each cell line in comparison to the housekeeping gene (Rao *et al.*, 2013), TBP. Evidence suggests that the lower the Ct value the higher the higher the RNA expression (Goni *et al.*, 2009). Therefore, ΔCt analysis of various cancer cell lines revealed that Raji, Daudi, DU528 and Jurkat lymphoma

cell lines had the highest ΔCt values (approximately 6 ΔCt) (Figure 5.8), suggesting that they were the lowest expressing of MORC4 mRNA. In addition, the pancreatic cancer cell line, PSN1, had a ΔCt value of -0.5 (Figure 5.8), suggesting that PSN1 was the highest expressing cell line of MORC4 mRNA. The negative ΔCt value of PSN1 also suggests that PSN1 was expressed at a higher level of MORC4 than the housekeeping gene, TBP.

To calculate the relative quantification (RQ) value (fold change), the ΔCt values were normalised using low expressing non-Hodgkin's Burkitts lymphoma cell lines Raji and Daudi to give $2^{\Delta\Delta Ct}$ values. Burkitts lymphoma cell lines were used for normalisation as no normal B-cells ($CD19^{+}/CD20^{+}$) or healthy tissue were available to calculate a true $2^{\Delta\Delta Ct}$ value. Initially, only Raji was used to normalise qPCR data, however, a combination of both Raji and Daudi was then used to ensure accuracy as it was a 'population' of similarly low expressing cell lines rather than a single cell line (Figure 5.9). Burkitts lymphoma expression was also similar to germinal center B-cell (GCB)-DLBCL expression, which is interesting as they both originate from centroblast cells in the B-cell maturation process (Weigert and Weinstock, 2012).

qRT-PCR analysis showed that MORC4 mRNA expression was highest in ABC-DLBCL of all of the lymphoma cell lines tested (Figure 5.9). It was also revealed that MORC4 mRNA expression in some ABC DLBCL cell lines (HLY-1 and OCI-Ly3) was approximately 20-fold higher than GCB DLBCL cell lines (Figure 5.9), as with a previous study using a smaller panel of lymphoma cell lines (Liggins *et al.*, 2007). In addition, the HBL-1 ABC-DLBCL lymphoma cell line expression was especially high (60-fold higher) in comparison to GCB-DLBCL cell lines. However, statistical analysis was performed using a Kruskal Wallis test (non-parametric data) on Statistical Product and Service Solutions (SPSS) software following data distribution analysis and revealed that there was no significant difference between any of the lymphoma cell line subgroups ($p = 0.344$). However, there was almost a significant difference between B-cell and T-cell lymphomas ($p = 0.055$). Conversely, MORC4 mRNA expression in colorectal

adenocarcinoma ($p = 0.014$), miscellaneous (e.g. HeLa and HEK293FT) ($p = 0.003$) and breast cancer cell lines ($p = 0.009$) were significantly higher than expression in lymphoma cell lines when grouped together (Figure 5.9). Relatively low expression was found for neuroblastoma cell lines except for the N-Myc proto-oncogene protein (MYCN) positive cell line, NB19. In addition, the difference between MYCN positive and MYCN negative neuroblastoma cell lines was close to the significant cut off point of $p = 0.05$ ($p = 0.055$). The pancreatic adenocarcinoma cancer cell line, PSN1, showed the highest expression all of the cell lines used in this study and had a relative quantification (RQ) value of approximately 110-140. However, the pancreatic carcinoma cell line, Mia-Pa-Ca-2 had low RNA expression with an RQ of approximately 10, which is low expression in comparison to PSN1 but high in comparison to the normalising Burkitts lymphoma cell lines, Raji and Daudi. This suggests that MORC4 could potentially play a role in pancreatic adenocarcinomas, although additional pancreatic cancer cell lines would need to be assessed to confirm this. Similar MORC4 mRNA expression was seen for all breast cancer cell lines, although MDA-MB-231 was the highest at approximately 60 RQ. In addition, the difference identified between oestrogen receptor (ER) positive or negative breast cancer cell lines was close to the significant cut off point ($p = 0.055$) (Figure 5.9).

The semi-quantitative RT-PCR results also supported the results from the qPCR analysis. RT-PCR found that there was higher expression of ABC-DLBCL cell lines compared to GCB-DLBCL. RT-PCR found that there was low expression of MORC4 transcripts in Daudi and Raji Burkitts cell lines (Section 5.2.1.2), supporting the proposal that these cell lines expressed low levels of MORC4 RNA also identified in qPCR. RT-PCR also determined that the highest expressing cell lines were derived from colorectal, miscellaneous (HeLa) and pancreatic cancers (PSN1), similar to qRT-PCR results. In addition, RT-PCR also identified MORC4 transcripts in neuroblastoma cell lines for the F2/F4 primer combination. However, the full-length MORC4 transcript was not identified in neuroblastoma cell lines. Nonetheless, qRT-

PCR analysis revealed that neuroblastoma cell lines expressed MORC4 mRNA at similar levels to T-ALL, GCB-DLBCL and myeloma cell lines (Figure 5.9). However, qRT-PCR detected these cell lines with the full-length MORC4 transcript as qRT-PCR only detects one exon boundary (13-14 exon) (Probe ID: Hs00226482_m1). qRT-PCR also found that the lowest MORC4 mRNA expressing cell lines were GCB-DLBCL subtypes, which was also identified in RT-PCR analysis (Section 5.2.1.2).This data supports and expands from the previous study, which identified MORC4 as a potential biomarker of ABC-DLBCL subtype (Liggins *et al.*, 2007). Overall, it was evident that both the RT-PCR and qRT-PCR analysis results had similar conclusions in relation to expression of MORC4 mRNA.

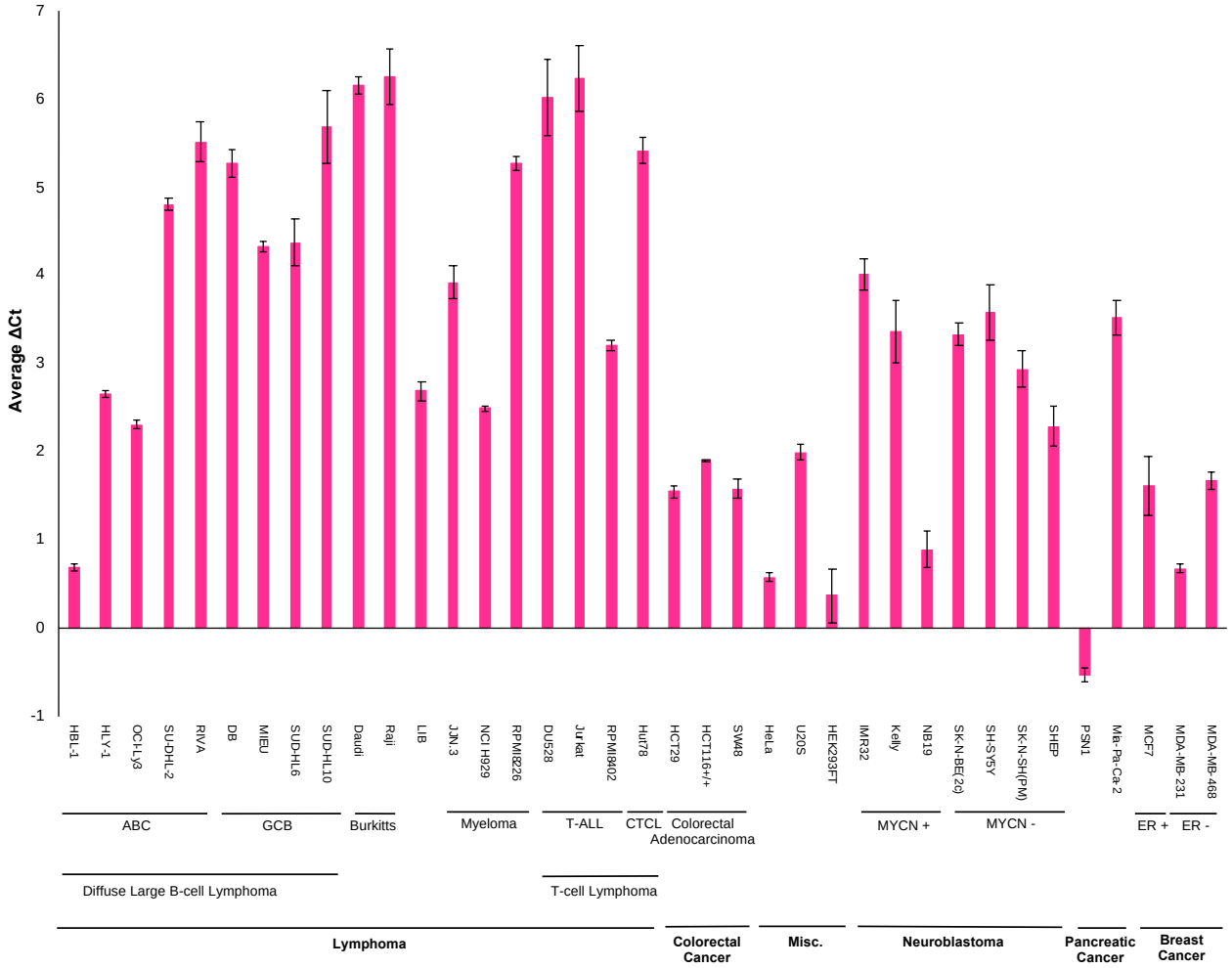


Figure 5.8: **Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) of cancerous cell lines to determine unormalised mRNA expression of MORC4 (Average ΔCt).** Cells were harvested and mRNA was extracted using a mRNA extraction kit. mRNA was used to synthesise cDNA and a dilution of 1/25 cDNA was used for qPCR. ΔCt was calculated using TBP as the house keeping gene (representative of technical replicates, n=4).

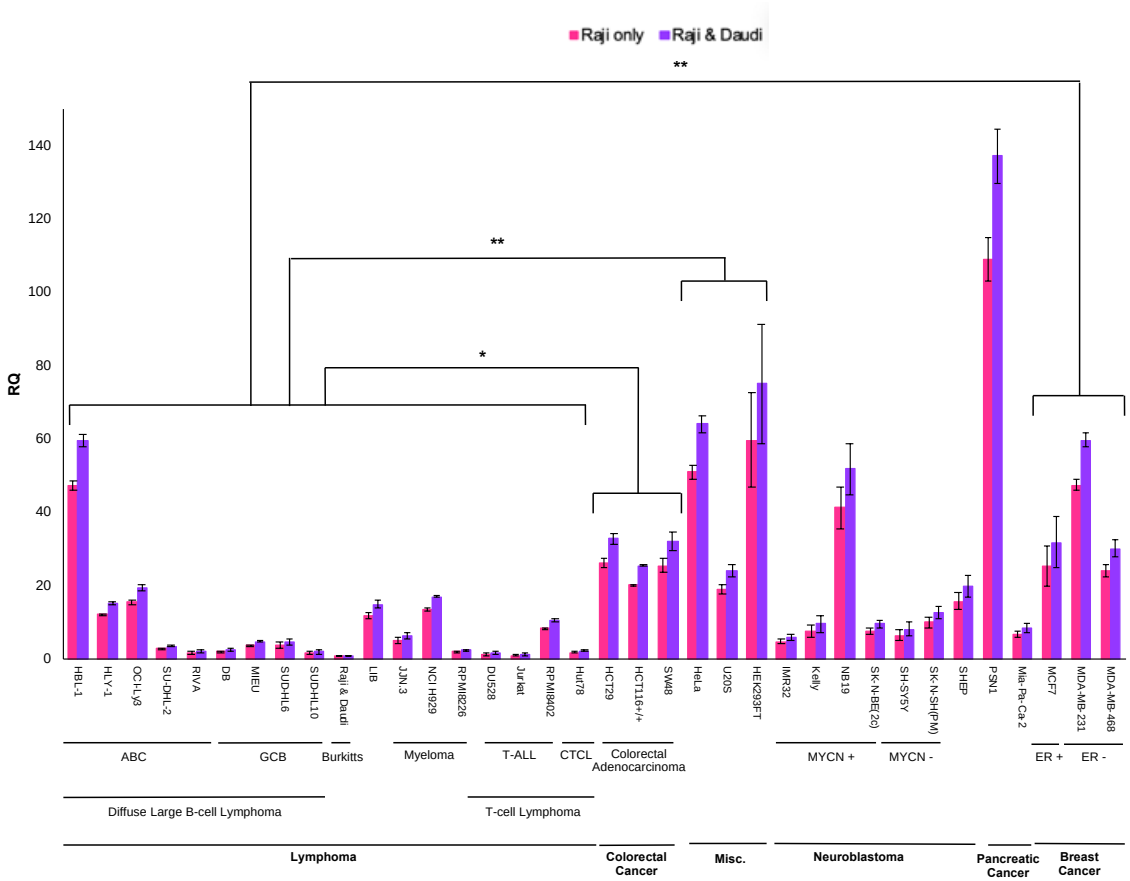


Figure 5.9: **Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR) of cancerous cell lines to determine normalised mRNA expression of MORC4 (RQ).** Cells were harvested and mRNA was extracted using a mRNA extraction kit. mRNA was used to synthesise cDNA and a dilution of 1/25 cDNA was used for qPCR. RQ was calculated using TBP as the house keeping gene and either Raji only (for normalisation) or both Raji and Daudi samples were used to normalise cell lines (representative of technical replicates, n=4 (* = $p \leq 0.05$ and ** = $p \leq 0.01$). Statistical analysis was performed using a Kruskal Wallis test (non-parametric data) using Statistical Product and Service Solutions (SPSS) software.

5.2.3 Bioinformatics analysis of MORC4 RNA expression

A number of RNA expression datasets were used to investigate the MORC4 gene expression across several normal and cancerous tissues. Gene expression analysis also determined survival and prognosis characteristics of several types of cancers expressing MORC4. RNA expression analysis could potentially help to determine MORC4 function and role in disease.

5.2.3.1 MORC4 RNA expression profiles

Several RNA expression databases were used to analyse MORC4 gene expression including the Gene Expression Profiling Interactive Analysis (GEPIA) (Tang *et al.,* 2017), the Gene Expression Database of Normal and Tumour Tissues 2 (GENT2) (Park *et al.,* 2019) and the RNA Expression Atlas by the European Bioinformatics Institute (EBI) (Petryszak *et al.*, 2017). The EBI RNA expression database revealed that there was variation of expression of MORC4 in non-cancerous tissues (Figure 5.10). Specifically, the placenta had the highest expression of MORC4 at approximately 170 transcripts per million (TPM), similar to a previous study which investigated mRNA expression levels in normal tissues from a radioactive dot plot (Liggins *et al.*, 2007). However, the pancreas and the colon expressed low levels of MORC4 at 5 and 10 TPM, respectively (Figure 5.10a). Pancreatic adenocarcinoma and colon adenocarcinoma cell lines were also investigated following high expression in RT-PCR and qRT-PCR experiments (Section 5.2.1.2 and 5.2.2), although, RT-PCR was an endpoint experiment so it could not be utilised to accurately determine expression levels, only for those cell lines which did not give much product even after a large number of cycles (40). The RNA expression datasets revealed that some pancreatic adenocarcinoma cell lines had increased expression of MORC4 in comparison to normal non-cancerous pancreatic tissue (Figure 5.10a and b). More specifically, a pancreatic adenocarcinoma cell line, PSN1, had a TPM of 85, unlike healthy pancreatic tissue which had an TPM of approximately 5 (Figure 5.10a and b). qRT-PCR analysis of the PSN1 cell line (Section 5.2.2) also showed evidence of high MORC4 RNA expression with an RQ

value of approximately 140. In addition, several colon adenocarcinoma cell lines also showed increased MORC4 RNA expression such as SW48, which increased 7-fold compared to non-cancerous colon tissue (Figure 5.10c).

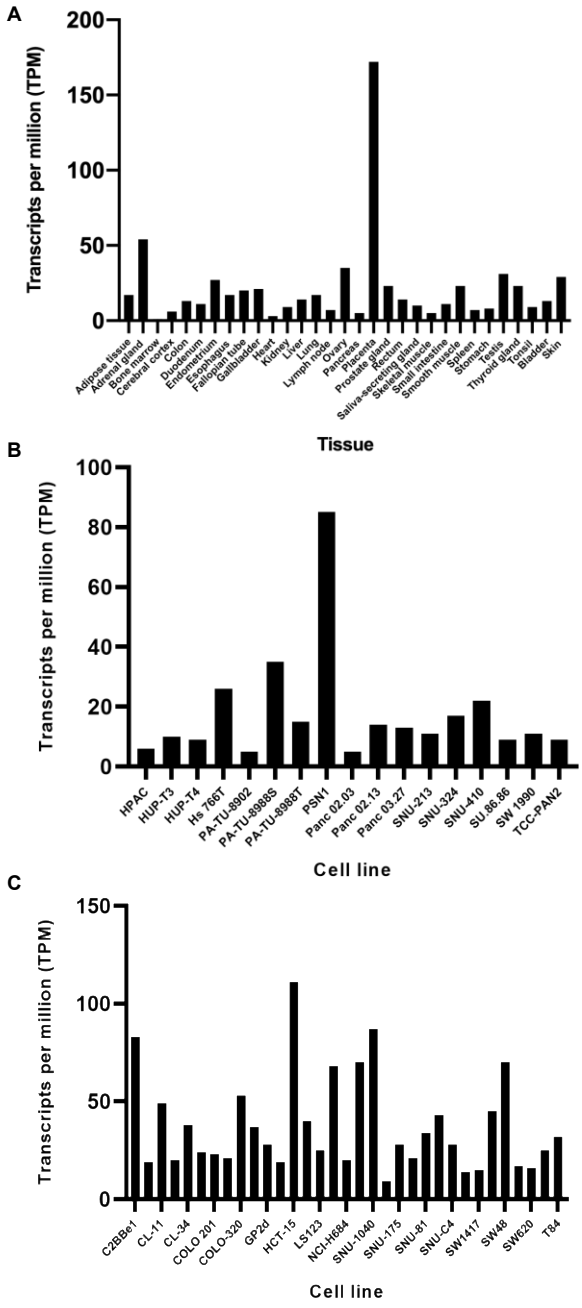


Figure 5.10: **MORC4 RNA expression dataset of normal non-cancerous tissue, pancreatic adenocarcinomas and colon adenocarcinoma cell lines using the RNA Expression Atlas by the European Bioinformatics Institute (EBI).** MORC4 RNA expression of **(A)** non-cancerous tissues; **(B)** Pancreatic adenocarcinoma cell lines; **(C)** Colon adenocarcinoma cell lines.

The GENT2 RNA expression database was also used to investigate MORC4 RNA expression in normal and cancerous tissues. Statistical analysis was performed using Prism software and initially involved the analysis of the distribution of data. This revealed that the data was normally distributed (parametric) and therefore led to an analysis of variance (ANOVA) test. The statistical analysis identified that there was a significant increase of MORC4 RNA expression found in the bladder ($p = 0.0005$), blood ($p = <0.0001$) brain (<0.0001), colon ($p = <0.0001$), ovary ($p = <0.0001$), skin ($p = <0.0001$), stomach ($p = <0.0001$), lung ($p = <0.0001$), testis ($p = 0.0007$), prostate ($p = <0.0001$), skin ($p = <0.0001$) and thyroid ($p = <0.0001$) cancerous tissue compared to normal non-cancerous tissues (Figure 5.11a). As with RNA expression analysis results from the EBI, cancerous pancreatic and colon cells were found to have higher expression of MORC4 RNA in the GENT2 database.

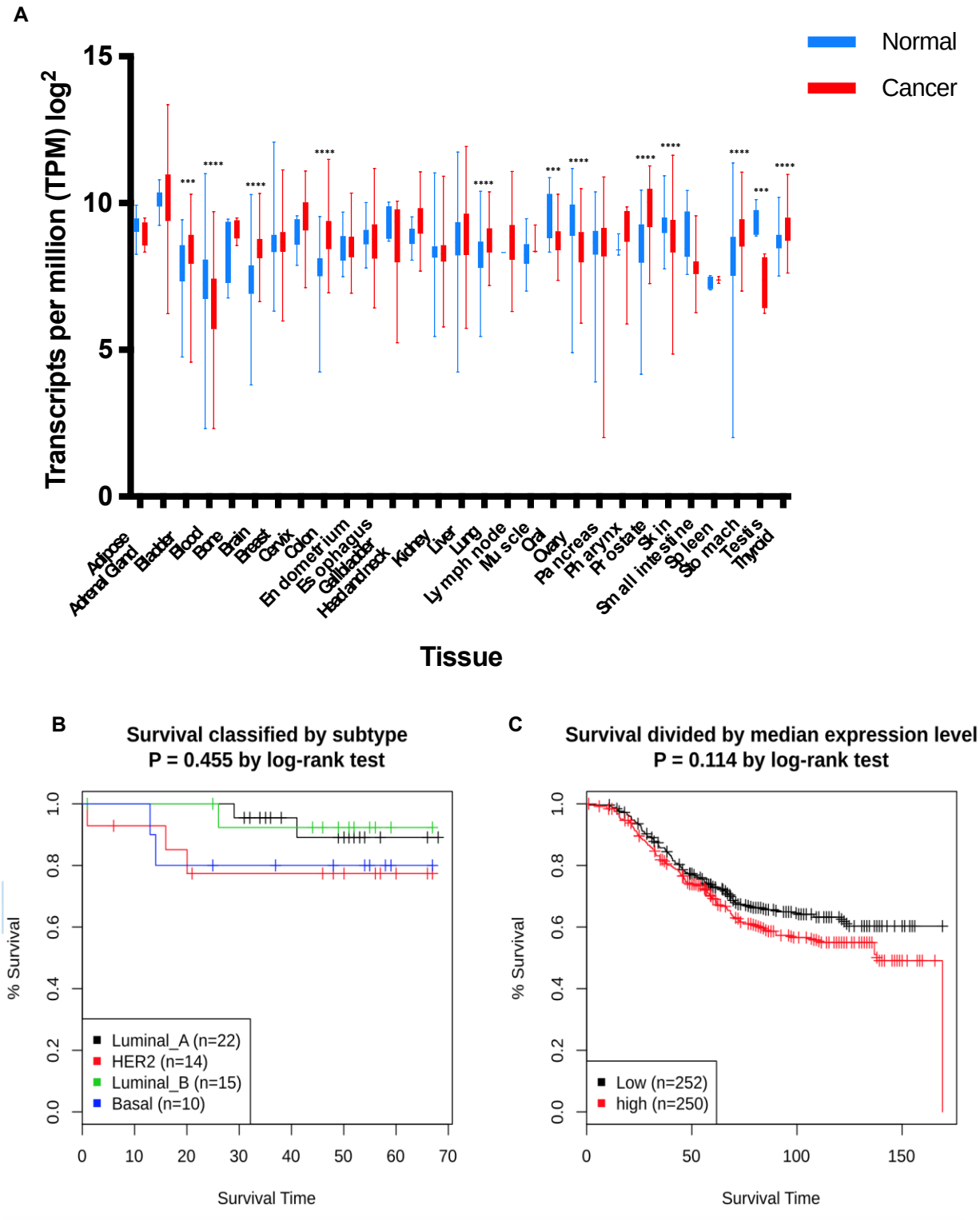


Figure 5.11: **MORC4 RNA expression dataset of normal non-cancerous tissue and cancerous tissue and breast adenocarcinoma survival analysis by type and expression level using the Gene Expression Database of Normal and Tumour Tissues 2 (GENT2).** MORC4 RNA expression of several non-cancerous and cancerous tissues (representative of n=256 (*** = $p \leq 0.001$ **** = $p \leq 0.0001$). Survival analysis of breast adenocarcinoma Kaplan-Meier plots **(B)** by subtype **(C)** by expression level of MORC4.

The GEPIA RNA expression database was used to investigate MORC4 RNA expression. GEPIA analysis revealed that there is higher expression of MORC4 RNA in several cancerous tissues compared to non-cancerous tissues (Figure 5.12a and b). Similar to the EBI and GENT2 RNA analysis, GEPIA also exhibited a higher expression of MORC4 RNA in colon adenocarcinoma (COAD) and pancreatic adenocarcinoma (PAAD) compared with normal non-cancerous tissue. In addition, GEPIA also found an increase of MORC4 expression in breast carcinoma (BRCA) and diffuse large B-cell lymphoma (DLBCL) (Figure 5.12b). A boxplot analysis of most common high expressing cancerous tissues revealed that DLBCL had significantly higher expression of MORC4 RNA compared to normal non-cancerous tissues (Figure 5.12c), similar to the data found in a previous study (Liggins *et al.*, 2007).

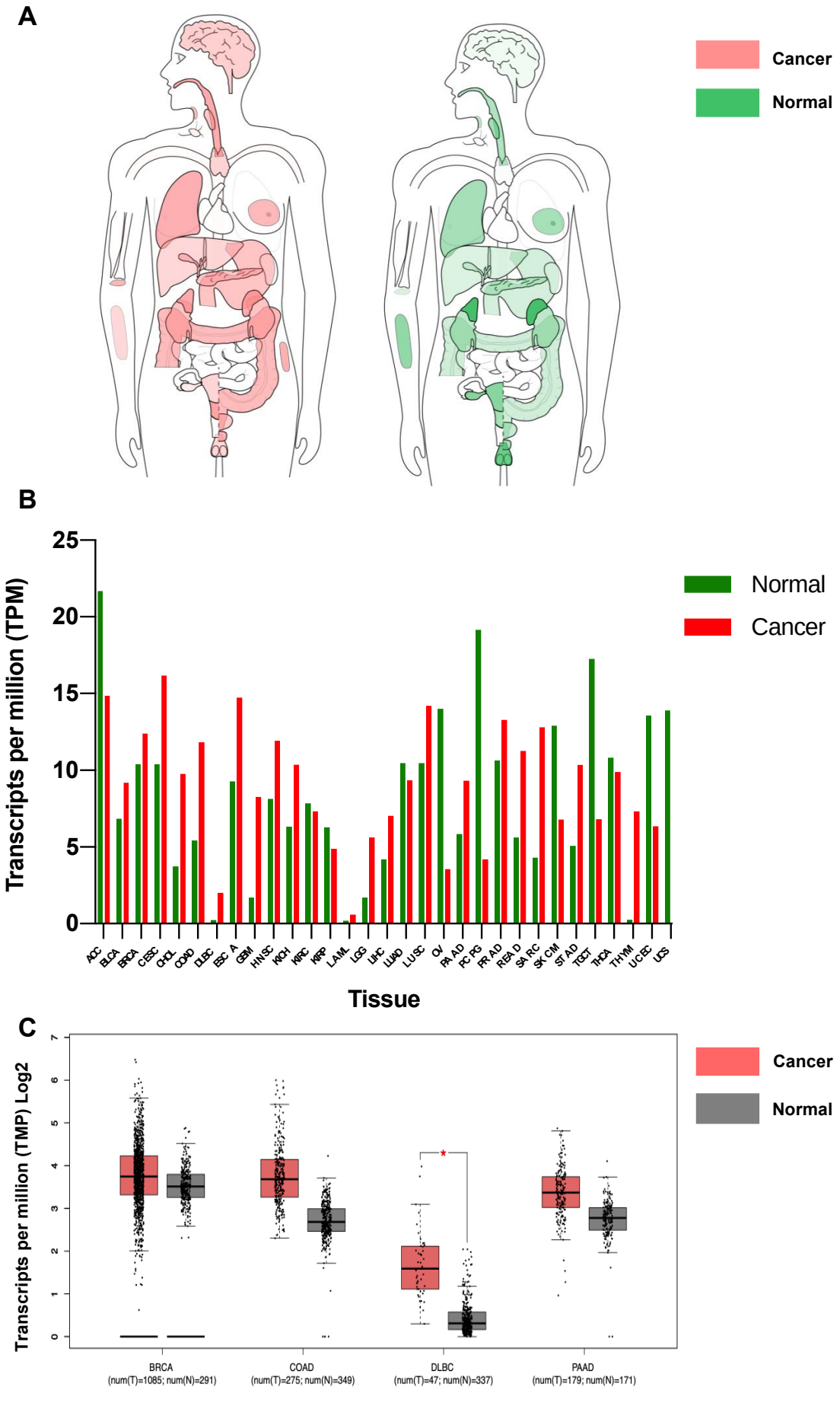

A
Cancer
Normal
B
Transcripts per million (TPM)
25
20
15
10
5
0
Normal
Cancer
ACC
BLCA
BRCA
CESC
CHOL
COAD
DLBC
ESCA
GBM
HNSC
KICH
KIRC
KIRP
LAML
LGG
LIHC
LUAD
LUSC
OV
PAAD
PCPG
PRAD
READ
SARC
SKCM
STAD
TGCT
THCA
THYM
UCEC
UCS
Tissue
C
Transcripts per million (TMP) Log2
Cancer
Normal
BRCA
(num(T)=1085; num(N)=291)
COAD
(num(T)=275; num(N)=349)
DLBC
(num(T)=47; num(N)=337)
PAAD
(num(T)=179; num(N)=171)

Figure 5.12: **MORC4 RNA expression dataset of normal non-cancerous tissue and cancerous tissue using the Gene Expression Profiling Interactive Analysis (GEPIA) database. (A)** Body map of MORC4 RNA expression in normal and cancerous tissue samples. **(B)** MORC4 RNA expression of several non-cancerous and cancerous tissues. ACC (adrenocortical carcinoma), BLCA (bladder carcinoma), BRCA (breast carcinoma), CESC (cervical carcinoma), CHOL (cholangio carcinoma), COAD (colon adenocarcinoma), DLBCL (diffuse large B-cell lymphoma), ESCA (esophageal carcinoma), GBM (glioblastoma multiforme), HNSC (head and neck carcinoma), KICH (kidney chromophobe), KIRC (kidney renal cell clear carcinoma), KIRP (kidney renal papillary cell carcinoma), LAML (acute myeloid leukaemia), LGG (brain lower grade glioma), LIHC (liver hepatocellular carcinoma), LUAD (lung adenocarcinoma), LUSC (lung carcinoma), MESO (mesothelioma), OV (ovarian carcinoma), PAAD (pancreatic adenocarcinoma), PCPG (pheochromocytoma and paraganglioma), PRAD (prostate adenocarcinoma), READ (rectum adenocarcinoma), SARC (sarcoma), SKCM (skin cutaneous melanoma), STAD (stomach adenocarcinoma), TGCT (testicular germ cell tumours), THCA (thyroid carcinoma), THYM (thymoma), UCEC (uterine carcinoma), UCS (uterine carcinosarcoma). **(C)** Boxplot of RNA expression of MORC4 in some non-cancerous and cancerous tissues (* = $p \leq 0.05$).

5.2.3.2 Survival and prognosis associations with MORC4 RNA expression

RNA expression databases can also be utilised to investigate the survival and prognosis of patients with cancer and high or low expression of specific genes. Survival analysis of pancreatic adenocarcinoma revealed that patients with high RNA expression of MORC4 had a high probability of overall survival compared to patients with low expression of MORC4 (Figure 5.13a). Specifically, PROGgeneV2 analysis (Goswami and Nakshatri, 2013) indicated that high expression of MORC4 increased overall probability of survival after approximately 600 days (1.6 years). After 700 days, patients with high expression of MORC4 had a 40% chance of survival compared to patients with low expression of MORC4, who only had a 20% chance of survival. Similar to PROGgeneV2, the Kaplan-Meier plotter database (Jia *et al.*, 2019) also demonstrated a higher probability of survival in pancreatic adenocarcinoma with a high expression level of MORC4. The Kaplan-Meier plotter, which is a survival estimator tool, also indicated that the average survival time with high expression of MORC4 was 24 months, whereas with low expression it was 19 months.

Survival analysis of breast adenocarcinoma using GENT2 predicted that the HER2 breast adenocarcinoma subtype had the lowest probability of survival following expression of MORC4 (Figure 5.13b). GENT2 survival analysis also suggested that high expression of MORC4 leads to a lower probability of survival (poorer prognosis). Both PROGgeneV2 and Kaplan-Meier plotter survival analysis also suggested similar results to GENT2 (Figure 5.15b), with PROGgeneV2 suggesting that high expression of MORC4 gives a 60% chance of survival after 3 years compared to low expression of MORC4, which shows an 80% chance of survival (Figure 5.13b). Survival analysis of colon adenocarcinoma also predicted that higher expression for MORC4 also leads to a lower probability of survival compared to lower expression of MORC4, in both PROGgeneV2 and GENT2 analysis (Figure 5.13c). PROGgeneV2 analysis revealed that there was only a 50% chance of survival after 5 years with high expression of MORC4 compared to low expression of MORC4, which had a 70% chance of survival after 5 years (Figure 5.13c).

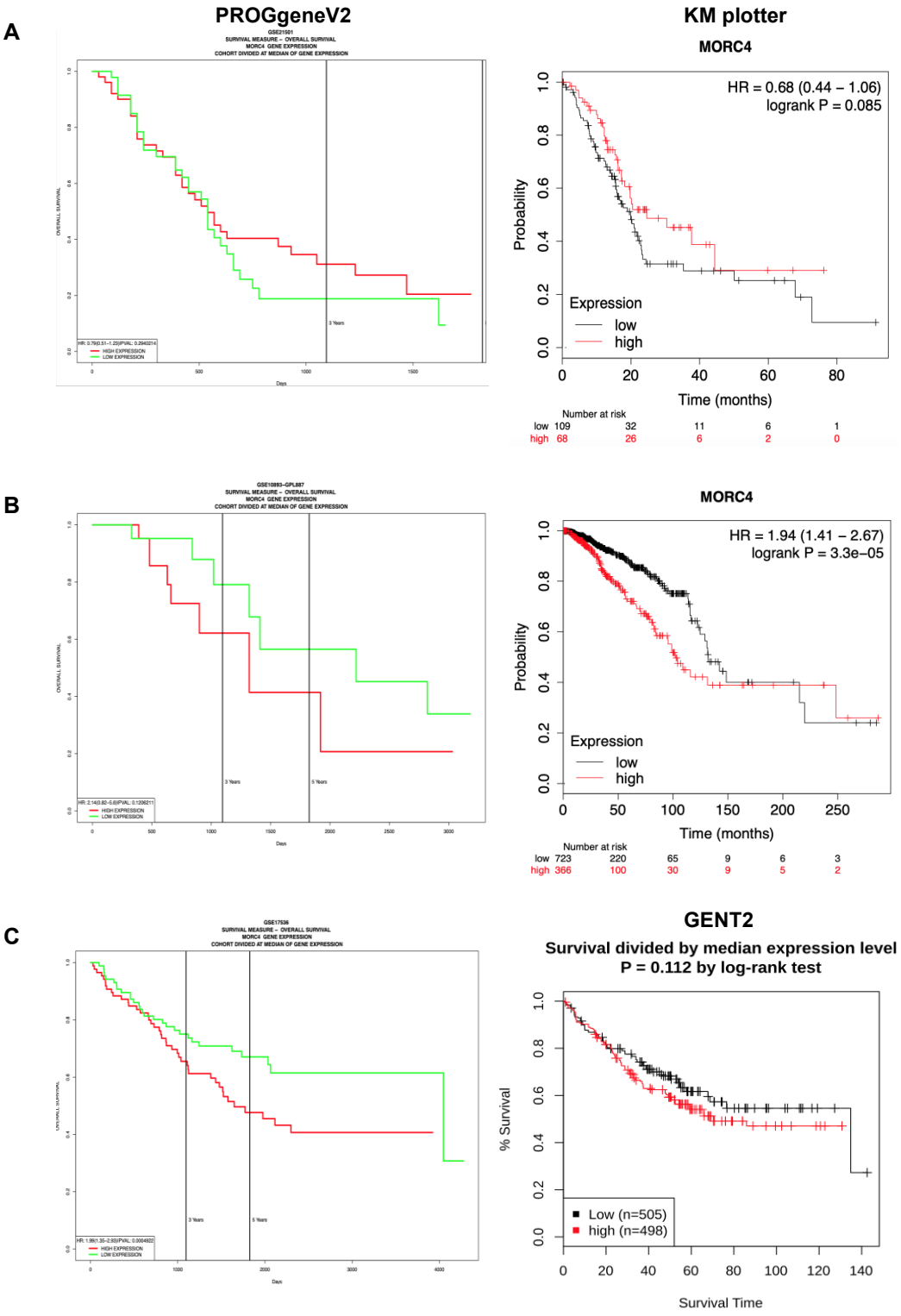


Figure 5.13: **MORC4 RNA expression survival analysis using Kaplan-Meier plots in pancreatic adenocarcinoma, breast adenocarcinoma and colon adenocarcinoma using several RNA expression databases. (A)** pancreatic adenocarcinoma survival analysis **(B)** breast adenocarcinoma survival analysis **(C)** colon adenocarcinoma survival analysis. All KM plots were created using the PROGgeneV2, the KM plotter or GENT2 databases.

5.3 MORC4 protein analysis

5.3.1 Cloning of full-length epitope-tagged *MORC* expression fragments

Ligation independent cloning (LIC) was utilised to create recombinant *MORC4* in mammalian expression vectors, with the addition of either FLAG (Einhauer and Jungbauer, 2001) or human influenza hemagglutinin (HA) polypeptide epitope tags (Kimple *et al.*, 2013). Cloning of FLAG-tagged and HA-tagged full-length *MORC* paralogues was also required for subsequent cell biology analysis, and verification of the specificity of MORC4 monoclonal antibodies (gift from Alison Banham, University of Oxford), following similarities between the paralogous sequences. If no bands were identified following western blotting of MORC4, MORC paralogues could be tested to ensure successful transfection of FLAG/HA-tagged paralogues using epitope antibodies. However, the MORC4 monoclonal antibodies were important for analysis of MORC4 expression in clinical lymphoma and other cancerous disease samples.

5.3.1.1 Polymerase chain reaction (PCR) amplification of full-length *MORC*

Full-length *MORC* DNA (*MORC1-4*) were cloned into mammalian expression vectors that included either pcDNA3-N-FLAG-LIC and pcDNA3-N-HA-LIC (Appendix 8.22). Full-length *MORC* (MORC1-4) were provided by the I.M.A.G.E consortium (Source Bioscience). Prior to cloning, each *MORC* template DNA was sequenced fully to confirm their nucleotide sequences.

5.3.1.1.1 Optimisation of *MORC4* PCR amplification

Full-length *MORC* open reading frames (Table 5.5) were amplified using polymerase chain reaction (PCR) and full-length *MORC* primers (Table 5.6). *MORC4* was amplified with a mix of high-fidelity Phusion® polymerase with standard MyTaq™ polymerase and either Phusion buffer or MyTaq™ red buffer. A 1:1 mix of Phusion® polymerase and MyTaq™ with MyTaq™ red buffer was necessary to amplify and to achieve a good yield of high fidelity

MORC4 PCR product (Figure 5.14). However, *MORC1*, *MORC2* and *MORC3* were successfully amplified using Phusion® buffer and Phusion® DNA polymerase only (Figure 5.14).

Table 5.5: **Full-length mammalian plasmid cloned fragments of *MORC1-4*** including the size of each clone, PCR products (bp) and protein size (kDa).

Residue boundary (aa)	Code	Primer F	Primer R	Length (bp)	Vector	PCR screen expected size (bp)	kDa
MORC1 (1-984)	NN057	MORC1f	MORC1r	2952	pcDNA3-N-FLAG-LIC	3237	114
MORC2 (24-1032)	NN058	MORC2f	MORC2r	3027	pcDNA3-N-FLAG-LIC	3312	117
MORC3 (1-938)	NN060	MORC3f	MORC3r	2814	pcDNA3-N-FLAG-LIC	3099	109
MORC4 (1-900)	NN061	MORC4f	MORC4r	2700	pcDNA3-N-FLAG-LIC	2985	104
MORC1 (1-984)	NN067	MORC1f	MORC1r	2952	pcDNA3-N-HA-LIC	3237	114
MORC2 (24-1032)	NN068	MORC2f	MORC2r	3027	pcDNA3-N-HA-LIC	3312	117
MORC3 (1-938)	NN070	MORC3f	MORC3r	2814	pcDNA3-N-HA-LIC	3099	109
MORC4 (1-900)	NN071	MORC4f	MORC4r	2700	pcDNA3-N-HA-LIC	2985	104

***Table 5.6: MORC* primers designed to created full-length MORCs (MORC1-4).**

Primer name	Primer sequence (5'-3')
MORC1f	TACTTCCAATCCATGGACGACAGGTACCCTGC
MORC1r	TATCCACCTTTACTGTCAATTTTCCGAAGTCTTTTC
MORC2f	TACTTCCAATCCATGACCACTCACGAATTCTTGTTTGG
MORC2r	TATCCACCTTTACTGTCAGTCCCCCTTGGTGATGAGGT
MORC3f	TACTTCCAATCCATGGCGGCGCAGCCACCCCGC
MORC3r	TATCCACCTTTACTGTCAAGTACTACTGATTTCAC
MORC4f	TACTTCCAATCCATGCTCCTGTACCGAGGGGC
MORC4r	TATCCACCTTTACTGTCACGAAGCTCCAAGTGAGG

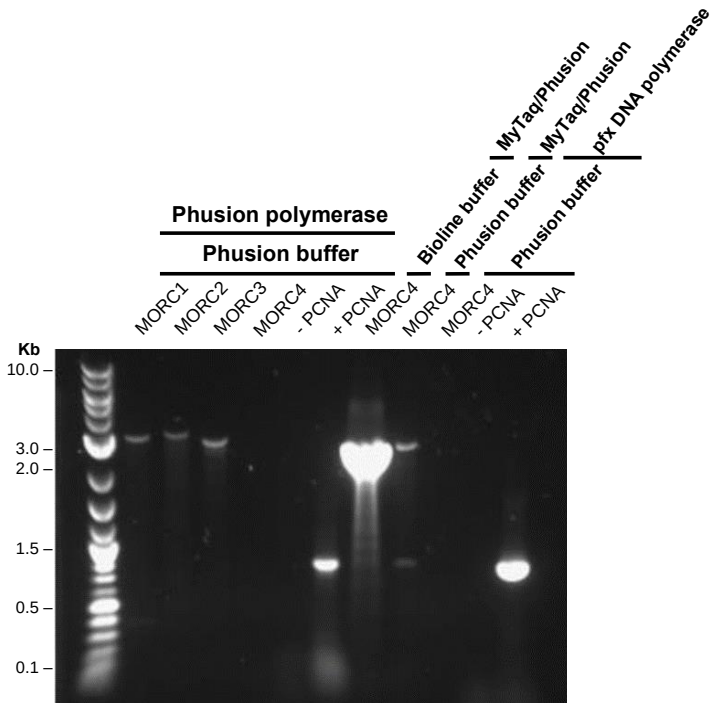


***Figure 5.14:* Amplification of full length *MORC* open reading frames (ORFs) for mammalian expression.** Full-length *MORC* open reading frames were amplified using PCR and were resolved on a 0.8% agarose gel. 1. *MORC1* 2. *MORC2* 3. *MORC3* 4. *MORC4* 5. negative control (PCNA primers) 6. PCNA positive control. Samples 1-6 were amplified using Phusion buffer and Phusion DNA polymerase. 7. *MORC4* (Bioline red buffer and a mix of MyTaq and Phusion DNA polymerase) 8. *MORC4* (Phusion buffer and a mix of MyTaq and Phusion DNA polymerase) 9. *MORC4* 10. PCNA negative control 11. PCNA positive control. Samples 9-11 were amplified using Phusion buffer and pfx DNA polymerase.

5.3.1.2 Cloning and PCR screen of full-length epitope-tagged *MORC* sequences

Following PCR amplification of full-length *MORC* sequences, PCR inserts were purified, and cohesive ends were produced by T4 DNA polymerase treatment (Savitsky *et al.*, 2010). Insert cohesive ends were ligated with vector (pcDNA3-N-HA-LIC or pcDNA3-N-FLAG-LIC) cohesive ends. Ligated products were transformed into MACH1 *E. coli* cells and a colony PCR screen, which utilised common primers binding to sites surrounding the cloning site (similar to *MORC4 E. coli* expression clones, (Chapter 3), allowing confirmation of the sizes of cloned full-length *MORC* sequences (Figure 5.15).

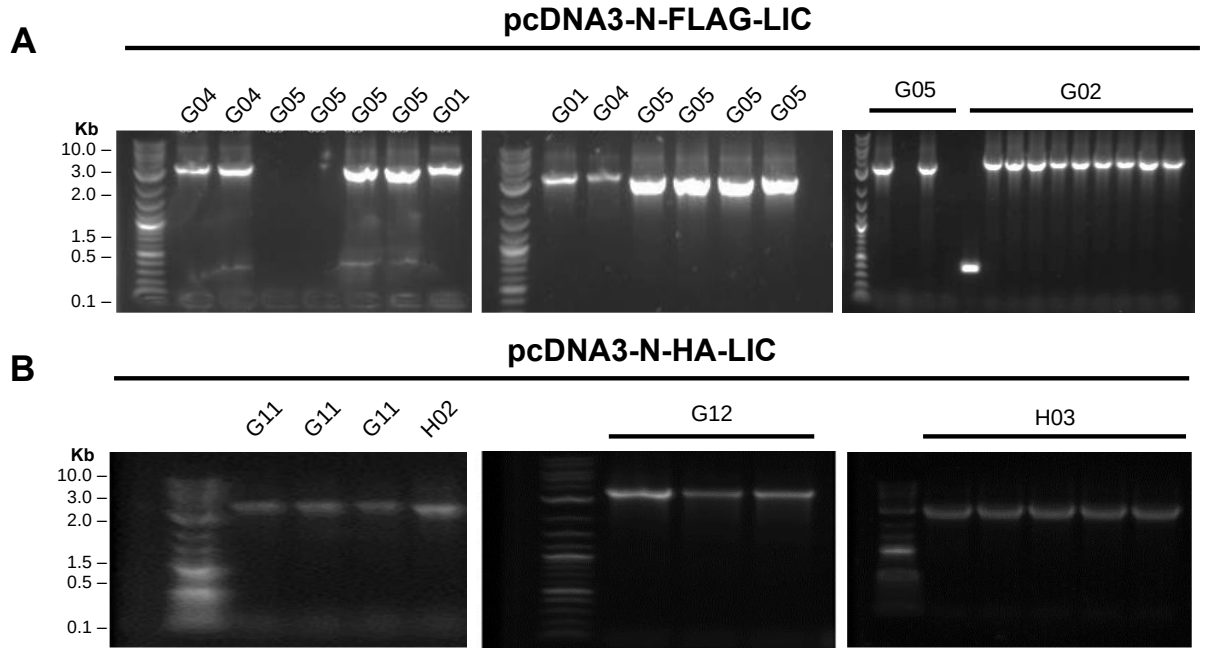


Figure 5.15: **Colony screen of full-length FLAG-tagged and HA-tagged *MORCs* in mammalian expression vectors.** Cloned MORC PCR products were screened on a 0.8% agarose gel to confirm sizes. **(A)** Includes full length MORC1, MORC2, MORC3 and MORC4 cloned into pcDNA3-N-FLAG-LIC. **(B)** Includes full length MORC1, MORC2, MORC3 and MORC4 cloned into pcDNA3-N-HA-LIC.

All screened clones of both FLAG-tagged and HA-tagged *MORCs* were of the correct size (Table 5.5 and figure 5.15), suggesting that *MORCs* were successfully cloned into MACH1 *E. coli* cells. To verify cloning boundaries, one successful clone from each sample from the colony PCR screen was sequenced, except for *MORC4*, where several successful clones had to be

sequenced to verify that there were no deletions or point mutations. It was especially important for full-length *MORC4* clones to be sequenced because of the difficulties with producing clones with a mixture of both MyTaq™ and Phusion® polymerase. Cloned full-length *MORC* truncated fragments were then used for subsequent transfections in mammalian cells.

5.3.2 Transfections of full-length *MORCs* into mammalian cell lines

Full-length FLAG-tagged and HA-tagged *MORC* plasmids were transfected into several mammalian cells lines for protein expression to aid the verification of monoclonal MORC4 antibodies. Transfecting all recombinant *MORC* plasmids into several cell lines was intended in an attempt to validate any observed MORC4 antibody binding specific to MORC4 and not the three closely related paralogues (MORC1, MORC2 and MORC3). In addition, the antibody verification required western blotting to ensure that any detected protein bands were MORC4 and not its paralogues. However, paralogues of MORC4 were manipulated with the addition of an epitope tag (FLAG-tagged or HA-tagged) to ensure detection if no MORC4 expression was identified. If no signal was observed for MORC4 with expression of MORC4 paralogues, this would confirm that the antibody was not binding to the paralogues, leading to the verification of MORC4 monoclonal antibodies.

To begin antibody characterisation, full-length *MORC* open reading frames (ORF) were cloned into either pcDNA3-N-Flag-LIC or pcDNA3-N-HA-LIC, mammalian expression vectors (Figure 5.14 and 5.15, section 5.3.1). FLAG-tagged or HA-tagged full-length *MORCs* were transfected into human embryonic kidney (HEK)293FT, human bone osteosarcoma (U2OS), cervical carcinoma (HeLa) and monkey kidney COS-1 cell lines. β-actin, a crucial component of the cytoskeleton of cells (Bunnell *et al.*, 2011), was used as a loading control to ensure protein loading was equivalent throughout each gel and across samples. The highly expressing chromodomain helicase DNA binding protein 1 Like (CHD1L), a DNA repair protein (Ahel *et*

al., 2009) was fused to a FLAG or HA epitope tag, which was used as a transfection and western blot positive control for all cell lines throughout transfections.

5.3.2.1 FLAG-tagged and HA-tagged MORC1 and MORC2 were expressed in HEK293FT cells

Following transfections of FLAG-tagged and HA-tagged full-length MORCs in HEK239FT and U2OS cell lines, FLAG-tagged and HA-tagged MORC1 and MORC2 were successfully expressed in HEK239FT cells (Figure 5.16a and c). The expected size of FLAG and HA-tagged MORC1 and MORC2 was 114 kDa (calculated using the Swiss Institute of bioinformatics (SIB) resource portal, ProtParam) and 117 kDa, respectively. Fluorescent detection western blotting determined expression of MORC1 and MORC2 at approximately 100 kDa, with MORC2 expression slightly larger, indicating successful expression. The results also showed that MORC1 and MORC2 had a stronger expression signal when a higher concentration of lipofectamine was used in addition to 2.5µg of DNA (Figure 5.16a and c). To determine a successful transfection, CHD1L was used as a positive control and was expressed at approximately 100 kDa (Figure 5.16). Housekeeping protein and loading control, β-actin (Ferguson *et al.*, 2005), was detected at approximately 42 kDa in every sample, suggesting a consistent and comparable level of protein in each sample, allowing the comparison of expression levels of transfected proteins. Contrary to MORC1 and MORC2, expression of FLAG and HA-tagged MORC3 and MORC4 was unsuccessful (Figure 5.16b and d) as they were expected to express at 109 kDa and 104 kDa, respectively, but were not observed on western blots.

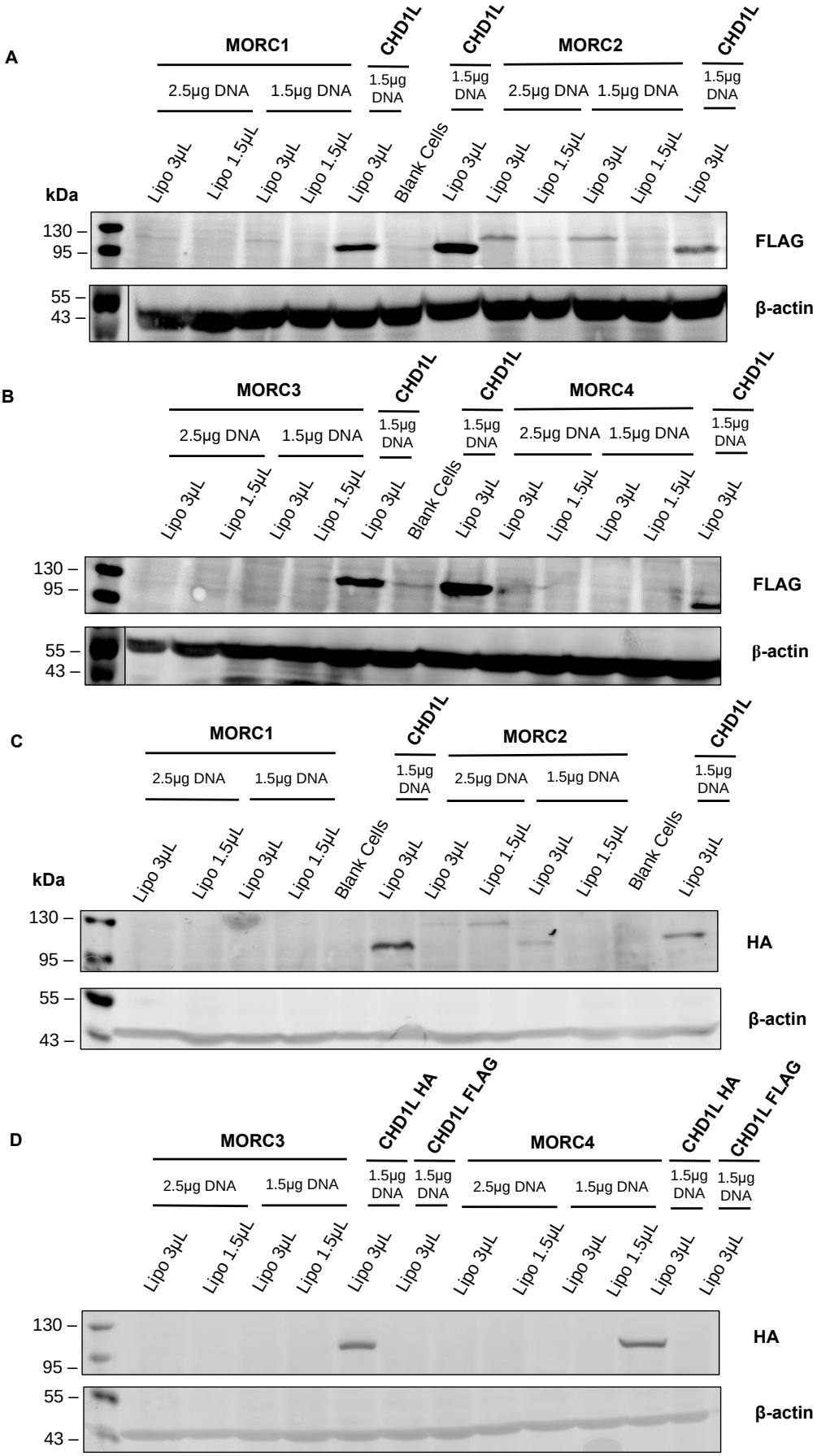

A
MORC1
CHD1L
CHD1L
MORC2
CHD1L
2.5μg DNA
1.5μg DNA
1.5μg DNA
2.5μg DNA
1.5μg DNA
1.5μg DNA
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Blank Cells
Lipo 3μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
kDa
130 –
95 –
FLAG
55 –
43 –
β-actin
B
MORC3
CHD1L
CHD1L
MORC4
CHD1L
2.5μg DNA
1.5μg DNA
1.5μg DNA
2.5μg DNA
1.5μg DNA
1.5μg DNA
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Blank Cells
Lipo 3μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
130 –
95 –
FLAG
55 –
43 –
β-actin
C
MORC1
CHD1L
MORC2
CHD1L
2.5μg DNA
1.5μg DNA
1.5μg DNA
2.5μg DNA
1.5μg DNA
1.5μg DNA
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Blank Cells
Lipo 3μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Blank Cells
Lipo 3μL
kDa
130 –
95 –
HA
55 –
43 –
β-actin
D
MORC3
CHD1L HA
CHD1L FLAG
MORC4
CHD1L HA
CHD1L FLAG
2.5μg DNA
1.5μg DNA
1.5μg DNA
1.5μg DNA
2.5μg DNA
1.5μg DNA
1.5μg DNA
1.5μg DNA
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 3μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 1.5μL
Lipo 3μL
Lipo 3μL
130 –
95 –
HA
55 –
43 –
β-actin

Figure 5.16: **Fluorescence detection western blotting of transfected full length MORCs into HEK293FT cell line**. HEK293FT cells were transfected with full length MORC plasmid DNA, harvested, sonicated and cell lysates were resolved on a 10% SDS-PAGE gel for fluorescence detection western blotting to determine protein expression. Membranes were probed with FLAG and β-actin (1:20,000 dilutions) primary antibodies in 5% milk and TBS-T. (A) FLAG-tagged MORC1 and MORC2. (B) FLAG-tagged MORC3 and MORC4. (C) HA-tagged MORC1 and MORC2. (D) HA-tagged MORC3 and MORC4. Membranes were detected using a Li-COR Odyssey imagining system. β-actin loading control was imaged using the green channel (800nm) or red channel (700nm) and the marker, FLAG and HA were imaged using the red channel (700nm).

5.3.2.2 FLAG-tagged and HA-tagged MORCs were not expressed in U2OS, HeLa or COS-1 cells

U2OS transfection of all MORCs was also unsuccessful (Appendix 8.23a and b). In an attempt to express all MORC proteins, the competent and commonly used mammalian cell line, HeLa (Asgharian *et al.*, 2014), was used for transfections, which resulted in no expression of any MORC protein (Appendix 8.24a and b). Furthermore, the COS-1 monkey kidney cell line (Mortlock *et al.*, 1993) was also used for transfections, however, they also failed to express full-length FLAG-tagged or HA-tagged MORCs (Appendix 8.25a-d). For each cell line which failed to express MORCs, the CHD1L and actin positive and loading controls were both detected, indicating that cells were present in each sample and the transfection was performed correctly. Although MORC1 and MORC2 proteins were expressed in HEK293FT cells, the paralogues could not be screened using this method as both MORC3 and MORC4 proteins could not be expressed, indicating that perhaps antibodies could be screening using paralogue recombinant protein instead.

5.3.3 Analysis of MORC4 protein expression in human cancer cell lines

5.3.3.1 Detection of MORC4 using monoclonal antibodies

In order to identify expression of MORC4 in a number of cells, MORC4 monoclonal antibodies were initially utilised. There were two clones of MORC4 monoclonal antibodies used throughout this study, 161A/3 and MORC4 97F/9 (gift from Professor Alison Banham,

University of Oxford). The epitope for these antibodies was designed from the C-terminal region of MORC4 (Figure 5.17) by the Banham laboratory (C. Cooper, *pers. comm.*). A multiple sequence alignment (MSA) of all human MORC proteins demonstrates that the epitope sequence was specific to MORC4 (Figure 5.17) and not the other paralogues. In addition, MORC4 monoclonal antibodies were also used to compare against the commercial rabbit polyclonal antibody used in a previous study to investigate MORC4 protein expression in breast cancer (Yang *et al.*, 2019). Furthermore, the MORC4 polyclonal antibody was also utilised for comparison to other MORC proteins using polyclonal antibodies of MORC1, MORC2 and MORC3.

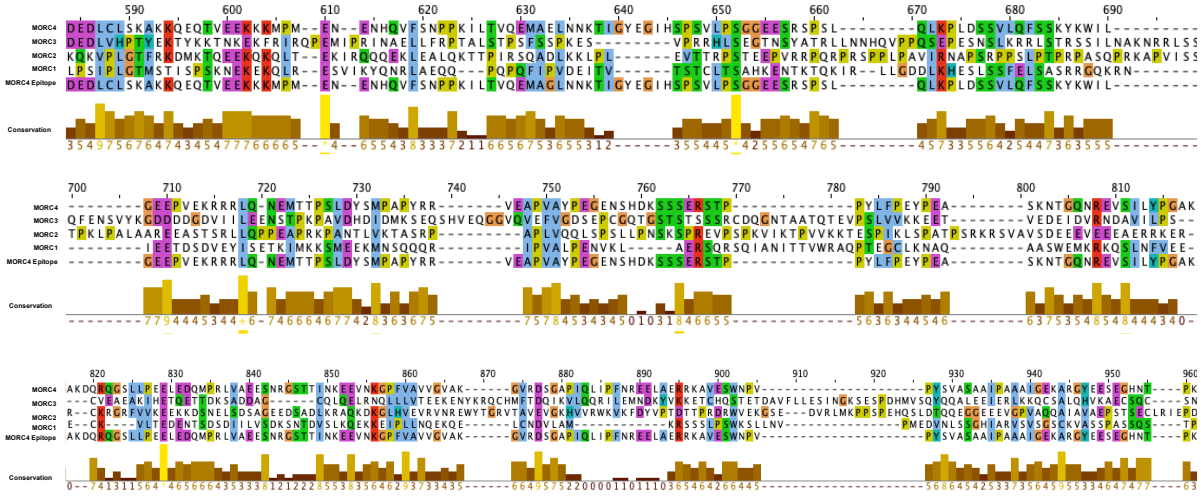


Figure 5.17: **Multiple sequence alignment of protein sequences of all human MORCs (MORC1-4) and the MORC4 epitope used to design monoclonal antibodies.** Human MORC protein sequences were used for a MUSCLE multiple sequence alignment using Jalview software. The conservation score across all MORCs was shown below the alignment (Yellow is the most conserved and brown or no colour is the least conserved). The multiple sequence alignment residues were coloured using ClustalX colours.

5.3.3.1.1 Optimisation of MORC4 monoclonal antibodies

Prior to use on several cancerous cell lines, optimisation of the antibodies was required to determine suitable conditions and concentrations of monoclonal antibodies for subsequent use in western blotting, including different methods of detection.

5.3.3.1.1.1 Western blotting of MORC4 using fluorescent detection

5.3.3.1.1.1.1 Optimising concentration of MORC4 antibodies

MORC4 monoclonal antibodies were initially optimised using fluorescence detection western blotting and using pure recombinant protein that contained the epitope used for the monoclonal antibody generation, $MORC4^{420\text{-}757}$ (Chapter 3, Section 3.2.2). To determine an optimal concentration of the MORC4 monoclonal antibodies, an initial titration experiment involved testing different amounts of blotted $MORC4^{420\text{-}757}$ with a range of antibody concentrations including 1:50, 1:500 and 1:5000, which were each prepared using 5% milk and TBS-T. Both antibodies, MORC4 161A/3 and MORC4 97F/9, detected pure $MORC4^{420\text{-}757}$ at ~40 kDa (expected size was 38 kDa) and several smaller protein bands, likely to result from protein degradation (Figure 5.18a and b). In addition, MORC4 97F/9 appeared to have a stronger signal (Figure 5.18b) than MORC4 161A/3 (Figure 5.18a), which was determined by the intensity of protein bands. Furthermore, the 97F/9 monoclonal antibody could detect MORC4 protein as little as 10 ng, suggesting that it was highly sensitive.

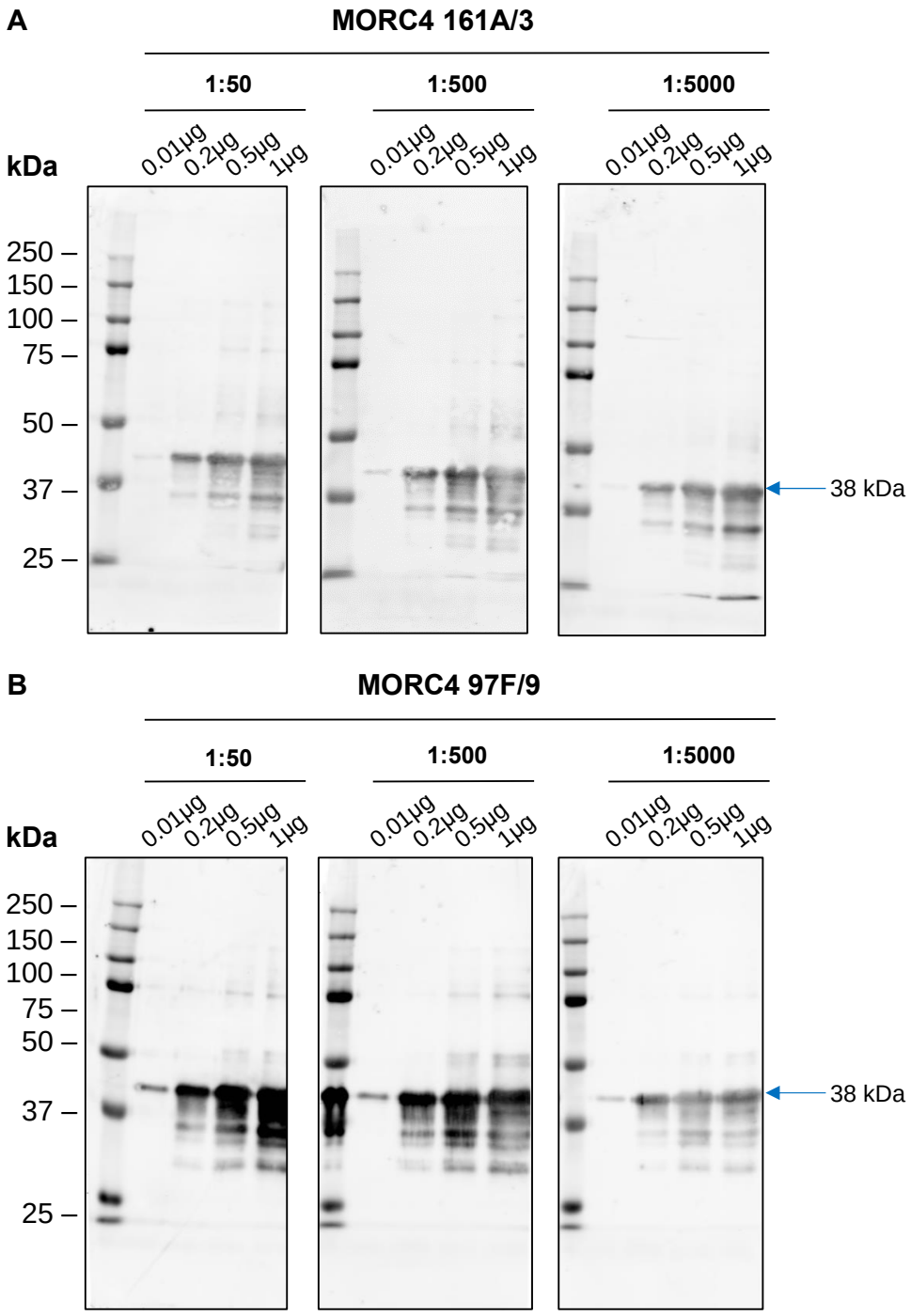


Figure 5.18: **Optimisation of fluorescence detection western blotting of MORC4 monoclonal antibodies and MORC4$^{420\text{-}757}$.** MORC4$^{420\text{-}757}$ was loaded onto an SDS-PAGE gel in varying concentrations of protein and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with (A) MORC4 161A/3 and (B) MORC4 97F/9, which were prepared in 5% milk and TBS-T. Membranes were detected using a Li-COR Odyssey imagining system and all membranes were imaged using the red channel (700nm). Blue arrows indicate the expected size of MORC4$^{420\text{-}757}$.

5.3.3.1.1.1.2 Determination of optimal antibody diluent and incubation conditions

To determine an optimal antibody diluent and incubation conditions, pure recombinant MORC4$^{420\text{-}757}$ protein was utilised for a similar titration experiment to the one involving the optimisation of the concentration of monoclonal antibodies. Monoclonal antibodies were prepared at a 1:500 dilution in both 5% milk and 5% bovine serum albumin (BSA) and TBS-T and were incubated at both 4°C for 12 hours or 22°C for 2 hours (Figure 5.19). 1µg of BSA was loaded onto each gel to ensure specificity of MORC4 antibodies and that any detected signal was protein-specific. The antibodies did not detect BSA in any blot, indicating that the MORC4 antibodies specifically recognised MORC4. MORC4$^{420\text{-}757}$ and BSA samples were also stained with Coomassie dye to determine the protein loaded onto the blots. The Coomassie stained SDS-PAGE gel result determined that there was protein present in each well including the control BSA sample (Figure 5.20), which was present at approximately 67 kDa. Overall, the results indicated that the MORC4 monoclonal antibodies produced optimal results when utilised in 5% milk at 4°C for 12 hours (Figure 5.19a), although there was not a great deal of difference between the tested conditions.

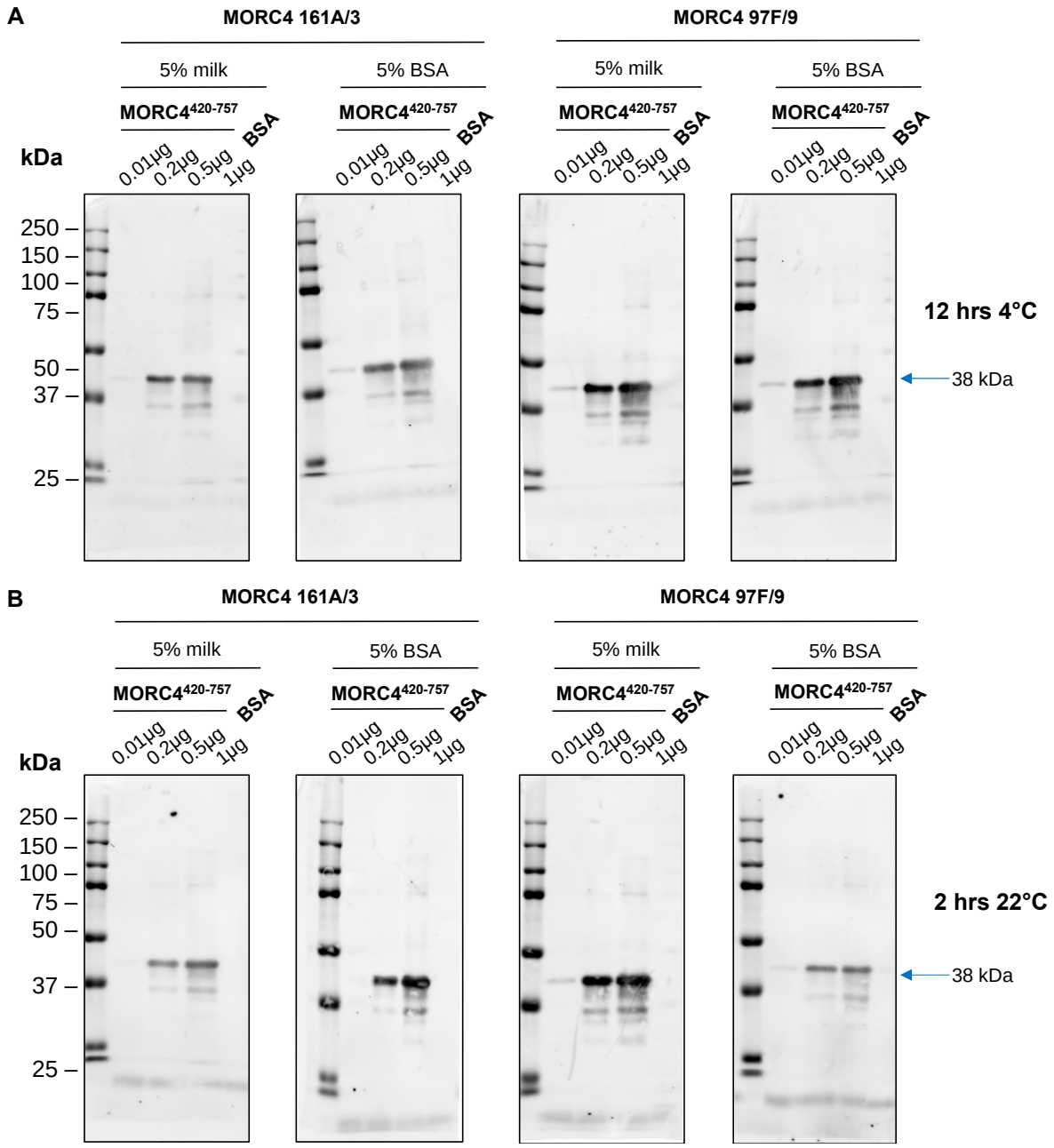


Figure 5.19: **Optimisation of fluorescence detection western blotting conditions of MORC4 monoclonal antibodies with MORC4$^{420-757}$ and bovine serine albumin (BSA).** MORC4$^{420-757}$ was loaded onto an SDS-PAGE gel in varying concentrations of protein and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with MORC4 161A/3 and MORC4 97F/9, which were prepared in a 1:500 dilution in 5% milk or 5% BSA and TBS-T. Antibodies were incubated at either (A) 4°C for 12 hours or (B) 22°C for 2 hours. Membranes were detected using a Li-COR Odyssey imagining system and all membranes were imaged using the red channel (700nm). Blue arrows indicate the expected size of MORC4$^{420-757}$.

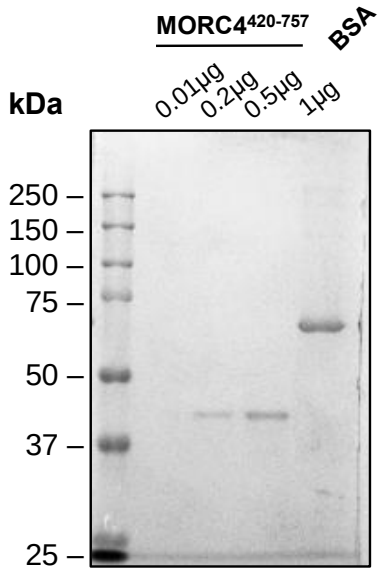


Figure 5.20: **Coomassie stained SDS-PAGE gel of MORC4$^{420-757}$ and bovine serum albumin (BSA).** MORC4$^{420-757}$ and BSA was resolved on an SDS-PAGE gel in varying concentrations of protein and stained with Coomassie dye for protein determination.

5.3.3.1.1.2 Western blotting of MORC4 using Enhanced chemiluminescence (ECL) detection

5.3.3.1.1.2.1 Determination of optimal antibody diluent and incubation conditions

In addition to western blotting detection by fluorescence, enhanced chemiluminescence (ECL) western blotting was used to optimise western blotting methods for using MORC4 monoclonal antibodies. ECL exposure on an X-ray film following secondary antibody incubation was approximately 1 hour, suggesting that low level of protein was detected. It also indicates that to decrease exposure time either a higher concentration of protein or antibody was required to detect MORC4, considering microgram levels of MORC4 protein were present on western blots. ECL detection was less sensitive than fluorescent western blotting, which was determined by the low signal and smaller protein bands that were visible following ECL (Figure 5.21). However, the most optimal conditions found when using ECL western blotting were with antibodies prepared with 5% milk and incubated for 12 hours at 4°C (Figure 5.21).

Overall, these results suggest that fluorescent detection western blotting was optimal for detection using MORC4 monoclonal antibodies.

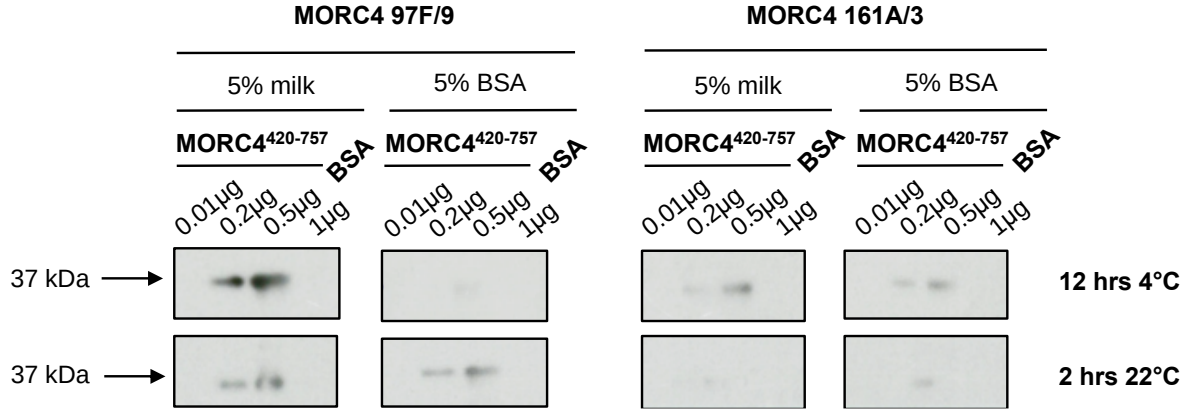


Figure 5.21: **Optimisation of enhanced chemiluminescence (ECL) western blotting conditions of MORC4 monoclonal antibodies with MORC4$^{420\text{-}757}$ and bovine serine albumin (BSA).** MORC4$^{420\text{-}757}$ was loaded onto an SDS-PAGE gel in varying concentrations of protein and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with MORC4 161A/3 and MORC4 97F/9, which were prepared in a 1:500 dilution in 5% milk or 5% BSA and TBS-T. Antibodies were incubated at either 4°C for 12 hours or 22°C for 2 hours. Membranes were detected using a BM ECL blotting substrate (POD) kit and exposed in a dark room for 1 hour, until protein bands appeared.

5.3.3.1.2 Detection of MORC protein expression in cancer cell lines

To investigate expression patterns of MORC proteins across several cancer types and to identify if MORC4 protein expression followed similar patterns seen at RNA level (Liggins *et al.*, 2007), several human cancer cells lines were utilised to examine protein expression of MORC proteins using western blotting. Expression patterns of MORC proteins were examined over an array of cancerous (Appendix 8.3) cell lines to identify any similarities or differences in expression of MORC proteins, in particular, MORC4. Identifying expression of MORC4 can give an insight into MORC4 function and role in disease.

5.3.3.1.2.1 Bicinchoninic acid (BCA) assay determined protein concentration

Following growth and maintenance of several cancer cell lines, cells were harvested and sonicated to extract the protein for analysis. Prior to protein analysis using western blotting, a

bicinchoninic acid (BCA) assay kit was used to quantify protein in order to calculate a specific concentration of protein to load onto an SDS-PAGE gel, to allow equivalence across cell lines, so MORC protein levels could be accurately compared. Several bovine serum albumin (BSA) controls of known concentration were prepared, and standard curves were created (Figure 5.22a-c) to aid determination of protein concentrations. Following the creation of a standard curve the equation of a straight line ($y = mx + c$) was used to calculate the protein concentration of each cell line sample, towards equivalence of protein loading between cell lines.

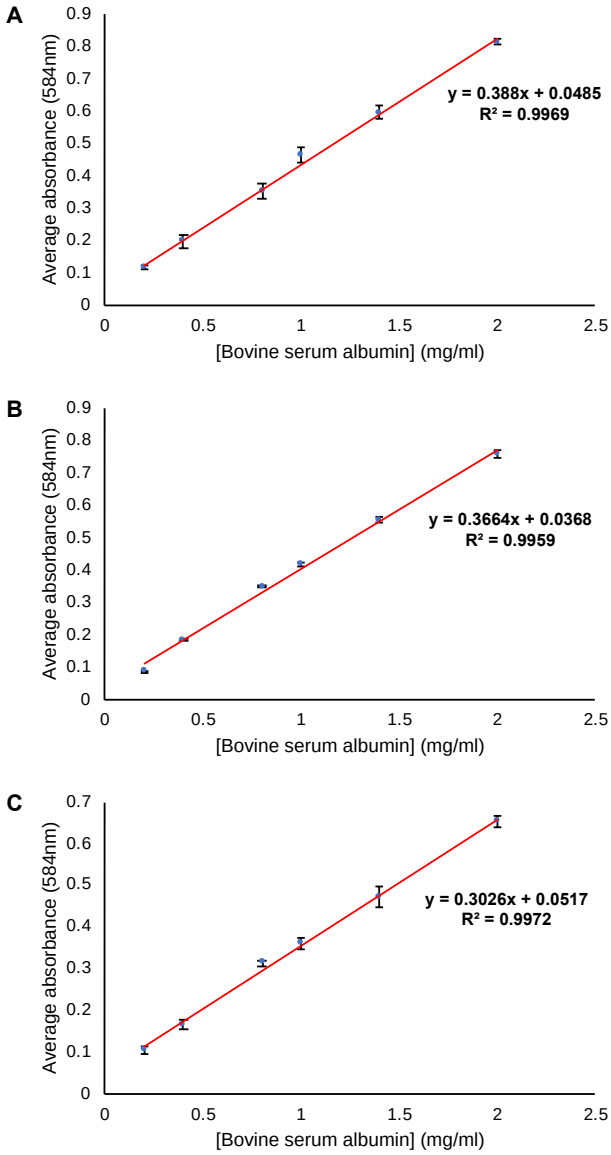


Figure 5.22: **Bicinchoninic acid (BCA) assay standards.** Cells were harvested, sonicated and a BCA assay kit was used to determine protein concentration. Several BSA standards were prepared and used to create a standard curve, which was used to calculate protein concentration. (A) standards for 1 in 4 protein dilutions of samples OCI-Ly3 to HEK293FT (1-25). (B) standards for 1 in 4 protein dilutions of samples SK-N-BE(2c) to MDA-MB-468 (26-38). (C) standards for 1 in 6 protein dilutions of samples SK-N-BE(2c), Kelly, SH-SY5Y, SK-A-NS and Mia-Pa-Ca-2. Each sample was prepared in triplicate.

Several cancer cells were utilised to investigate MORC4 protein expression using monoclonal antibodies. Detection of MORC4 protein expression was required and used to compare to qRT-PCR mRNA expression. Therefore, identical cell lines were used in addition to the same

passage number across multiple batches to ensure comparisons between RNA and protein were equivalent. Prior to the MORC4 protein analysis of an array of cancer cells, lymphoma cell lines were used for an initial experiment to determine suitable conditions required for fluorescent western blotting. MORC4 monoclonal antibodies were diluted 1:50 in 5% milk and membranes were probed for either 22°C for 3 hours or 4°C for 12 hours (Section 5.3.3.1.1.1.1 and 5.3.3.1.1.1.2). Pure MORC4$^{420\text{-}757}$ protein was used as a positive control together with MDA-MB-231, a breast adenocarcinoma cell line, which had been previously found to express detectable MORC4 protein (Yang *et al.*, 2019). Lymphoma cell lines and the control cell line MDA-MB-231, did not show any expression of MORC4 however, the recombinant positive control protein MORC4$^{420\text{-}757}$ was detected, which was more intense for membranes that had been incubated for 12 hours at 4°C in monoclonal antibodies (Figure 5.23).

Undiluted cell supernatant from the monoclonal hybridoma cultures were utilised in an attempt to investigate MORC4 protein expression following the lack of detectable MORC4 signal. Several human cancer cell lines were probed with the undiluted MORC4 monoclonal antibodies (supernatants), which had been incubated at 4°C for 12 hours. The results showed that there was no detectable expression of MORC4 in any cancer cell line, yet the positive recombinant MORC4 control was detected (Figure 5.24). This could suggest that there was no, or very low, levels of MORC4 protein in the cancer cell lines, which is surprising considering it was detected in the MDA-MB-231 breast cancer cell line previously (Yang *et al.*, 2019).

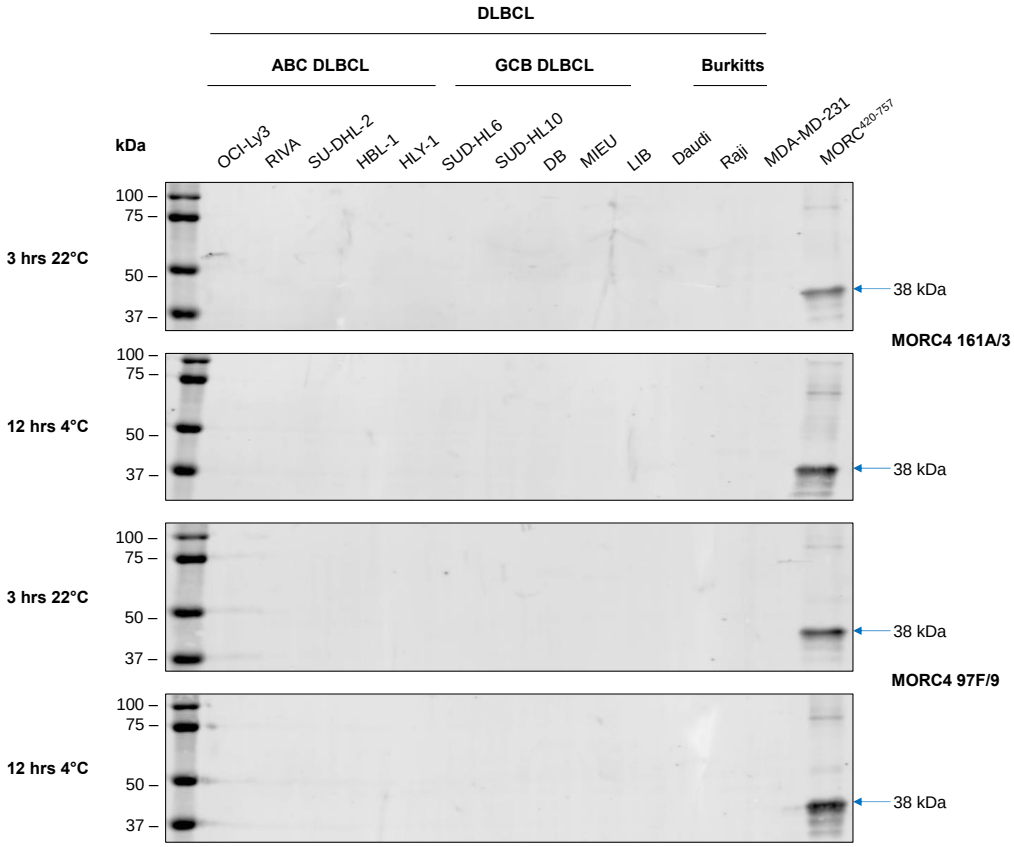


Figure 5.23: **Optimisation of fluorescence detection western blotting conditions of MORC4 monoclonal antibodies with lymphoma cell lines and MORC4$^{420\text{-}757}$.** Cells were harvested, sonicated and 20µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. MORC4$^{420\text{-}757}$ was loaded at 0.2µg. Membranes were probed with MORC4 161A/3 and MORC4 97F/9, which were prepared in a 1:50 dilution in 5% milk or 5% BSA and TBS-T. Monoclonal antibodies were incubated at either 4°C for 12 hours or 22°C for 2 hours. Membranes were detected using a Li-COR Odyssey imagining system and all membranes were imaged using the red channel (700nm). Blue arrows indicate the expected size of MORC4$^{420\text{-}757}$.

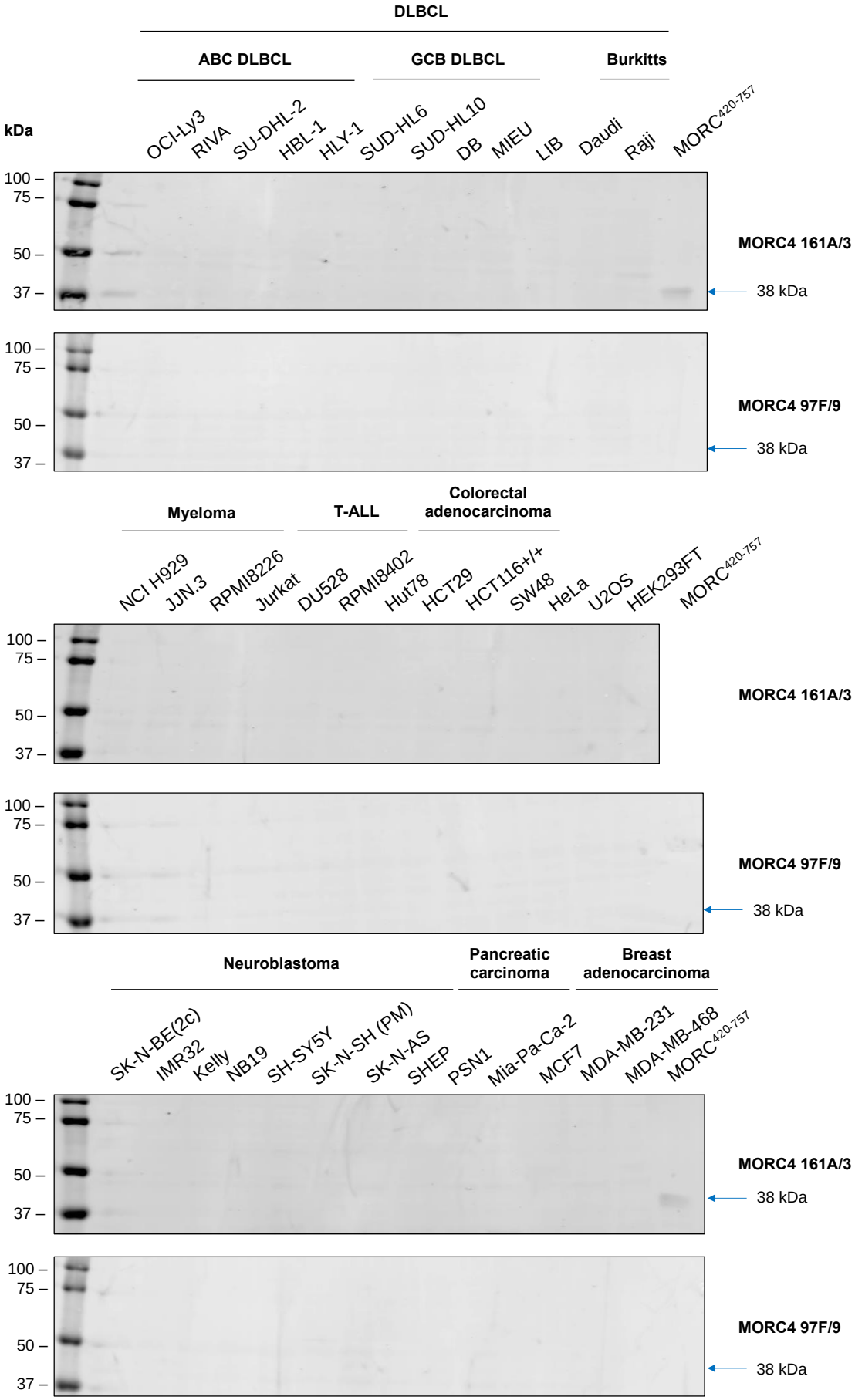

DLBCL
ABC DLBCL
GCB DLBCL
Burkitts
kDa
OCI-Ly3
RIVA
SU-DHL-2
HBL-1
HLY-1
SUD-HL6
SUD-HL10
DB
MIEU
LIB
Daudi
Raji
MORC420-757
100 –
75 –
50 –
37 –
MORC4 161A/3
38 kDa
MORC4 97F/9
38 kDa
Myeloma
T-ALL
Colorectal adenocarcinoma
NCI H929
JJN.3
RPMI8226
Jurkat
DU528
RPMI8402
Hut78
HCT29
HCT116+/+
SW48
HeLa
U2OS
HEK293FT
MORC420-757
Neuroblastoma
Pancreatic carcinoma
Breast adenocarcinoma
SK-N-BE(2c)
IMR32
Kelly
NB19
SH-SY5Y
SK-N-SH (PM)
SK-N-AS
SHEP
PSN1
Mia-Pa-Ca-2
MCF7
MDA-MB-231
MDA-MB-468
MORC420-757

Figure 5.24: **Optimisation of fluorescence detection western blotting conditions of MORC4 monoclonal antibodies with cell lines and MORC4$^{420-757}$.** Cells were harvested, sonicated and 20µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. MORC4$^{420-757}$ was loaded at 0.2µg. Membranes were probed with 11mL of undiluted MORC4 161A/3 and MORC4 97F/9. Monoclonal antibodies were incubated at 4°C for 12 hours. Membranes were detected using a Li-COR Odyssey imagining system and all membranes were imaged using the red channel (700nm). Blue arrows indicate the expected size of MORC4$^{420-757}$.

5.3.3.2 Detection of MORC protein expression using polyclonal antibodies

5.3.3.2.1 Detection of MORC proteins using fluorescence detection western blotting

5.3.3.2.1.1 MORC1, MORC2 and MORC3 protein expression

In addition to MORC4, polyclonal antibodies against MORC1, MORC2 and MORC3, were also used to investigate expression of MORC proteins in a number of cancer cell lines.

5.3.3.2.1.1.1 MORC1 expression was potentially observed in DLBCL and neuroblastoma

MORC1 was found to express at approximately 250 kDa (Figures 5.25, 5.26 and 5.27), which was double the size of the expected molecular weight of 112 kDa, suggesting that MORC1 could run as an unresolved dimer in these cell lines, or a non-specific protein was detected. If the former, low expression of MORC1 was found in several types of diffuse large B-cell lymphomas (Figure 5.25) and neuroblastomas (Figure 5.27).

5.3.3.2.1.1.2 MORC2 expression in neuroblastoma and pancreatic carcinoma

MORC2 expected size was 118 kDa, however, MORC2 expression showed two bands of approximately 120 and 130 kDa in several cancerous cell lines (Figure 5.25, 5.26 and 5.27). From this analysis, the smaller band of 120 kDa is more likely to be MORC2 as it is similar to the expected size. There was low expression of MORC2 in several DLBCL subtypes (Figure 5.25) however, there was a higher level of expression for neuroblastomas and pancreatic carcinomas (Figure 5.27).

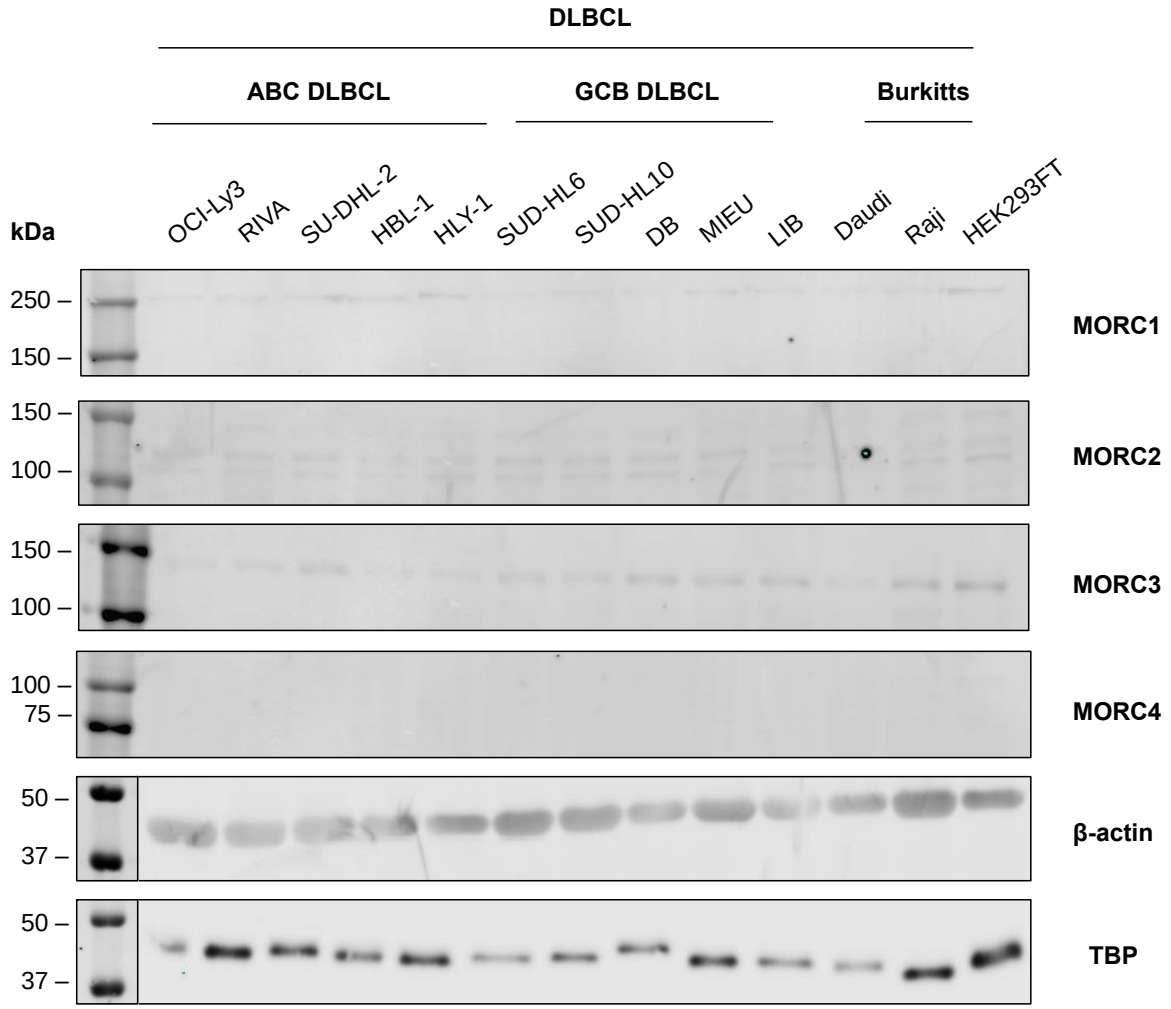


Figure 5.25: **Fluorescence detection western blotting using MORC polyclonal antibodies with lymphoma cell lines (20 µg).** Cells were harvested, sonicated and 20µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with polyclonal MORC, β-actin and TBP antibodies, which were diluted in 5% milk and TBS-T. Membranes were detected using a Li-COR Odyssey imagining system. MORC membranes and markers were imaged using the red channel (700nm) and control samples, TBP and β-actin, were imaged using the green channel (800nm).

5.3.3.2.1.1.3 MORC3 was expressed in DLBCL, neuroblastoma and pancreatic carcinomas

MORC3 expected size was 107 kDa, however, the fluorescent detection western blotting results show MORC3 expression at between approximately 120-140 kDa (Figures 5.25, 5.26 and 5.27). MORC3 was highest expressing in DLBCLs and HEK293FT (Figure 5.25). MORC3 was also expressed in some neuroblastomas and two types of pancreatic carcinomas, PSN1 and

Mia-Pa-Ca-2 (Figure 5.27). However, these polyclonal antibodies may not be validated suggesting that protein bands could be non-specific.

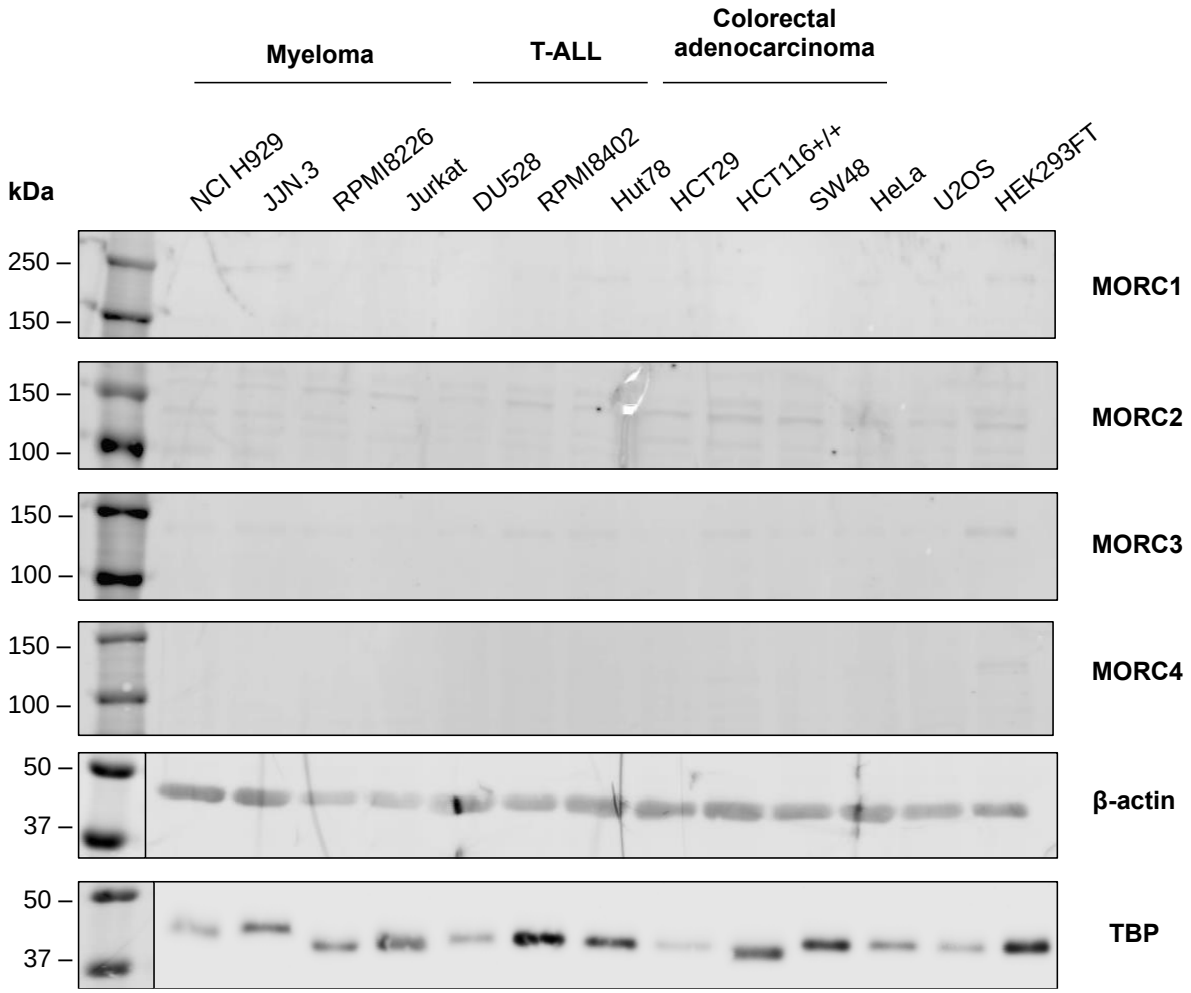


Figure 5.26: **Fluorescence detection western blotting using MORC polyclonal antibodies with cancer cell lines (20 µg).** Cells were harvested, sonicated and 20µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with polyclonal MORC, β-actin and TBP antibodies, which were diluted in 5% milk and TBS-T. Membranes were detected using a Li-COR Odyssey imagining system. MORC membranes and markers were imaged using the red channel (700nm) and control samples, TBP and β-actin, were imaged using the green channel (800nm).

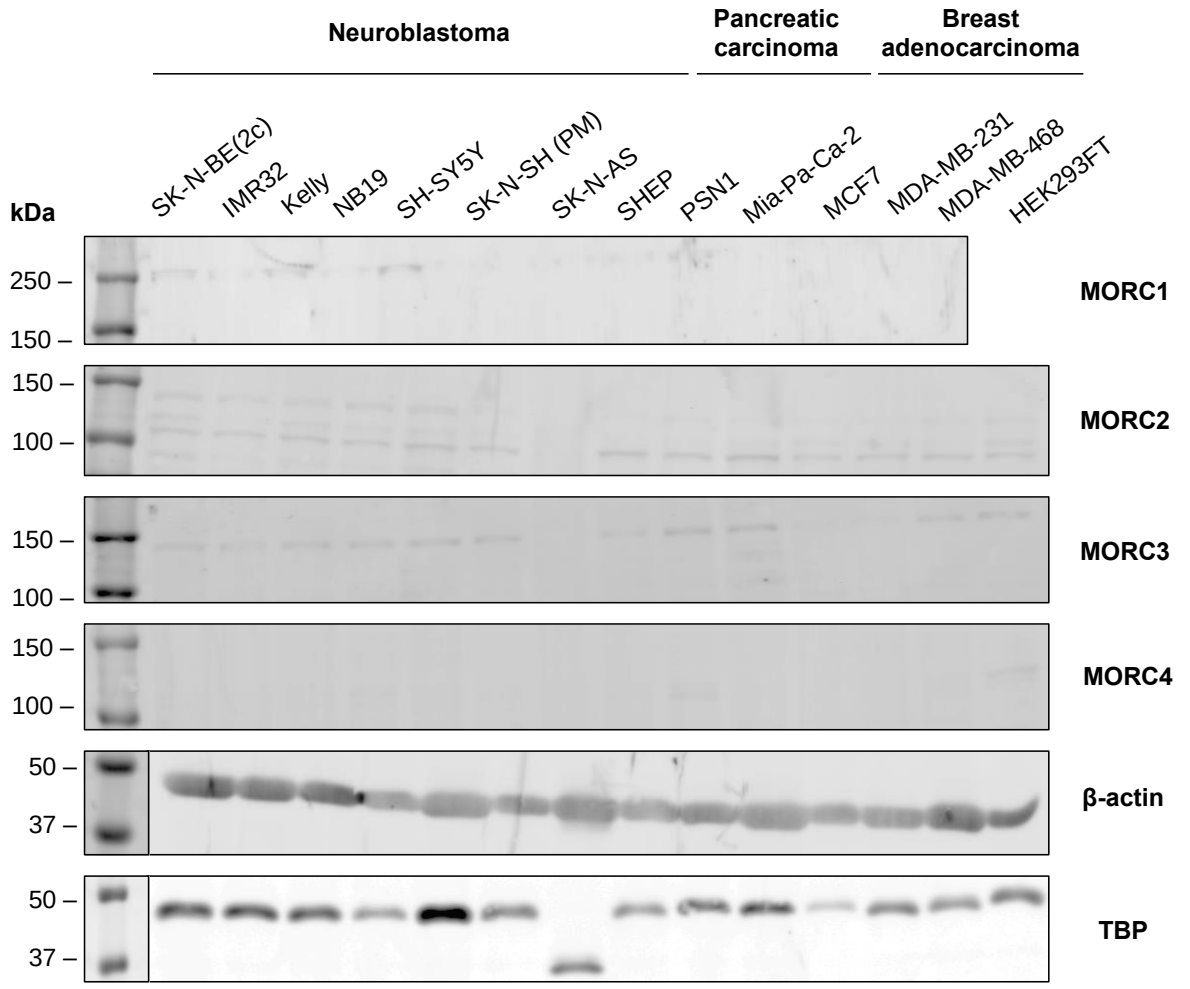


Figure 5.27: **Fluorescence detection western blotting using MORC polyclonal antibodies with cancer cell lines (20 µg).** Cells were harvested, sonicated and 20µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with polyclonal MORC, β-actin and TBP antibodies, which were diluted in 5% milk and TBS-T. Membranes were detected using a Li-COR Odyssey imagining system. MORC membranes and markers were imaged using the red channel (700nm) and control samples, TBP and β-actin, were imaged using the green channel (800nm).

5.3.3.2.1.2 MORC4 protein expression

Following unsuccessful detection of expression of MORC4 using monoclonal antibodies, a commercial anti-MORC4 polyclonal antibody (Sigma) was used in an attempt to identify MORC4 protein expression. This particular MORC4 polyclonal antibody was previously used in another study where high expression of MORC4 was detected in the breast cancer cell line, MDA-MB-231 (Yang *et al.*, 2019). In contrast to this previous study, there was no detectable MORC4 protein expression of MDA-MB-231 found in this study although the positive

MORC4 recombinant protein control was detected (Figure 5.27 and 5.28). When utilising 20 µg of protein extract, fluorescent detection western blotting showed that there was no detectable MORC4 protein in any lymphoma cell line, although it was observed for the non-cancerous transformed human embryonic kidney cell line, HEK293FT, with a very faint band at approximately 125 kDa (Figures 5.25, 5.26 and 5.27). Hence, HEK293FT was also used across all western blotting experiments to assess and compare MORC4 expression across multiple western blots. In addition, there was also faint detection of MORC4 in the pancreatic adenocarcinoma cell line, PSN1 (Figure 5.27). Both β-actin and TATA-binding protein (TBP) (Lossos *et al.*, 2003) were used as loading controls throughout MORC4 fluorescent detection western blotting. TBP is a nuclear protein transcription factor (Davidson, 2003), which has a molecular weight between 37-48 kDa. Both β-actin and TBP were detected at generally equivalent levels in all cell lines (although β-actin was more consistent) allowing for comparison. However, variations of TBP molecular weight were seen between each cell line (Figures 5.25, 5.26 and 5.27), in particular, the neuroblastoma SK-N-AS cell line, which expressed TBP at approximately 37 kDa instead of 50 kDa.

In an attempt to identify MORC4 protein expression, fluorescent detection western blotting was also carried out using a greater quantity of protein per blotted sample (40 µg). (Figure 5.28). MORC4 expression was also seen in the HEK293FT cell line, similar to the lower protein concentration western blot. However, two protein bands were visible (approximately 100 kDa and 75 kDa), suggesting that there were two possible isoforms of MORC4 present in the HEK293FT cell line. It could also suggest degradation or non-specific background detection. As with HEK293FT, multiple bands were also seen in some myeloma, T-cell acute

lymphoblastic leukaemia (T-ALL) and colorectal adenocarcinoma cell lines, which could also suggest non-specific background detection or degradation (Figure 5.28).

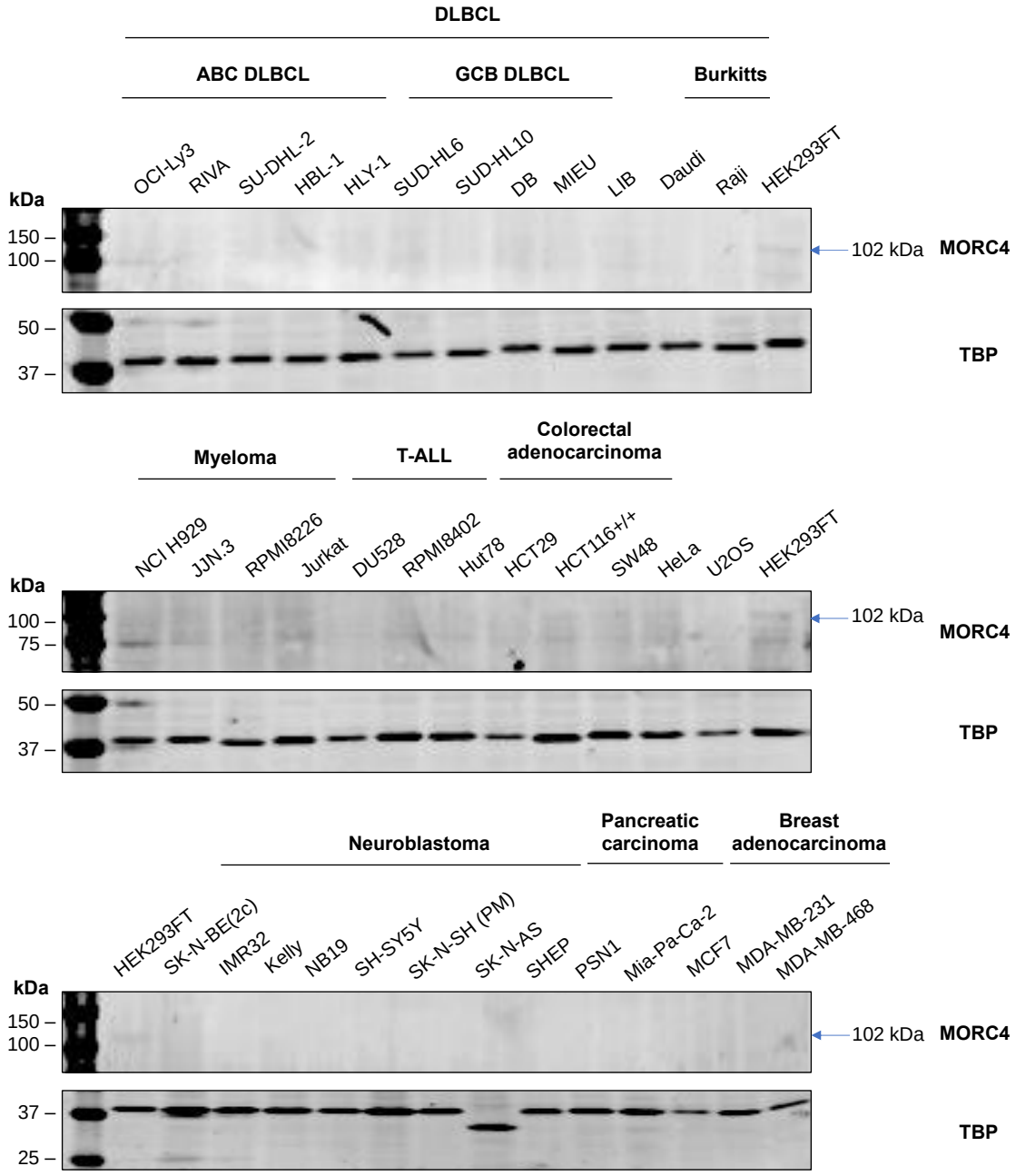


Figure 5.28: **Fluorescence detection western blotting using a MORC4 polyclonal antibody with cancer cell lines (40 µg).** Cells were harvested, sonicated and 40µg was loaded onto an SDS-PAGE gel and used for fluorescence detection western blotting to determine the presence of protein. Membranes were probed with either 1:1000 of commercial MORC4 or 1:2000 TBP antibodies, which were both diluted in 5% milk and TBS-T. Primary antibodies were incubated at room temperature for 4 hours. Membranes were detected using a Li-COR Odyssey imagining system and all membranes were imaged using the red channel (700nm). Blue arrows represent the MORC4 protein band in the HEK293FT cell line (102 kDa).

5.3.3.2.2 Detection of MORC proteins using enhanced chemiluminescence (ECL) western blotting

5.3.3.2.2.1 MORC4 protein expression

Enhanced chemiluminescence (ECL) western blotting was also used to investigate MORC4 protein expression in several human cancer cell lines, using the same commercial polyclonal MORC4 antibody utilised in fluorescence detection western blotting. This was to assess if ECL was more sensitive as MORC4 was previously only detected in the HEK293FT cell line by fluorescent detection western blotting (Section 5.3.3.2.1). ECL detection showed fairly even detection of the loading control protein TBP, with expression at approximately 40 kDa for the majority of cell lines except for SK-N-AS (Figure 5.31), which expressed TBP at approximately 37 kDa. ECL western blotting revealed low MORC4 protein expression for OCI-Ly3, HBL-1 and HLY-1 and ABC-DLBCL lymphoma cell lines at 100 kDa, however, no bands were detected in GCB-DLBCL lines or RIVA and SU-DHL-2 ABC DLBCL cell lines (Figure 5.29). This data had a similar pattern to results seen in qRT-PCR analysis (Section 5.2.2), which is also supported by a previous study that identified MORC4 as a potential biomarker of ABC-DLBCL (Liggins *et al.*, 2007). LIB, an unidentified DLBCL cell line and HEK293FT showed high levels of expression of MORC4, also present at 100 kDa (Figure 5.29), with HEK293FT also showing bands at 25 kDa and 37 kDa, although likely reflecting protein possible degradation. This result may suggest that perhaps the DLBCL cell line, LIB, belongs to the ABC subtype as no expression of GCB subtypes was detected. Overall, this also suggests that OCI-Ly3, HBL-1, HLY-1, LIB and HEK293FT cell lines express the full-length MORC4 isoform.

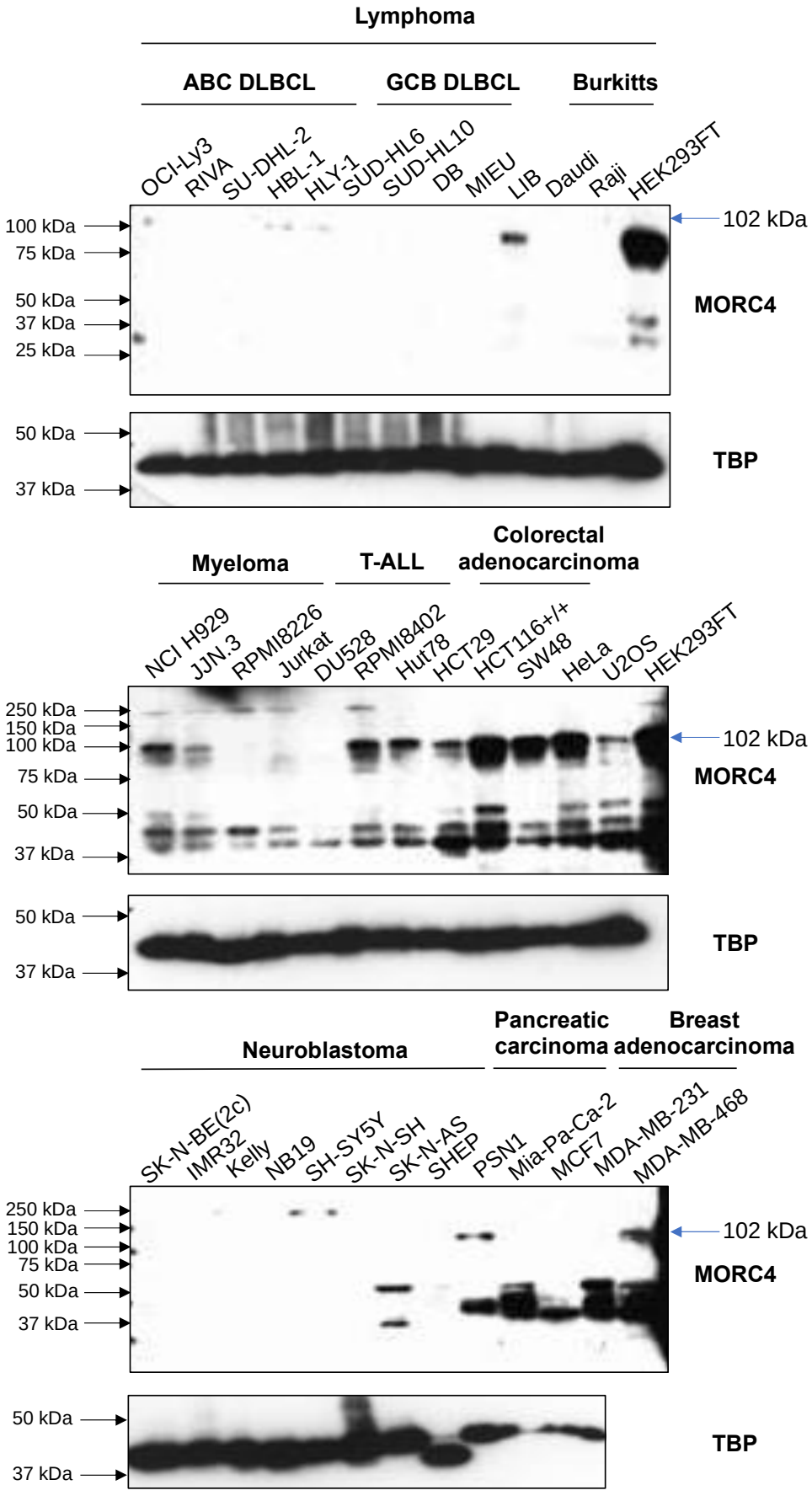

Lymphoma
ABC DLBCL
GCB DLBCL
Burkitts
OCI-Ly3
RIVA
SU-DHL-2
HBL-1
HLY-1
SUD-HL6
SUD-HL10
DB
MIEU
LIB
Daudi
Raji
HEK293FT
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
102 kDa
MORC4
50 kDa
37 kDa
TBP
Myeloma
T-ALL
Colorectal adenocarcinoma
NCI H929
JJN.3
RPMI8226
Jurkat
DU528
RPMI8402
Hut78
HCT29
HCT116+/+
SW48
HeLa
U2OS
HEK293FT
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
102 kDa
MORC4
50 kDa
37 kDa
TBP
Neuroblastoma
Pancreatic carcinoma
Breast adenocarcinoma
SK-N-BE(2c)
IMR32
Kelly
NB19
SH-SY5Y
SK-N-SH
SK-N-AS
SHEP
PSN1
Mia-Pa-Ca-2
MCF7
MDA-MB-231
MDA-MB-468
250 kDa
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
102 kDa
MORC4
50 kDa
37 kDa
TBP

***Figure 5.29*: Enhanced chemiluminescence (ECL) western blotting conditions of MORC4 polyclonal antibody and TBP with cancer cell lines (40 µg).** Cells were harvested, sonicated and 40µg was loaded onto an SDS-PAGE gel and used for enhanced chemiluminescence detection western blotting to determine the presence of protein. Membranes were probed with either MORC4 polyclonal (1:1000) or TBP (1:2000), which had been diluted 5% milk and TBS-T. Primary antibodies were incubated at 4°C for 12 hours. MORC4 and TBP membranes were detected using a BM ECL blotting substrate (POD) kit and exposed in a dark room for 30 minutes and 100 minutes, respectively. Blue arrows represent the expected size of MORC4 (102 kDa).

Colorectal adenocarcinoma cell lines, HCT116+/+ and SW48 also exhibited protein expression of MORC4 (Figure 5.29). In addition, the HCT29 colorectal adenocarcinoma cell line also expressed MORC4 but at a lower level (Figure 5.29). Cell lines expressing lower levels of MORC4 included myeloma and T-ALL cell lines. However, all of these cell lines exhibited a potential MORC4 band at approximately 100 kDa, suggesting expression full-length MORC4 was present in these cell lines. All myeloma cell lines and one T-ALL cell line (RPMI8402) also showed a potential MORC4 band at 250 kDa, suggesting a large novel MORC4 isoform, perhaps unresolved dimeric MORC4 similar to other MORC proteins (Douse *et al.*, 2018; Li *et al.*, 2016) or non-specific background detection (Figure 5.29). In addition, several smaller protein bands were visible below the 100 kDa MORC4 band, suggesting protein degradation. Finally, MORC4 protein bands were also seen between approximately 37-50 kDa, which could also be either several isoforms of MORC4 or degradation of MORC4.

In addition, low expression of the 250 kDa MORC4 putative dimer was also present in neuroblastoma cell line, SH-SY5Y. This also suggests that a large MORC4 isoform or perhaps unresolved dimeric MORC4 is present in this sample, as with myeloma and T-ALL cell lines. Higher expression of the 100 kDa MORC4 isoform was visible in both pancreatic carcinoma cell line PSN1 and breast adenocarcinoma cell line MDA-MB-468 (Figure 5.29). Similar to other cell lines, pancreatic adenocarcinoma and breast adenocarcinoma cell lines also showed MORC4 protein bands between 37-50 kDa, suggesting degradation of MORC4.

The ECL western blot analysis using polyclonal antibodies seemed to be more sensitive than fluorescent detection western blotting. HEK239FT, PSN1, HBL-1, HCT116+/+, SW48 and HeLa cell lines exhibited higher MORC4 protein levels than other cancer cell lines probed. This result correlated with the qRT-PCR data, which was particularly noticeable for the ABC-DLBCL cell lines, suggesting a relationship between RNA and protein levels as expected, although, MORC4 levels are generally low overall. This could indicate that MORC4 protein levels could be detected using immunohistochemistry

5.4 Discussion

MORC4 mRNA and protein expression analysis was utilised throughout this chapter in an attempt to indicate potential involvement of MORC4 in human cancers, which may ultimately be reflected in MORC4 function. This chapter included a combination of both RNA and protein analysis to investigate MORC4 expression in a number of diseases, which could give an insight into the function of MORC4 in relation to human cancers.

5.4.1 MORC4 RNA analysis

In an attempt to study mRNA transcripts of MORC4 in several cancer cell lines, non-quantitative (endpoint) reverse transcription (RT)-PCR was used to identify potential MORC4 splice variants following sequence analysis of known splice variants. RT-PCR determined that of the cell lines used in this study, all of them expressed transcripts containing exons that encode the MORC4 N-terminal ATPase region. However, some cell lines appeared to encode the C-terminus of MORC4 encoding terminal ‘GAS’ residues, but the majority of cell lines expressed transcripts with the encoding C-terminal end in common with both ‘HILD’ and ‘GAS’ isoforms. The majority of cell lines tested detected nearly full-length MORC4 (F1/R4 primer combination) except for neuroblastoma cell lines. In addition, some cell lines such as SW48 and PSN1 detected a larger transcript of MORC4 at 2.5 kb instead of the more

commonly detected 2.1 kb full-length transcript. This suggests that both SW48 and PSN1 may contain a MORC4 isoform comprising the coiled coil region. This also indicates that there were functional differences in MORC4 isoforms as there were differences in the MORC4 transcript sizes between cell lines. HEK293FT cell line was also found to have the large 2.5 kb splice variant in addition to high protein expression, this could suggest that perhaps loss of MORC4 is associated with cancer cells. Some cell lines also had multiple PCR products, which could be from non-specific amplification or multiple different MORC4 transcripts expressed simultaneously, for example the 2.1 and 2.5 kb products found in some cell lines. This may suggest that there could be multiple but different coinciding roles for MORC4. In an attempt to confirm the identity of MORC4 transcripts and to rule out any non-specific products, future work may include the extraction of the amplified RT-PCR products, which would then be cloned and-sequenced to confirm identity.

Quantitative (q)RT-PCR analysis identified several cancer cell lines with relatively high MORC4 mRNA expression. Lymphoma cell lines expressed MORC4 mRNA at lower levels when compared to other types of cancer cell lines, such as colorectal adenocarcinoma and breast cancer cell lines. However, ABC-DLBCL cell lines (e.g. HBL-1, HLY-1 and OCI-Ly3) appeared to express MORC4 at higher levels than GCB-DLBCL cell lines (e.g. DB, SUD-HL10 and MIEU), although it was not statistically significant following statistical analysis. These results are similar to a previous study, which also found that ABC-DLBCL (HLY-1 and OCI-Ly3 cell lines were higher expressing than GCB-DLBCL (DB, SUD-HL10 and SUD-HL6) (Liggins *et al.*, 2007). In addition, ER-negative breast cancer cell lines had higher expression of MORC4 than ER-positive breast cancer cell lines, suggesting that perhaps MORC4 plays a role in ER-negative cancer.

Ideally, qPCR analysis would have also included mRNA expression of normal non-cancerous B-cells such as $CD19^+$ and $CD20^+$ cell lines, similar to a previous study (Liggins *et al.*, 2007).

However, normal B-cells were not available for this analysis. Analysis of non-cancerous B-cells is important for future work to this study, in order for true normalisation of B-cell lymphomas and for full normalisation with normal cells. In addition to lymphoma cell lines, colon, pancreatic and breast cancer cell lines also showed relatively high expression of MORC4. Furthermore, it would be important to produce biological replicates of this experiment by using cells grown from a separate passage. In addition, the identity of the cells should be tested using short tandem repeat (STR) typing (Butler, 2007; Dirks *et al.*, 2005) and mycoplasma testing to eliminate the possibility of contamination as mycoplasma can affect the metabolic function and cell morphology (Nikfarjam and Farzaneh, 2012). It would also be important to produce technical replicates performed in this study. In addition, another qPCR probe could be utilised to determine expression of MORC4 as qPCR probes only span one exon boundary, in particular, the 13-14 exon boundary at 1779 bp in MORC4.

A previous study suggested that MORC4 mRNA is expressed significantly higher in DLBCL cells compared to normal healthy B-cells, which indicates that MORC4 is a potential DLBCL biomarker (Liggins *et al.*, 2007). Similarly, bioinformatic analysis using RNA expression datasets revealed differences in expression of MORC4 between non-cancerous tissues and cancerous tissues, including DLBCL. The RNA expression atlas database (EBI) suggested that MORC4 is expressed at low levels in several types of tissue. The placenta was the highest expressing tissue followed by adrenal gland and the testis. Similar to RNA expression database results, one study also suggested that MORC4 is expressed at low levels in several non-cancerous tissues specifically, the placenta and the testis (Liggins *et al.*, 2007). The EBI RNA expression atlas also suggested that MORC4 is expressed at low levels in colon tissue. GENT2 analysis also indicates that there is a significant increase in RNA expression of MORC4 in colon adenocarcinoma compared to non-cancerous colon tissue. EBI analysis also revealed that there is high expression of MORC4 in the colon adenocarcinoma cell line, SW48. Kaplan-

Meier plots also indicate that high expression of MORC4 decreases survival in patients with colon adenocarcinoma whereas, patients with low expression of MORC4 have a higher probability of survival. Furthermore, evidence also suggests that there is an SNP on the CLDN2-MORC4 locus, which has been thought to cause inflammatory bowel disease (Soderman *et al.*, 2013). Overall, this data could suggest that MORC4 may play a role in colon adenocarcinoma. However, more evidence is required to identify MORC4 role in cancer.

In addition to colon adenocarcinoma, breast adenocarcinoma was also analysed using RNA expression databases. Both GEPIA and GENT2 databases revealed that there was higher expression of MORC4 in breast adenocarcinoma tissue compared to normal non-cancerous breast tissue. Survival and prognosis probability were also analysed using Kaplan-Meier plots. The results revealed that there was a poor prognosis of patients with breast adenocarcinoma with high expression of MORC4, similar to (Yang *et al.*, 2019).

EBI RNA expression atlas revealed that MORC4 was expressed at low levels in the pancreas, similar to another study which investigated expression levels of MORC4 in normal tissues (Liggins *et al.*, 2007). Analysis of the EBI dataset also revealed high expression of MORC4 in the pancreatic adenocarcinoma cell line, PSN1. Both GENT2 and GEPIA databases revealed that pancreatic adenocarcinoma had significantly higher expression of MORC4 compared to normal non-cancerous pancreatic tissue. Interestingly, some evidence also suggests that the (CLDN2)-MORC4 locus has an SNP associated with chronic pancreatitis (Monique H Derikx *et al.*, 2015). Survival and prognosis analysis using PROGgeneV2 and the Kaplan-Meier plotter both demonstrated that high levels of MORC4 expression have a higher probability of survival, unlike a breast cancer study, which suggested that MORC4 could be a breast cancer oncogene after concluding that high levels of MORC4 expression correlated with a poorer outcome, in addition to knockdown of MORC4 promoting apoptosis (Yang *et al.*, 2019).

Following RNA analysis using a combination of RT-PCR, qRT-PCR and RNA expression databases, several similarities were found across all analyses. The RNA expression database, EBI, revealed that there was high expression of MORC4 in the colon adenocarcinoma cell lines, SW48, similar to results from qPCR. In addition, the EBI database also showed high expression of MORC4 in the pancreatic adenocarcinoma cell line, PSN1, which was also seen in the qPCR results herein. Conversely, the Mia-Pa-Ca-2 pancreatic cancer cell line did not show such a high level of RNA MORC4 expression as PSN1, suggesting that more pancreatic cell lines are required for analysis of MORC4 RNA and protein expression, in addition to a normal transformed pancreatic cell line in order to determine that MORC4 expression is upregulated in pancreatic cancer. Of several cancer cell lines tested, PSN1, had the highest MORC4 expression of all cell lines following qRT-PCR. Some evidence suggests that healthy non-cancerous pancreatic tissue expresses a low level of MORC4 (Liggins *et al.*, 2007). Evidence also suggests that the X chromosomal linked Claudin 2 (CLDN2)-MORC4 locus has a single nucleotide polymorphism (SNP) (when a specific single nucleotide varies between species or humans), which is associated with chronic pancreatitis (Monique H Derikx *et al.*, 2015). Similarly, an SNP on the CLDN2-MORC4 locus was also found to cause inflammatory bowel disease (Soderman *et al.*, 2013). The qPCR results also showed high MORC4 expression of HCT116+/+ and SW48 colorectal adenocarcinoma cell lines. These results suggest that MORC4 could play a role in both colorectal cancer and pancreatic adenocarcinomas, although more evidence is required to determine MORC4 function in these diseases. In addition to pancreatic and colon cancer, breast adenocarcinoma cell line MDA-MB-231 showed relatively high expression of MORC4 following qPCR and several RNA databases showed evidence that MORC4 expression was higher in breast cancer than normal healthy breast cells. The evidence found in this study was also similar to a previous study, which used RNA analysis to come to the conclusion that MORC4 is a breast cancer oncogene (Yang *et al.*, 2019).

5.4.2 MORC4 protein analysis

MORC4 protein analysis was important for helping to determine function of MORC4 as proteins are functional molecules and clinical samples can only be analysed using protein analysis and antibodies, hence the optimisation and characterisation of novel MORC4 monoclonal antibodies was an important part of this study. In an attempt to verify MORC monoclonal antibodies in mammalian cell lines, full-length *MORCs* were cloned using ligation independent cloning (LIC) into mammalian expression plasmids, pCDNA3-N-FLAG-LIC and pCDNA3-N-HA-LIC to ensure the antibodies did not recognise other MORC paralogues. Prior to cloning of full-length *MORC* DNA, *MORCs* were amplified using PCR. *MORC1-3* were amplified using a standard PCR protocol. However, *MORC4* only amplified with a mix of Phusion® high fidelity DNA polymerase and hot start standard MyTaq™ DNA polymerase with addition of hot start DNA polymerase buffer to amplify full-length *MORC4* DNA without any mutations. Some evidence suggests that fast-acting Taq DNA polymerase has the ability to introduce errors into synthesised DNA 10-fold, compared with Phusion® high fidelity DNA polymerase (McInerney *et al.*, 2014). This suggests that a mixture of both DNA polymerases could indicate that MyTaq™ DNA polymerase is able to initiate DNA synthesis and Phusion® high fidelity DNA polymerase prevents the formation of mutations by using 3'-5' exonuclease activity (Bębenek and Ziuzia-Graczyk, 2018) to remove misincorporated nucleotides by Taq polymerase. Similar experiments have been demonstrated where DNA polymerases with different properties have been mixed in an attempt to overcome issues associated with amplification of DNA during PCR, especially when amplifying large DNA templates (Śpibida *et al.*, 2017). Furthermore, MyTaq™ DNA polymerase buffer was also used, which contained dNTPs, $MgCl_2$ and a number of stabilisers and enhancers at an optimal concentration, unlike Phusion® high fidelity DNA polymerase where additional additives were required manually.

This could suggest that the use of MyTaq™ DNA polymerase buffer resulted in less error and therefore generated the amplification of full-length *MORC4*.

Following cloning of full-length *MORCs* with the use of mammalian plasmids, FLAG-tagged and HA-tagged full-length *MORCs* were transfected into several mammalian cell lines including U2OS, HEK293FT, HeLa and COS-1 monkey cell line. Only successful expression of MORC1 and MORC2 was seen in HEK293FT cell line. No expression was found for MORC3 and MORC4 in any of the cell lines, suggesting that perhaps the isoforms of MORC4 used to create epitope-tagged MORCs were not stable in these cell lines. Overall, this indicates that this method could not be utilised for verification of MORC4 monoclonal antibodies. Future work may include validation of antibodies using recombinant MORC4 protein and western blotting.

Several cancer cell lines were studied to investigate MORC4 protein expression using the anti-MORC4 monoclonal antibodies. However, there was no detected expression of MORC4 found in any cell line even with the use of undiluted antibody from hybridoma cultures. Although monoclonal antibodies were validated to recognise MORC4 protein (as BSA was not detected), it could suggest that the antibodies were too weak (or dilute) to detect the low levels of MORC4 in cell lines. Future work for this study could include western blotting against recombinant MORC1, MORC2 and MORC3 in addition to purifying the monoclonal antibodies from the hybridoma supernatants and concentrating them, in an attempt to detect the low level of MORC4. To overcome this issue using MORC4 monoclonal antibodies, an anti-MORC4 polyclonal antibody, which binds to multiple epitopes of MORC4 and has recently been used in another published study (Yang *et al.*, 2019), was used to investigate MORC4 protein expression. Low expression of MORC4 protein was found in some of the lymphoma cell lines when using 40 µg of blotted protein extract, suggesting that MORC4 protein expression was low in lymphomas. Although, ECL western blotting did identify higher levels of MORC4

expression in ABC-DLBCL compared to GCB-DLBCL, which was also reflected in the qRT-PCR results. Conversely, SU-DHL-2 and RIVA ABC-DLBCL cell lines did not show a high level of MORC4 expression compared to GCB-DLBCL, however, this could reflect disease heterogeneity. Interestingly, a study which investigated MORC4 RNA expression levels in DLBCL subtypes, also found that MORC4 had higher expression in ABC-DLBCL compared to GCB-DLBCL (Liggins *et al.*, 2007). Higher MORC4 protein expression was found for neuroblastoma, pancreatic carcinoma and breast adenocarcinoma cell lines, than in lymphoma cell lines. Fluorescent detection western blotting requires an SDS wash stage for the secondary antibody to reduce the background. This detergent wash may be responsible for removal of some of the primary antibody as SDS is a strong detergent. This could explain why the ECL detection western blotting was more sensitive even though the same primary polyclonal antibody was used for both detection methods.

Using the MORC4 polyclonal antibody, MORC4 protein expression was detected in HEK293FT transformed human embryonic kidney cell line, HCT116+/+ and SW48 colorectal adenocarcinoma cell lines, pancreatic carcinoma PSN1 cell line and MDA-MB-468 breast adenocarcinoma cell line. In some cell lines it was evident there was a protein band at 250 kDa, suggesting a large novel MORC4 isoform. This could indicate a retained intron from splicing, which can be a common occurrence in cancers (Dvinge and Bradley, 2015; El Marabti and Younis, 2018), with some studies suggesting that intron retention is important for gene regulation (Jacob and Smith, 2017).

In addition, the housekeeping protein, β-actin, was detected in all cell lines, with equivalent loading observed for comparisons to be made between sample. TATA box-binding protein (TBP) control was also detected cell lines, albeit with more variable expression levels. There was, however, molecular weight variation in each sample ranging from 37 to 50 kDa. Evidence suggests that the N-terminal of TBP consists of a long chain of glutamines (CAG repeats),

which regulate the DNA binding activity of the C-terminus. The number of N-terminal CAG repeats varies depending on health and disease. Healthy humans have approximately 27-44 CAG repeats whereas, there can be approximately 50-55 CAG repeats in a human with a disease (Zühlke *et al.*, 2001). Some evidence also suggests that the longer the length of the CAG repeat, the more serious the disease (Xiang *et al.*, 2018). This suggests an explanation for the molecular weight variation seen in TBP protein expression. It also indicates that neuroblastoma (except for SK-N-AS), pancreatic carcinoma and breast adenocarcinoma cancers are more severe than other cancers tested in this experiment as they have a larger expression of TBP.

5.4.3 Overview of MORC4 RNA and protein expression analysis

5.4.3.1 Diffuse large B-cell lymphoma (DLBCL)

The GEPIA RNA database suggested that MORC4 RNA was significantly more expressed in DLBCL cells compared to normal cells. However, the expression level in transcripts per million (TPM) was low compared to several other types of diseases. Interestingly, this was also seen in a previous study (Liggins *et al.*, 2007) and in this study, where MORC4 RNA was expressed at low levels in comparison to other types of diseases. More specifically, low expression of MORC4 mRNA in lymphoma cell lines was seen, with the highest expression in HBL-1 and HLY-1 ABC-DLBCL cell lines. This relatively low RNA expression could explain why there appeared to be very low detection of MORC4 protein in DLBCL cells following western blotting, although some protein expression was seen in ABC lines HLY-1 and HBL-1 but at much lower levels than other cell lines. Furthermore, high mRNA expression does not necessarily correlate with a high expression of protein (Greenbaum *et al.*, 2003), suggesting that although there is a significant increase of MORC4 mRNA in cancerous cell lines, it does not automatically correspond to a higher level of MORC4 protein (Yansheng Liu *et al.*, 2016). An example is a study which investigated three genes associated with prostate cancer and

showed no significant relationship between RNA expression and protein expression of these genes (Lichtinghagen *et al.*, 2002). However, another study found a strong correlation between RNA expression in several human carcinomas and protein expression levels (Ørntoft *et al.*, 2002).

There are several potential reasons for poor correlation between levels of RNA and protein (Greenbaum *et al.*, 2003). One example is that there are multiple post-transcriptional modifications that can occur in the process of translating mRNA into protein (Nachtergaele and He, 2017) such as methylation of mRNA with N6- methyladenosine (m^6A), 5-methylcytosine (m^5C) (Qiu Li *et al.*, 2017) or 2`-O-methylation (Nm) (Dimitrova *et al.*, 2019) chemical modifications. The most common RNA post-transcriptional modification in stable RNA is pseudouridine (ψ), formed by isomerisation of uridine (Zhao *et al.*, 2018).

5.4.3.2 Colon adenocarcinoma

RNA database analysis suggested that colon adenocarcinoma cell lines, specifically SW48, expressed high levels of MORC4. In addition, qRT-PCR analysis also revealed that colon adenocarcinoma cell lines also expressed relatively high levels of MORC4, which was significantly higher than in lymphoma cell lines. Protein analysis also showed high MORC4 expression in SW48 and HCT116+/+ colon adenocarcinoma cell lines, indicating that there was a correlation between RNA expression and protein expression in colon adenocarcinoma.

5.4.3.3 Breast adenocarcinoma

RNA databases suggested that there was higher expression of MORC4 in breast cancer cells compared to normal healthy breast cancer cells. RNA databases Kaplan-Meier plots are showed that the higher the expression of MORC4, the lower the chance of survival. Similar to these RNA expression database results, one study revealed over expression of MORC4 in breast adenocarcinoma (Yang *et al.*, 2019). The previous study also demonstrated high protein

expression of breast adenocarcinoma cell lines MDA-MB-231 and MCF7 using western blotting (Yang *et al.*, 2019). Contrary to this study, very low expression was seen for both of these cell lines herein, which were more likely degradation or non-specific bands. The previous study also found evidence that MORC4 protein expression was upregulated in breast cancer, more specifically, the MDA-MB-231 breast adenocarcinoma cell line (Yang *et al.*, 2019). Similar to the qRT-PCR result in this study, MDA-MB-231 and MCF7 were highly expressed when examined by western blotting and were found to be playing a crucial role in breast cancer in the previous study, as when MORC4 was knocked down, cancer cell growth was inhibited (Yang *et al.*, 2019).

5.4.3.4 Pancreatic carcinoma

RNA databases suggested that there was low RNA expression of MORC4 in normal pancreatic tissue. However, there was an increase in expression of MORC4 in pancreatic carcinoma cells, specifically, the PSN1 cell line. Similar to this RNA database analysis, PSN1 had the highest MORC4 mRNA expression of all cell lines tested following qRT-PCR. Western blotting also determined that PSN1 expressed MORC4, suggesting that there was a correlation between expression of MORC4 mRNA and protein in pancreatic carcinoma. Overall, this could suggest that MORC4 could play a role in pancreatic carcinoma. Although, it is not clear if this would be a negative or positive role as RNA analysis Kaplan-Meier plotter revealed that higher expression of MORC4 in pancreatic carcinoma led to a higher chance of survival in patients. Future work would be important to test additional cell lines as the other pancreatic cell line, Mia-Pa-Ca-2, was low in comparison to the PSN1 cell line.

5.4.4 Summary and future work

Colon adenocarcinoma, breast adenocarcinoma, pancreatic carcinoma and DLBCL cell lines all showed high level of expression in both mRNA and protein analyses (Figure 5.30). This may suggest that enhanced MORC4 expression could be playing an important role in these

diseases. For example, it appeared that MORC4 was at higher levels in ABC-DLBCL cell lines than GCB-DLBCL (Figure 5.30), supported by protein and mRNA analysis in this study and in previous RNA studies (Liggins *et al.*, 2007). It may also suggest that perhaps MORC4 is a biomarker for ABC-DLBCL, which could also be involved in the aggression of ABC-DLBCL.

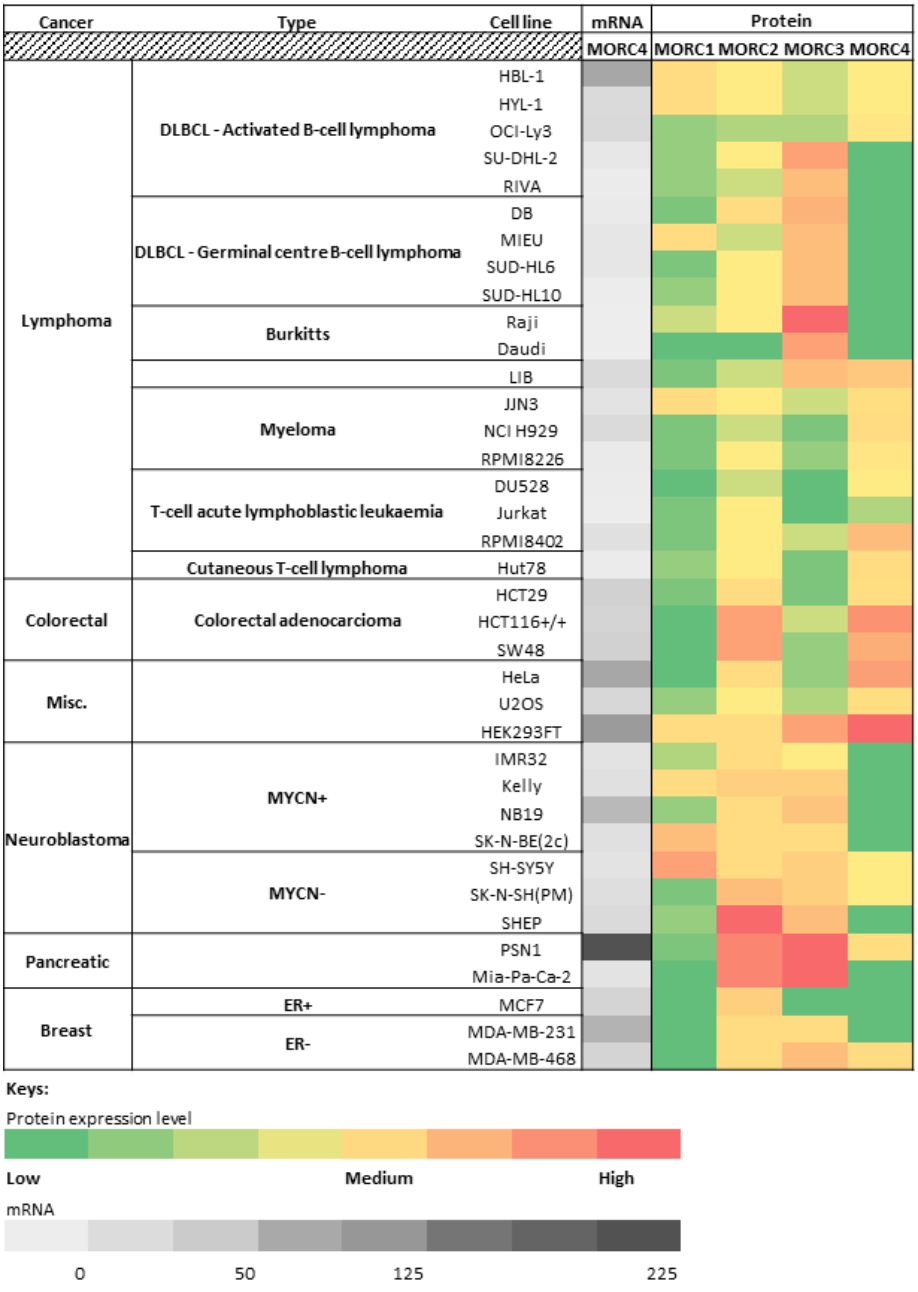


Figure 5.30: **Summary heatmap representation of MORC mRNA expression and protein expression levels.** MORC4 mRNA expression and protein expression levels of MORC1, MORC2, MORC3 and MORC4 in cancer cell lines utilised throughout this study. Data utilised for the mRNA heatmap was taken from the quantitative qPCR of all cell lines. Data for the protein heatmap was taken from the semi-quantitative western blots.

Future work may include additional analysis of additional cell lines such as pancreatic and colon cancer as they were found to express MORC4 at relatively high levels. More cell lines would be required to support evidence that MORC4 is highly expressing in these specific types of cancers. In addition, a different probe could be utilised for qPCR analysis as qPCR probes only amplify one exon boundary. Therefore, it would be interesting to investigate other exon boundaries in an attempt to further analyse the expression of MORC4. In regard to RT-PCR analysis, it would be ideal to extract RT-PCR transcripts and clone them in order to identify sequences in addition to examining if products were non-specific, which could be investigated using next generation sequencing (Kukurba and Montgomery, 2015). Following identification of a stable isoform of MORC4, functional cell biological studies such as CRISPR/siRNA knockdown of *MORC4* or exogenous *MORC4* overexpression could be performed in addition to analysis of subsequent effects on proliferation, migration and resistance to chemotherapeutic drugs. In addition to future work regarding RNA analysis, it would also be beneficial to expand on protein analysis such as investigating expression of MORC 1-3 proteins using ECL western blotting and polyclonal antibodies, as ECL seemed to be more sensitive with the use of the commercial MORC4 polyclonal antibody. Furthermore, it would be useful to optimise MORC4 polyclonal and monoclonal antibodies for immunohistochemistry analysis on clinical formalin-fixed, paraffin-embedded (FFPE) sections to assess if there is any correlation in expression with clinical features.

CHAPTER 6

General Discussion

Chapter 6 General discussion

Microrchidia (MORC) is a highly conserved family of proteins (Iyer *et al.*, 2008), which have been found to bind to methylated histones in chromatin, suggesting a role in chromatin remodelling (Li *et al.*, 2013). There are four members of the MORC protein family in humans, including MORC1, MORC2, MORC3 and MORC4 (Li *et al.*, 2013). There is also a fifth related protein known as structural maintenance of chromosomes flexible hinge domain containing 1 (SMCHD1) (Blewitt *et al.*, 2008), which is thought to function as a tumour suppressor (Leong *et al.*, 2013). The structures and biochemistry of MORC2 and MORC3 are relatively well-characterised, with recent evidence suggesting that they both dimerise in the presence of ATP (Douse *et al.*, 2018; Li *et al.*, 2016). However, MORC4 is currently uncharacterised, with some evidence suggesting that it is a potential biomarker of specific subtypes of lymphoma (Liggins *et al.*, 2007). Domain architecture prediction and sequence analysis has also revealed that MORC3 and MORC4 appear related, suggesting that they could have a similar function. Whereas, MORC1 and MORC2 are different in structure and sequence (Zhang *et al.*, 2019). This study predominantly aimed to biochemically and structurally characterise MORC4 with an aim to determine MORC4 function.

Previous studies also suggest that MORC4 is highly expressed in a specific type of non-Hodgkin lymphoma known as diffuse large B-cell lymphoma (DLBCL). More specifically, MORC4 mRNA levels were found to be upregulated in activated B-cell (ABC)-DLBCL compared to germinal center B-cell (GCB)-DLBCL and healthy B-cells, suggesting that MORC4 is a potential biomarker of ABC-DLBCL (Liggins *et al.*, 2007). However, MORC4 protein expression in lymphoma cell lines is required to further assess the expression with monoclonal antibodies. The second part of this study investigated MORC4 mRNA and protein expression in several cancer cell lines in an attempt to investigate expression in cancer, which

could lead to an insight into MORC4 as a potential biomarker. In addition, MORC4 has recently been identified as a breast cancer oncogene and evidence suggests that MORC4 is overexpressed in breast cancer cell lines, giving rise to a poor prognosis. Therefore, this study also aims to study a number of cancer cell lines to investigate mRNA and protein expression levels of MORC4 to investigate its potential as a cancer biomarker.

Ligation independent cloning (LIC) was used to produce truncated fragments of *MORC4* with the addition of an affinity tag in an *Escherichia coli* expression vector, in an attempt to produce soluble and highly expressing MORC4 proteins. MORC4 proteins included the ATPase domain only, Zf-CW domain only and both domains in tandem. These MORC4 proteins were expressed at a large-scale in *E. coli* and purified to milligram quantities. Pure MORC4 protein was then utilised for use in structural and biochemical characterisation studies of MORC4. Although several MORC4 truncated proteins were produced via this high-throughput method, the tandem domain ATPase-Zf-CW protein MORC4$^{29\text{-}480}$ was used throughout this study for characterisation of MORC4, as it was a high yielding and could be purified to homogeneity.

Structural characterisation was attempted using crystallisation studies and pure MORC4$^{29\text{-}480}$ protein where several crystals were formed during crystallisation trials. However, MORC4 crystals were only present in samples containing AMP-PNP, suggesting that MORC4 dimerises in the presence of AMP-PNP, as with MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016). Both MORC2 and MORC3 structures were solved as dimeric structures with the addition of AMP-PNP, providing evidence that perhaps MORC4 may also dimerise in the presence of AMP-PNP. Following X-ray diffraction screening, attempts at optimising MORC4$^{29\text{-}480}$ crystals did not result in diffraction at a sufficient resolution to solve the structure of MORC4$^{29\text{-}480}$, suggesting that a more powerful X-ray source such as high intensity synchrotron radiation (Lindley, 1995) may have been required to diffract the protein crystals at a higher resolution. Furthermore, other MORC4 fragments could be utilised in the future in

addition to wider crystal screen optimisations and reductive methylation to achieve a higher-resolution structure MORC4$^{29\text{-}480}$. Another issue faced throughout crystallisation trials was the limited quantity of available recombinant protein. Robotic handling systems would have helped to reduce the volumes of protein required for crystallisation trials, which would allow a greater range of crystallisation space to be screened with the limited amount of protein available.

Chemical crosslinking also indicated that wild type (WT) MORC4$^{29\text{-}480}$ could dimerise in presence of AMP-PNP (Figure 6.1). Whereas, several mutants were expected to prevent binding of ATP or AMP-PNP following similar mutations in MORC2, did not show any evidence of dimerisation, also suggesting that AMP-PNP (and hence, ATP) is required for MORC4 homodimerisation as with MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016). Although both of these experiments suggested that MORC4 has the potential to dimerise, it was not clear from the size exclusion chromatography-multiple angle laser light scattering (SEC-MALLS) and analytical SEC experiments with MORC4$^{29\text{-}480}$ in the presence of AMP-PNP. However, MORC2, demonstrated dimerisation using similar techniques (Douse *et al.*, 2018). SEC-MALLS suggested that MORC4$^{29\text{-}480}$ molecular weight was at least 12 times the known molecular weight of MORC4$^{29\text{-}480}$, suggesting that SEC-MALLS could not be used to accurately determine oligomeric status of MORC4. Although, the low amount of dimeric MORC4 formed in chemical crosslinking experiments suggests that dimerisation is not a tight interaction.

Analytical SEC accurately determined the molecular weight of the MORC4$^{29\text{-}480}$ monomer in the absence of AMP-PNP. However, in the presence of AMP-PNP the protein did not resolve at the expected molecular weight of 100 kDa, instead the protein was resolved at 50 kDa, suggesting that other methods may be required to detect this weak interaction with AMP-PNP. Other techniques such as analytical ultracentrifugation, which can be used to analyse the oligomeric status of proteins by investigating the sedimentation of biological molecules using

centrifugation (Lebowitz *et al.*, 2002), could be used to assess MORC4$^{29\text{-}480}$ oligomerisation in the presence of AMP-PNP/ATP in the future. Furthermore, native electrophoresis could have also been used to analyse the oligomeric status of MORC4$^{29\text{-}480}$, as native electrophoresis allows proteins to separate in their native state, leading to resolution of protein oligomers (Nowakowski *et al.*, 2014).

Despite the fact that the structure of MORC4 could not be solved over the course of this project, protein homology modelling allowed some interpretation of the predicted structure of MORC4 based on MORC2 and MORC3 structures. Multiple sequence alignments (MSAs) of human MORCs revealed that the ATPase domain is highly conserved across all human MORC paralogues. Homology modelling also demonstrated that the ATPase domain of MORC3 and MORC4 are structurally similar. However, the C-terminal domains of all human MORC paralogues were not conserved, suggesting that the C-terminal domain may have different roles and functions for each MORC paralogue. SWISS-MODEL homology modelling predicted that MORC4 could dimerise, as with MORC2 (Douse *et al.*, 2018) and MORC3 (Li *et al.*, 2016). The respective residues which bind to AMP-PNP in MORC3 were present in the MORC4 homology model, suggesting that MORC4 does bind to (and hydrolyse) ATP (Figure 6.1). Furthermore, the behaviour of the protein during crystallisation studies and chemical crosslinking also suggested that MORC4 dimerises in the presence of AMP-PNP. Homology modelling and biochemical characterisation of MORC4 gave an insight into the function of MORC4 and suggested that as with MORC2 and MORC3, MORC4 could also dimerise in the presence of the ligand AMP-PNP (Figure 6.1). This suggests that MORC4 also binds to ATP, leading to ATP hydrolysis. This may also lead to the hypothesis that MORC4 could utilise ATP hydrolysis to function as a chromatin remodeller, as with MORC2 (Xie *et al.*, 2019) and MORC3 (H Andrews *et al.*, 2016). However, further experiments would be required to determine if MORC4 functions as a chromatin remodeller. Chromatin association assays were

used to determine that MORC3 recruits chromatin, specifically, via H3K4me3 (H Andrews *et al.*, 2016; Zhang *et al.*, 2019), suggesting that MORC3 is an ATP-dependent chromatin remodeller. In addition, chromatin association assays could be used to investigate MORC4 binding to chromatin as well as ATP-dependent nucleosome remodelling assays to assess MORC4 chromatin remodelling activity (Chen *et al.*, 2014). ATPase assays would also be required to investigate the ATPase activity of MORC4, similar to those experiments utilised to confirm MORC2 (D Q Li *et al.*, 2012) and MORC3 (H Andrews *et al.*, 2016) as ATPases. Although an initial ATPase assay was optimised with a bacterial control DNA-dependent ATPase, (RecA) in this study it was not possible to investigate the ATPase activity of MORC4 due to time constraints.

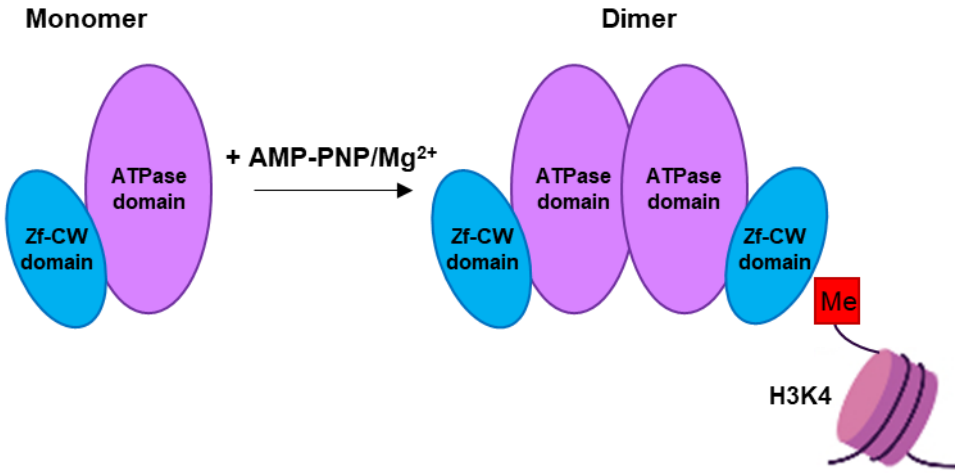


Figure 6.1: **Schematic representation of the predicted quaternary structure of MORC4 homodimerisation and the potential CW domain interaction with methylated H3K4.** MORC4 could bind to the of non-hydrolysable ligand, AMP-PNP/Mg^{2+}, leading to homodimerisation. The CW domain of MORC4 interacts with methylated H3K4. ATPase domains are shown as purple ovals; Zf-CW domains are shown as blue ovals and methylated histone is highlighted in red.

In addition, the MORC3 Zf-CW domain interacts with the H3K4me3 peptide, allowing MORC3 to bind to chromatin (H Andrews *et al.*, 2016). In an attempt to assess MORC4 as a chromatin remodeller and binding to methylated histones, isothermal titration calorimetry (ITC) assays could be utilised to investigate the strength of the binding interaction (Duff Jr *et*

al., 2011) between MORC4 protein and histone peptides, similar to MORC3 (Li *et al.*, 2016). A previous study has found that MORC4 CW domain is involved binding to histone tails, particularly methylated H3K4 (Figure 6.1) (Hoppmann *et al.*, 2011; Yanli Liu *et al.*, 2016), suggesting that MORC4 plays a role in binding of chromatin as with MORC2 and MORC3. In addition, the low expressing W435K mutation (W419K in MORC3) (H Andrews *et al.*, 2016), which prevents binding to H3K4me3, could be used as a negative control to investigate MORC4 binding activity to H3K4me3 using ITC. Overall, this study has identified that MORC4 likely binds ATP (AMP-PNP) to dimerise and that it has the required residues of an ATPase as with other MORC proteins. However, further evidence is required both structurally and biochemically to conclude that MORC4 functions as an ATPase, interacts with methylated histones and functions as a chromatin remodeller.

To investigate MORC4 function in disease, epitope-tagged MORC expression fragments were LIC cloned and transfected into human cell lines, in an attempt to verify MORC4 monoclonal antibodies. Transfected and epitope-tagged MORC4 could not be expressed in a number of human cell lines, which would be required to assess the suitability of MORC4 antibodies to study MORC4 expression by immunohistochemistry (IHC)/immunofluorescence (IF). Hence, verification of the novel monoclonal antibodies was performed using recombinant protein instead. Following this, both monoclonal antibodies and a commercial polyclonal antibody were used to detect expression MORC4 in cancer cell lines by western blotting.

MORC4 has previously been identified as a breast cancer oncogene, where overexpression of MORC4 protein was observed in MDA-MB-231 and MCF7 cell lines (Yang *et al.*, 2019). Overexpression of MORCR4 in breast cancer was also found to indicate a poor prognosis in clinical samples (Yang *et al.*, 2019). The previous study also identified high level of protein expression of MORC4 in both MDA-MB-231 and MCF7 cell lines as well as MDA-MB-468 (Yang *et al.*, 2019), supporting the hypothesis that MORC4 is a breast cancer oncogene.

However, this project only identified expression in the MDA-MB-468 breast cancer cell line, although qPCR data suggested expression in all breast cancer cell lines. In addition, it was observed here that colorectal carcinoma and some pancreatic carcinoma cell lines also expressed MORC4 protein at high levels. qPCR experiments also confirmed that MORC4 mRNA was expressed at high levels in breast adenocarcinoma, colorectal adenocarcinoma and pancreatic carcinoma cancer cell lines. These results suggested that MORC4 could be a potential biomarker of colorectal adenocarcinoma, pancreatic carcinoma and breast adenocarcinoma.

No prior studies have identified MORC4 as highly expressing in pancreatic cancer or colorectal cancer as yet. However, some evidence suggests that the X chromosomal linked Claudin 2 (CLDN2)-MORC4 locus has a single nucleotide polymorphism (SNP), which is associated with chronic pancreatitis (Giri *et al.*, 2016; M H Derikx *et al.*, 2015). RNA expression databases also revealed that high expression of MORC4 increases the chance of survival in patients with pancreatic cancer. This SNP and high expression of MORC4 protein in pancreatic carcinoma could suggest that MORC4 plays a role in pancreatic carcinoma. Similarly, some evidence also suggests that an SNP located on the CLDN2-MORC4 region is also associated with inflammatory bowel disease (Soderman *et al.*, 2013). In addition, analysis of RNA expression databases revealed that high expression of MORC4 in colon adenocarcinoma reduced the overall chance of survival in patients. This SNP, high expression of MORC4 protein in colon adenocarcinoma cells and survival analysis suggests that MORC4 could also play a role in pancreatic cancer. In addition, RNA expression of MORC4 was highest in ABC-DLBCL compared to GCB-DLBCL, supporting the previous study investigating MORC4 as a potential lymphoma biomarker (Liggins *et al.*, 2007). Furthermore, the anti-MORC4 commercial polyclonal antibody identified a higher level of protein expression in ABC-DLBCL compared to GCB-DLBCL cell lines, suggesting that the high RNA level correlates

with protein expression in these cell lines, further supporting MORC4 as a biomarker in DLBCL. Overall, MORC4 protein expression was found at higher levels in cell lines from breast adenocarcinoma, colorectal adenocarcinoma and pancreatic carcinoma, than other cancer cell lines. Lower expression of MORC4 was seen GCB-DLBCL, Burkitts and T-cell acute lymphoblastic leukaemia (T-ALL) lymphomas in addition to some neuroblastoma cell lines. In future, MORC4 protein expression could also be assessed using immunohistochemistry to investigate expression of MORC4 using both monoclonal and polyclonal antibodies (Duraiyan *et al.*, 2012) to aid characterisation of MORC4 expression patterns in clinical cancer patient samples. This would require the addition of the transfected epitope-tagged MORC4 expression to ensure it was the correct paralogue detected by the antibody, as a control. Further studies would include the assessment of such clinical sample studies, to identify if MORC4 plays a role in development of cancer in pancreatic, colorectal, breast cancers, and also specific subtypes of DLBCLs indicated here such as ABC-DLBCL. Such comparisons of MORC4 expression with prognostic or diagnostic factors (e.g. survival) performed such as IHC, which would for instance, require a haemato-oncology pathologist for DLBCL to assess staining patterns in the samples.

Overall, this study has contributed towards the characterisation of MORC4 from a number of perspectives, including structurally, biochemically and in relation to its RNA and protein expression in cancer cell lines. However, additional experiments will in future contribute to further understanding of MORC4 function and potential role in cancer development.

Chapter 7 References

Achilleos, A., Trainor, P.A. (2012) Neural crest stem cells: discovery, properties and potential for therapy. *Cell Research*. **22**(2), 288–304.

Ahel, D., Hořejší, Z., Wiechens, N., Polo, S.E., Garcia-Wilson, E., Ahel, I., Flynn, H., Skehel, M., West, S.C., Jackson, S.P., Owen-Hughes, T., Boulton, S.J. (2009) Poly(ADP-ribose)-dependent regulation of DNA repair by the chromatin remodeling enzyme ALC1. *Science*. **325**(5945), 1240–1243.

Al-Sukhni, W., Aronson, M., Gallinger, S. (2008) Hereditary Colorectal Cancer Syndromes: Familial Adenomatous Polyposis and Lynch Syndrome. *Surgical Clinics of North America*. **88**(4), 819–844.

Alfert, A., Moreno, N., Kerl, K. (2019) The BAF complex in development and disease. *Epigenetics and Chromatin*. **12**(1), 1–15.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson, W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., Staudt, L.M. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. **403**(6769), 503–511.

Alkhatib, S.G., Landry, J.W. (2011) The nucleosome remodeling factor. *FEBS letters*. **585**(20), 3197–3207.

Allec, N., Choi, M., Yesupriya, N., Szychowski, B., White, M.R., Kann, M.G., Garcin, E.D., Daniel, M.C., Badano, A. (2015) Small-angle X-ray scattering method to characterize molecular interactions: Proof of concept. *Scientific Reports*. **5**, 1–12.

Allen, H.F., Wade, P.A., Kutateladze, T.G. (2013) The NuRD architecture. *Cellular and Molecular Life Sciences*. **70**(19), 3513–3524.

Allfrey, V.G., Faulkner, R., Mirsky, A.E. (1964) Acetylation and Methylation of Histones and Their Possible Role in the. *Proceedings of the National Academy of Sciences of the United States of*. **51**(5), 786–794.

ALT, F.W., Blackwell, T.K., Depinho, R.A., Reth, M.G., Yancopoulos, G.D. (1986) Regulation of Genome Rearrangement Events during Lymphocyte Differentiation. *Immunological Reviews*. **89**(1), 5–30.

Altahan, A., Martin, M.G. (2018) Association between pesticides use and incidence of diffuse large B cell lymphoma (DLBCL). *Journal of Clinical Oncology*. **36**(15_suppl), 1570.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*. **215**(3), 403–410.

Amin, A.R.M.R., Karpowicz, P.A., Carey, T.E., Arbiser, J., Nahta, R., Chen, Z.G., Dong, J.T., Kucuk, O., Khan, G.N., Huang, G.S., Mi, S., Lee, H.Y., Reichrath, J., Honoki, K., Georgakilas, A.G., Amedei, A., Amin, A., Helferich, B., Boosani, C.S., Ciriolo, M.R., Chen, S., Mohammed, S.I., Azmi, A.S., Keith, W.N., Bhakta, D., Halicka, D., Niccolai, E., Fujii, H., Aquilano, K., Ashraf, S.S., Nowsheen, S., Yang, X., Bilsland, A., Shin, D.M. (2015) Evasion of anti-growth signaling: A key step in tumorigenesis and potential target for treatment and prophylaxis by natural compounds. *Seminars in Cancer Biology*. **35**, S55–S77.

Anders, C.K., Johnson, R., Litton, J., Phillips, M., Bleyer, A. (2009) Breast cancer before age 40 years. *Seminars in oncology*. **36**(3), 237–249.

Andrews, H, Tong, Q., Sullivan, K.D., Cornett, E.M., Zhang, Y., Ali, M., Ahn, J., Pandey, A., Guo, A.H., Strahl, B.D., Costello, J.C., Espinosa, J.M., Rothbart, S.B., Kutateladze, T.G. (2016) Multivalent Chromatin Engagement and Inter-domain Crosstalk Regulate MORC3 ATPase. *Cell Rep*. **16**(12), 3195–3207.

Armaghany, T., Wilson, J.D., Chu, Q., Mills, G. (2012) Genetic alterations in colorectal cancer.

Gastrointestinal cancer research : GCR. **5**(1), 19–27.
Arora, B., Tandon, R., Bhatia*, P.A. and R. (2017) Chemical Crosslinking: Role in Protein and Peptide Science. *Current Protein & Peptide Science*. **18**(9), 946–955.
Asgharian, A., Banan, M., Najmabadi, H. (2014) Optimizing A Lipocomplex-Based Gene Transfer Method into HeLa Cell Line. *Cell Journal (Yakhteh)*. **15**(4), 372–377.
Avvakumov, N., Côté, J. (2007) The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene*. **26**(37), 5395–5407.
Ayala, R., Willhoft, O., Aramayo, R.J., Wilkinson, M., McCormack, E.A., Ocloo, L., Wigley, D.B., Zhang, X. (2018) Structure and regulation of the human INO80–nucleosome complex. *Nature*. **556**(7701), 391–395.
Aydin, Ö.Z., Vermeulen, W., Lans, H. (2014) ISWI chromatin remodeling complexes in the DNA damage response. *Cell Cycle*. **13**(19), 3016–3025.
Badenhorst, P., Voas, M., Rebay, I., Wu, C. (2002) Biological functions of the ISWI chromatin remodeling complex NURF. *Genes and Development*. **16**(24), 3186–3198.
Banerjee, T., Chakravarti, D. (2011) A Peek into the Complex Realm of Histone Phosphorylation. *Molecular and Cellular Biology*. **31**(24), 4858 LP – 4873.
Bannister, A.J., Kouzarides, T. (2011) Regulation of chromatin by histone modifications. *Cell Research*. **21**(3), 381–395.
Bannister, A.J., Schneider, R., Kouzarides, T. (2002) Histone methylation: Dynamic or static? *Cell*. **109**(7), 801–806.
Bao, Y., Shen, X. (2009) INO80 subfamily of Chromatin Remodeling Complexes Yunhe. *Mutat. Res.*. **618**(1–2), 18–29.
Barneda-Zahonero, B., Parra, M. (2012) Histone deacetylases and cancer. *Molecular Oncology*. **6**(6), 579–589.
Barnum, K.J., O'Connell, M.J. (2014) Cell cycle regulation by checkpoints. *Methods in molecular biology (Clifton, N.J.)*. **1170**, 29–40.
Barthel, T.K., Zhang, J., Walker, G.C. (2001) ATPase-defective derivatives of Escherichia coli DnaK that behave differently with respect to ATP-induced conformational change and peptide release. *Journal of bacteriology*. **183**(19), 5482–5490.
Bartholomew, B. (2014) ISWI chromatin remodeling: one primary actor or a coordinated effort? *Current opinion in structural biology*. **24**, 150–155.
Basson, M.A., van Ravenswaaij-Arts, C. (2015) Functional Insights into Chromatin Remodelling from Studies on CHARGE Syndrome. *Trends in genetics : TIG*. **31**(10), 600–611.
Basta, J., Rauchman, M. (2009) The nucleosome remodeling deacetylase (NuRD) complex in Development and Disease. *Translational research : the journal of laboratory and clinical medicine*. **165**(May), 36–47.
Bauer, K.R., Brown, M., Cress, R.D., Parise, C.A., Caggiano, V. (2007) Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: A population-based study from the California Cancer Registry. *Cancer*. **109**(9), 1721–1728.
Bębenek, A., Ziuzia-Graczyk, I. (2018) Fidelity of DNA replication—a matter of proofreading. *Current Genetics*. **64**(5), 985–996.
Becker, S. (2015) A historic and scientific review of breast cancer: The next global healthcare challenge. *International Journal of Gynecology and Obstetrics*. **131**, S36–S39.
Beckwith, S.L., Schwartz, E.K., García-Nieto, P.E., King, D.A., Gowans, G.J., Wong, K.M., Eckley, T.L., Paraschuk, A.P., Peltan, E.L., Lee, L.R., Yao, W., Morrison, A.J. (2018) The INO80 chromatin remodeler sustains metabolic stability by promoting TOR signaling and regulating histone acetylation. *PLoS Genetics*. **14**(2), e1007216–e1007216.
Bell, C.E. (2005) Structure and mechanism of Escherichia coli RecA ATPase. *Molecular Microbiology*. **58**(2), 358–366.

Bellush, J.M., Whitehouse, I. (2017) DNA replication through a chromatin environment. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372**(1731).

Beneke, S. (2012) Regulation of chromatin structure by poly(ADP-ribosyl)ation. *Frontiers in Genetics*. **3**(SEP), 1–16.

Benkert, P., Tosatto, S.C.E., Schomburg, D. (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Genetics*. **71**(1), 261–277.

Berek, D. (2010) Size exclusion chromatography - A blessing and a curse of science and technology of synthetic polymers. *Journal of Separation Science*. **33**(3), 315–335.

Bergholtz, H., Lien, T.G., Swanson, D.M., Frigessi, A., Bathen, T.F., Borgen, E., Børresen-Dale, A.L., Engebråten, O., Garred, Ø., Geisler, J., Geitvik, G.A., Hartmann-Johnsen, O.J., Hofvind, S., Kristensen, V.N., Langerød, A., Lingjærde, O.C., Mælandsmo, G.M., Naume, B., Russnes, H., Sauer, T., Schlichting, E., Skjerven, H.K., Daidone, M.G., Tost, J., Wärnberg, F., Sørlie, T. (2020) Contrasting DCIS and invasive breast cancer by subtype suggests basal-like DCIS as distinct lesions. *npj Breast Cancer*. **6**(1), 1–9.

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T. (2017) Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*. **7**(1), 1–15.

Besson, C., Lancar, R., Prevot, S., Algarte-Genin, M., Delobel, P., Bonnet, F., Meyohas, M.-C., Partisani, M., Oberic, L., Gabarre, J., Goujard, C., Boue, F., Coppo, P., Costello, R., Hendel-Chavez, H., Mekerri, N., Dos Santos, G., Recher, C., Delarue, R., Casasnovas, R.-O., Taoufik, Y., Mounier, N., Costagliola, D., Cohort, the A.-C.L. (2017) Outcomes for HIV-associated diffuse large B-cell lymphoma in the modern combined antiretroviral therapy era. *AIDS*. **31**(18).

Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E., Gianni, L. (2016) Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature reviews. Clinical oncology*. **13**(11), 674–690.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T. (2014) SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*. **42**(W1), 252–258.

Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. (2017) The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Research*. **45**(D1), D313–D319.

Bird, A. (2001) Methylation Talk Between Histones and DNA. *Science*. **294**(5549), 2113 LP – 2115.

Blewitt, M.E., Gendrel, A.-V., Pang, Z., Sparrow, D.B., Whitelaw, N., Craig, J.M., Apedaile, A., Hilton, D.J., Dunwoodie, S.L., Brockdorff, N., Kay, G.F., Whitelaw, E. (2008) SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nature Genetics*. **40**(5), 663–669.

Bocangel, D., Sengupta, S., Mitra, S., Bhakat, K.K. (2009) p53-mediated down-regulation of the human DNA repair gene O 6-methylguanine-DNA methyltransferase (MGMT) via interaction with Sp1 transcription factor. *Anticancer Research*. **29**(10), 3741–3750.

Bojuwoye, M.O., Olokoba, A.B., Ogunlaja, O.A., Agodirin, S.O., Ibrahim, O.K., Okonkwo, K.C., Aliyu, A.M. (2018) Familial adenomatous polyposis syndrome with colorectal cancer in two Nigerians: a report of two cases and review of literature. *The Pan African medical journal*. **30**, 6.

Booth, W.T., Schlachter, C.R., Pote, S., Ussin, N., Mank, N.J., Klapper, V., Offermann, L.R., Tang, C., Hurlburt, B.K., Chruszcz, M. (2018) Impact of an N-terminal Polyhistidine Tag on Protein Thermal Stability. *ACS omega*. **3**(1), 760–768.

Bornhorst, J.A., Falke, J.J. (2000) Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology*. **326**, 245–254.

Bosse, K.R., Maris, J.M. (2016) Advances in the translational genomics of neuroblastoma: From improving risk stratification and revealing novel biology to identifying actionable genomic alterations. *Cancer*. **122**(1), 20–33.
Bottomley, M.J. (2004) Structures of protein domains that create or recognize histone modifications. *EMBO Rep*. **5**(5), 464–469.
Bradley, Christina Marchetti, Ronning, D.R., Ghirlando, R., Craigie, R., Dyda, F. (2005) Structural basis for DNA bridging by barrier-to-autointegration factor. *Nature Structural & Molecular Biology*. **12**(10), 935–936.
Bradley, Philip, Misura, K.M.S., Baker, D. (2005) Biochemistry: Toward high-resolution de novo structure prediction for small proteins. *Science*. **309**(5742), 1868–1871.
Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A. (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*. **68**(6), 394–424.
Brennan, B., Stiller, C., Bourdeaut, F. (2013) Extracranial rhabdoid tumours: What we have learned so far and future directions. *The Lancet Oncology*. **14**(8), e329–e336.
Brewer, H.R., Jones, M.E., Schoemaker, M.J., Ashworth, A., Swerdlow, A.J. (2017) Family history and risk of breast cancer: an analysis accounting for family structure. *Breast Cancer Research and Treatment*. **165**(1), 193–200.
Bronner, M.E. (2012) Formation and migration of neural crest cells in the vertebrate embryo. *Histochemistry and cell biology*. **138**(2), 179–186.
Buchan, D.W.A., Jones, D.T. (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research*. **47**(W1), W402–W407.
Bujanda, L., Cosme, A., Gil, I., Arenas-Mirave, J.I. (2010) Malignant colorectal polyps. *World journal of gastroenterology*. **16**(25), 3103–3111.
Bunnell, T.M., Burbach, B.J., Shimizu, Y., Ervasti, J.M. (2011) β-Actin specifically controls cell growth, migration, and the G-actin pool. *Molecular Biology of the Cell*. **22**(21), 4047–4058.
Butler, J.M. (2007) Short tandem repeat typing technologies used in human identity testing. *BioTechniques*. **43**(4).
Byrne, A.B., McWhirter, R.D., Sekine, Y., Strittmatter, S.M., Miller, D.M., Hammarlund, M. (2016) Inhibiting poly(ADP-ribosylation) improves axon regeneration. *eLife*. **5**(OCTOBER2016), 1–14.
Cai, Q., Medeiros, L.J., Xu, X., Young, K.H. (2015) MYC-driven aggressive B-cell lymphomas: biology, entity, differential diagnosis and clinical management. *Oncotarget*. **6**(36), 38591–38616.
Cao, J., Yan, Q. (2012) Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Frontiers in Oncology*. **2**(MAR), 1–9.
Carlson, S., Glass, K.C. (2014) The MOZ histone acetyltransferase in epigenetic signaling and disease. *Journal of cellular physiology*. **229**(11), 1571–1574.
Carter, A.P., Diamant, A.G., Urnavicius, L. (2016) How dynein and dynactin transport cargos: a structural perspective. *Current Opinion in Structural Biology*. **37**, 62–70.
Cavalli, G., Paro, R. (1998) Chrome-domain proteins: Linking chromatin structure to epigenetic regulation. *Current Opinion in Cell Biology*. **10**(3), 354–360.
Di Cerbo, V., Schneider, R. (2013) Cancers with wrong HATs: The impact of acetylation. *Briefings in Functional Genomics*. **12**(3), 231–243.
Chaikuad, A., Knapp, S., Von Delft, F. (2015) Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens. *Acta Crystallographica Section D: Biological Crystallography*. **71**, 1627–1639.
Chandy, M., Gutiérrez, J.L., Prochasson, P., Workman, J.L. (2006) SWI/SNF displaces SAGA-acetylated nucleosomes. *Eukaryotic Cell*. **5**(10), 1738–1747.
Chang, L., Takada, S. (2016) Histone acetylation dependent energy landscapes in tri-nucleosome

revealed by residue-resolved molecular simulations. *Scientific Reports*. **6**(September), 1–13.
Chen, J. (2016) The Cell-Cycle Arrest and Apoptotic and Progression. *Cold Spring Harbor perspectives in biology*, 1–16.
Chen, J., Herlong, F.H., Stroehlein, J.R., Mishra, L. (2016) Mutations of Chromatin Structure Regulating Genes in Human Malignancies. *Current protein & peptide science*. **17**(5), 411–437.
Chen, L., Cai, Y., Jin, G., Florens, L., Swanson, S.K., Washburn, M.P., Conaway, J.W., Conaway, R.C. (2011) Subunit organization of the human INO80 chromatin remodeling complex: An evolutionarily conserved core complex catalyzes ATP-dependent nucleosome remodeling. *Journal of Biological Chemistry*. **286**(13), 11283–11289.
Chen, L., Conaway, R.C., Conaway, J.W. (2013) Multiple modes of regulation of the human Ino80 SNF2 ATPase by subunits of the INO80 chromatin-remodeling complex. *Proceedings of the National Academy of Sciences*. **110**(51), 20497 LP – 20502.
Chen, L., Ooi, S.K., Conaway, J.W., Conaway, R.C. (2014) Biochemical assays for analyzing activities of ATP-dependent chromatin remodeling enzymes. *Journal of Visualized Experiments*. (92), 1–10.
Chen, X., Jensen, P.E. (2008) The role of B lymphocytes as antigen-presenting cells. *Archivum Immunologiae et Therapiae Experimentalis*. **56**(2), 77.
Chiraniya, A., Finkelstein, J., O'Donnell, M., Bloom, L.B. (2013) A novel function for the conserved glutamate residue in the Walker B motif of replication factor C. *Genes*. **4**(2), 134–151.
Chiu, L.Y., Gong, F., Miller, K.M. (2017) Bromodomain proteins: Repairing DNA damage within chromatin. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **372**(1731).
Chopra, V.S., Srinivasan, A., Kumar, R.P., Mishra, K., Basquin, D., Docquier, M., Seum, C., Pauli, D., Mishra, R.K. (2008) Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. *Developmental Biology*. **317**(2), 660–670.
Clapier, C.R., Iwasa, J., Cairns, B.R., Peterson, C.L. (2017) Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*. **18**(7), 407–422.
Cohen, S.A., Leininger, A. (2014) The genetic basis of Lynch syndrome and its implications for clinical practice and risk management. *The application of clinical genetics*. **7**, 147–158.
Concetta, G.M., Imbalzano, N.A. (2007) The Chd Family of Chromatin Remodelers Concetta. *The Chd Family of Chromatin Remodelers*. **605**(1), 105–112.
Cooper, C.D.O., Marsden, B.D. (2017) N- and C-Terminal Truncations to Enhance Protein Solubility and Crystallization: Predicting Protein Domain Boundaries with Bioinformatics Tools. In N. A. Burgess-Brown, ed. *Heterologous Gene Expression in E.coli: Methods and Protocols*. New York, NY: Springer New York, pp. 11–31.
Cooper, D.R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., Derewenda, Z. (2007) Protein crystallization by surface entropy reduction: Optimization of the SER strategy. *Acta Crystallographica Section D: Biological Crystallography*. **63**(5), 636–645.
Corbett, K.D., Berger, J.M. (2005) Structural dissection of ATP turnover in the prototypical GHL ATPase TopoVI. *Structure*. **13**(6), 873–882.
Corbett, K.D., Berger, J.M. (2003) Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *The EMBO Journal*. **22**(1), 151–163.
Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M., Stier, G. (2008) Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Science*. **17**(12), 2070–2079.
Costa, S., Almeida, A., Castro, A., Domingues, L. (2014) Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Frontiers in Microbiology*. **5**(FEB), 1–20.
Crowley, J.A., Wang, Y., Rapoport, A.P., Ning, Y. (2005) Detection of MOZ-CBP fusion in acute myeloid leukemia with 8;16 translocation [8]. *Leukemia*. **19**(12), 2344–2345.
Dalla-Favera, R. (2012) Lymphoid malignancies: Many tumor types, many altered genes, many

therapeutic challenges. *Journal of Clinical Investigation*. **122**(10), 3396–3397.
Datta, A., Brosh, R.M. (2018) New insights into DNA helicases as druggable targets for cancer therapy. *Frontiers in Molecular Biosciences*. **5**(JUN), 1–21.
Davidson, I. (2003) The genetics of TBP and TBP-related factors. *Trends in Biochemical Sciences*. **28**(7), 391–398.
Davis, R.E., Brown, K.D., Siebenlist, U., Staudt, L.M. (2001) Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *The Journal of experimental medicine*. **194**(12), 1861–1874.
Deb, D.K., Chen, Y., Sun, J., Wang, Y., Li, Y.C. (2017) ATP-citrate lyase is essential for high glucose-induced histone hyperacetylation and fibrogenic gene upregulation in mesangial cells. *American Journal of Physiology - Renal Physiology*. **313**(2), F423–F429.
Demeret, C., Vassetzky, Y., Méchali, M. (2001) Chromatin remodelling and DNA replication: From nucleosomes to loop domains. *Oncogene*. **20**(24 REV. ISS. 3), 3086–3093.
Derewenda, Z.S., Vekilov, P.G. (2006) Entropy and surface engineering in protein crystallization. *Acta Crystallographica Section D: Biological Crystallography*. **62**(1), 116–124.
Derikx, Monique H, Kovacs, P., Scholz, M., Masson, E., Chen, J.-M., Ruffert, C., Lichtner, P., M te Morsche, R.H., Martina Cavestro, G., Férec, C., H Drenth, J.P., Witt, H., Rosendahl, J. (2015) Polymorphisms at PRSS1–PRSS2 and CLDN2–MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study PanEuropean Working group on Alcoholic Chronic Pancreatitis members and collaborators. *Gut*. **64**, 1426–1433.
Derikx, M H, Kovacs, P., Scholz, M., Masson, E., Chen, J.M., Ruffert, C., Lichtner, P., Te Morsche, R.H., Cavestro, G.M., Ferec, C., Drenth, J.P., Witt, H., Rosendahl, J., PanEuropean Working group on Alcoholic Chronic Pancreatitis, M., Collaborators (2015) Polymorphisms at PRSS1-PRSS2 and CLDN2-MORC4 loci associate with alcoholic and non-alcoholic chronic pancreatitis in a European replication study. *Gut*. **64**(9), 1426–1433.
Dessau, M.A., Modis, Y. (2011) Protein Crystallization for X-ray Crystallography. *Journal of Visualized Experiments : JoVE*. (47), 2285.
Dey, P. (2006) Chromatin remodeling, cancer and chemotherapy. *Current medicinal chemistry*. **13**, 2909–2919.
Diamandopoulos, K., Green, J. (2018) Down syndrome: An integrative review. *Journal of Neonatal Nursing*. **24**(5), 235–241.
Dimitrova, D.G., Teysset, L., Carré, C. (2019) RNA 2′-O-Methylation (Nm) modification in human diseases. *Genes*. **10**(2), 1–23.
Dirks, W.G., Fähnrich, S., Estella, I.A.J., Drexler, H.G. (2005) Short tandem repeat DNA typing provides an international reference standard for authentication of human cell lines. *Altex*. **22**(2), 103–109.
Doherty, L.S., Kiely, J.L., Deegan, P.C., Nolan, G., McCabe, S., Green, A.J., Ennis, S., McNicholas, W.T. (2007) Late-onset central hypoventilation syndrome: a family genetic study. *European Respiratory Journal*. **29**(2), 312 LP – 316.
Dong, W., Vannozzi, A., Chen, F., Hu, Y., Chen, Z., Zhang, L. (2018) MORC domain definition and evolutionary analysis of the MORC gene family in green plants. *Genome Biology and Evolution*. **10**(7), 1730–1744.
Dong, X., Weng, Z. (2013) The correlation between histone modifications and gene expression. *Epigenomics*. **5**(2), 113–116.
Dorigo, B., Schalch, T., Kulangara, A., Duda, S., Schroeder, R.R., Richmond, T.J. (2004) Nucleosome Arrays Reveal the Two-Start Organization of the Chromatin Fiber. *Science*. **306**(5701), 1571 LP – 1573.
Dotan, E., Aggarwal, C., Smith, M.R. (2010) Impact of Rituximab (Rituxan) on the Treatment of B-Cell Non-Hodgkin's Lymphoma. *P & T : a peer-reviewed journal for formulary management*. **35**(3), 148–157.

Douse, C.., Bloor, S., Liu, Y., Shamin, M., Tchasovnikarova, I.A., Timms, R.T., Lehner, P.J., Modis, Y. (2018) Neuropathic MORC2 mutations perturb GHKL ATPase dimerization dynamics and epigenetic silencing by multiple structural mechanisms. *Nature Communications*. **9**(1).

Drazic, A., Myklebust, L.M., Ree, R., Arnesen, T. (2016) The world of protein acetylation. *Biochimica et Biophysica Acta - Proteins and Proteomics*. **1864**(10), 1372–1401.

Duan, G., Walther, D. (2015) The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Computational Biology*. **11**(2), 1–23.

Duff Jr, M.R., Grubbs, J., Howell, E.E. (2011) Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *Journal of visualized experiments : JoVE*. (55), 2796.

van Duin, M., Broyl, A., de Knegt, Y., Goldschmidt, H., Richardson, P.G., Hop, W.C., van der Holt, B., Joseph-Pietras, D., Mulligan, G., Neuwirth, R., Sahota, S.S., Sonneveld, P. (2011) Cancer testis antigens in newly diagnosed and relapse multiple myeloma: prognostic markers and potential targets for immunotherapy. *Haematologica*. **96**(11), 1662–1669.

Dunleavy, K. (2015) Aggressive B cell Lymphoma: Optimal Therapy for MYC-positive, Double-Hit, and Triple-Hit DLBCL. *Current Treatment Options in Oncology*. **16**(12), 58.

Dunleavy, K., Wilson, W.H. (2014) Appropriate management of molecular subtypes of diffuse large B-cell lymphoma.

Duraiyan, J., Govindarajan, R., Kaliyappan, K., Palanisamy, M. (2012) Applications of immunohistochemistry. *Journal of pharmacy & bioallied sciences*. **4**(Suppl 2), S307–S309.

Dürr, H., Flaus, A., Owen-Hughes, T., Hopfner, K.P. (2006) Snf2 family ATPases and DExx box helicases: Differences and unifying concepts from high-resolution crystal structures. *Nucleic Acids Research*. **34**(15), 4160–4167.

Dutta, R., Inouye, M. (2000) GHKL, an emergent ATPase/kinase superfamily. *Trends in Biochemical Sciences*. **25**(1), 24–28.

Dvinge, H., Bradley, R.K. (2015) Widespread intron retention diversifies most cancer transcriptomes. *Genome Medicine*. **7**(1), 1–13.

Eberharter, A., Becker, P.B. (2002) Histone acetylation: A switch between repressive and permissive chromatin. Second in review on chromatin dynamics. *EMBO Reports*. **3**(3), 224–229.

Edgar, R.C. (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. **32**(5), 1792–1797.

Einhauer, A., Jungbauer, A. (2001) The FLAG^{TM} peptide, a versatile fusion tag for the purification of recombinant proteins. *Journal of Biochemical and Biophysical Methods*. **49**(1–3), 455–465.

El-Aouar Filho, R.A., Nicolas, A., De Paula Castro, T.L., Deplanche, M., De Carvalho Azevedo, V.A., Goossens, P.L., Taieb, F., Lina, G., Le Loir, Y., Berkova, N. (2017) Corrigendum: Heterogeneous Family of Cyclomodulins: Smart Weapons That Allow Bacteria to Hijack the Eukaryotic Cell Cycle and Promote Infections. *Frontiers in cellular and infection microbiology*. **7**(May), 364.

Elfring, L.K., Deuring, R., McCallum, C.M., Peterson, C.L., Tamkun, J.W. (1994) Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. *Molecular and Cellular Biology*. **14**(4), 2225–2234.

Euskirchen, G., Auerbach, R.K., Snyder, M. (2012) SWI/SNF chromatin-remodeling factors: multiscale analyses and diverse functions. *The Journal of biological chemistry*. **287**(37), 30897–30905.

Fan, C.H., Liu, W.L., Cao, H., Wen, C., Chen, L., Jiang, G. (2013) O6-methylguanine DNA methyltransferase as a promising target for the treatment of temozolomide-resistant gliomas. *Cell Death and Disease*. **4**(10).

Farria, A., Li, W., Dent, S.Y.R. (2015) KATs in cancer: Functions and therapies. *Oncogene*. **34**(38), 4901–4913.

Ferguson, R.E., Carroll, H.P., Harris, A., Maher, E.R., Selby, P.J., Banks, R.E. (2005) Housekeeping proteins: A preliminary study illustrating some limitations as useful references in protein

expression studies. *Proteomics*. **5**(2), 566–571.
Fernandez-Capetillo, O., Lee, A., Nussenzweig, M., Nussenzweig, A. (2004) H2AX: The histone guardian of the genome. *DNA Repair*. **3**(8–9), 959–967.
Fields, P., Wrench, D. (2017) Hodgkin lymphoma. *Medicine*. **45**(5), 305–310.
Fiorentino, D.F., Chung, L.S., Christopher-Stine, L., Zaba, L., Li, S., Mammen, A.L., Rosen, A., Casciola-Rosen, L. (2013) Most patients with cancer-associated dermatomyositis have antibodies to nuclear matrix protein NXP-2 or transcription intermediary factor 1γ. *Arthritis and rheumatism*. **65**(11), 2954–2962.
Fisher, R.I., Gaynor, E.R., Dahlberg, S., Oken, M.M., Grogan, T.M., Mize, E.M., Glick, J.H., Coltman, C.A.J., Miller, T.P. (1993) Comparison of a Standard Regimen (CHOP) with Three Intensive Chemotherapy Regimens for Advanced Non-Hodgkin's Lymphoma. *New England Journal of Medicine*. **328**(14), 1002–1006.
Fleming, M., Ravula, S., Tatishchev, S.F., Wang, H.L. (2012) Colorectal carcinoma: Pathologic aspects. *Journal of Gastrointestinal Oncology*. **3**(3), 153–173.
Fletcher-Etherington, A., Nightingale, K., Lin, K.-M., Ravenhill, B., Nobre, L., Weekes, M. (2019) Determining the function of proteins degraded by Human Cytomegalovirus. *Access Microbiology*. **1**(1A).
Foster, I. (2008) Cancer: A cell cycle defect. *Radiography*. **14**(2), 144–149.
Frick, D.N., Banik, S., Rypma, R.S. (2007) Role of divalent metal cations in ATP hydrolysis catalyzed by the hepatitis C virus NS3 helicase: magnesium provides a bridge for ATP to fuel unwinding. *Journal of molecular biology*. **365**(4), 1017–1032.
Friedberg, J.W., Fisher, R.I. (2008) Diffuse large B-cell lymphoma. *Hematology/oncology clinics of North America*. **22**(5), 941–ix.
Fulda, S. (2010) Evasion of apoptosis as a cellular stress response in cancer. *International Journal of Cell Biology*. **2010**.
Gao, Y., Teschendorff, A.E. (2017) Epigenetic and genetic deregulation in cancer target distinct signaling pathway domains. *Nucleic Acids Research*. **45**(2), 583–596.
Gataulin, D. V., Carey, J.N., Li, J., Shah, P., Grubb, J.T., Bishop, D.K. (2018) The ATPase activity of E. Coli RecA prevents accumulation of toxic complexes formed by erroneous binding to undamaged double stranded DNA. *Nucleic Acids Research*. **46**(18), 9510–9523.
Geiman, T.M., Robertson, K.D. (2002) Chromatin remodeling, histone modifications, and DNA methylation - How does it all fit together? *Journal of Cellular Biochemistry*. **87**(2), 117–125.
Geller, J.I., Roth, J.J., Biegel, J.A. (2015) Biology and treatment of rhabdoid tumor. *Critical Reviews in Oncogenesis*. **20**(3–4), 199–216.
Gilabert, M., Launay, S., Ginestier, C., Bertucci, F., Audebert, S., Pophillat, M., Toiron, Y., Baudelet, E., Finetti, P., Noguchi, T., Sobol, H., Birnbaum, D., Borg, J.P., Charafe-Jauffret, E., Gonçalves, A. (2014) Poly(ADP-Ribose) polymerase 1 (PARP1) overexpression in human breast cancer stem cells and resistance to olaparib. *PLoS ONE*. **9**(8), 6–11.
Giono, L.E., Manfredi, J.J. (2006) The p53 tumor suppressor participates in multiple cell cycle checkpoints. *Journal of Cellular Physiology*. **209**(1), 13–20.
Giri, A.K., Midha, S., Banerjee, P., Agrawal, A., Mehdi, S.J., Dhingra, R., Kaur, I., G, R.K., Lakhotia, R., Ghosh, S., Das, K., Mohindra, S., Rana, S., Bhasin, D.K., Garg, P.K., Bharadwaj, D., Indipan, Consortium, I. (2016) Common Variants in CLDN2 and MORC4 Genes Confer Disease Susceptibility in Patients with Chronic Pancreatitis. *PLoS One*. **11**(1), e0147345.
Gisselbrecht, C., Glass, B., Mounier, N., Singh Gill, D., Linch, D.C., Trneny, M., Bosly, A., Ketterer, N., Shpilberg, O., Hagberg, H., Ma, D., Brière, J., Moskowitz, C.H., Schmitz, N. (2010) Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. **28**(27), 4184–4190.
Gnanapragasam, M.N., Scarsdale, J.N., Amaya, M.L., Webb, H.D., Desai, M.A., Walavalkar, N.M.,

Wang, S.Z., Zu Zhu, S., Ginder, G.D., Williams Jr, D.C. (2011) p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proceedings of the National Academy of Sciences of the United States of America*. **108**(18), 7487–7492.

Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E. (2018) UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science*. **27**(1), 14–25.

Gong, F., Miller, K.M. (2013) Mammalian DNA repair: HATs and HDACs make their mark through histone acetylation. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis*. **750**(1–2), 23–30.

Goni, R., García, P., Foissac, S. (2009) The qPCR data statistical analysis. *Control*. **1**(September), 1–9.

Gonneaud, A., Turgeon, N., Jones, C., Couture, C., Lévesque, D., Boisvert, F.-M., Boudreau, F., Asselin, C. (2019) HDAC1 and HDAC2 independently regulate common and specific intrinsic responses in murine enteroids. *Scientific Reports*. **9**(1), 5363.

Gonzalez-Fernandez, R., Morales, M., Avila, J., Martin-Vasallo, P. (2012) Changes in leukocyte gene expression profiles induced by antineoplastic chemotherapy. *Oncol Lett*. **3**(6), 1341–1349.

Gordon, C.T., Xue, S., Yigit, G., Filali, H., Chen, K., Rosin, N., Yoshiura, K., Oufadem, M., Beck, T.J., McGowan, R., Magee, A.C., Altmüller, J., Dion, C., Thiele, H., Gurzau, A.D., Nürnberg, P., Meschede, D., Mühlbauer, W., Okamoto, N., Varghese, V., Irving, R., Sigaudy, S., Williams, D., Ahmed, S.F., Bonnard, C., Kong, M.K., Ratbi, I., Fejjal, N., Fikri, M., Elalaoui, S.C., Reigstad, H., Bole-Feysot, C., Nitschké, P., Ragge, N., Lévy, N., Tunçbilek, G., Teo, A.S.M., Cunningham, M.L., Sefiani, A., Kayserili, H., Murphy, J.M., Chatdokmaiprai, C., Hillmer, A.M., Wattanasirichaigoon, D., Lyonnet, S., Magdinier, F., Javed, A., Blewitt, M.E., Amiel, J., Wollnik, B., Reversade, B. (2017) De novo mutations in SMCHD1 cause Bosma arhinia microphthalmia syndrome and abrogate nasal development. *Nature Genetics*. **49**(2), 249–255.

Gorrec, F. (2009) The MORPHEUS protein crystallization screen. *Journal of Applied Crystallography*. **42**(6), 1035–1042.

Gorrec, F., Löwe, J. (2018) Automated protocols for macromolecular crystallization at the MRC laboratory of molecular biology. *Journal of Visualized Experiments*. **2018**(131), 1–18.

Goswami, C.P., Nakshatri, H. (2013) PROGgene: Gene expression based survival analysis web application for multiple cancers. *Journal of Clinical Bioinformatics*. **3**(1), 1–9.

Gottschalk, A.J., Trivedi, R.D., Conaway, J.W., Conaway, R.C. (2012) Activation of the SNF2 family ATPase ALC1 by poly(ADP-ribose) in a stable ALC1.PARP1.nucleosome intermediate. *J Biol Chem*. **287**(52), 43527–43532.

Gouveia, G.R., Siqueira, S.A.C., Pereira, J. (2012) Pathophysiology and molecular aspects of diffuse large B-cell lymphoma. *Revista Brasileira de Hematologia e Hemoterapia*. **34**(6), 447–451.

Granchi, C. (2018) ATP citrate lyase (ACLY) inhibitors: An anti-cancer strategy at the crossroads of glucose and lipid metabolism. *European Journal of Medicinal Chemistry*. **157**, 1276–1291.

Graslund, S., Nordlund, P., Weigelt, J., Hallberg, B.M., Bray, J., Gileadi, O., Knapp, S., Oppermann, U., Arrowsmith, C., Hui, R., Ming, J., Dhe-Paganon, S., Park, H.W., Savchenko, A., Yee, A., Edwards, A., Vincentelli, R., Cambillau, C., Kim, R., Kim, S.H., Rao, Z., Shi, Y., Terwilliger, T.C., Kim, C.Y., Hung, L.W., Waldo, G.S., Peleg, Y., Albeck, S., Unger, T., Dym, O., Prilusky, J., Sussman, J.L., Stevens, R.C., Lesley, S.A., Wilson, I.A., Joachimiak, A., Collart, F., Dementieva, I., Donnelly, M.I., Eschenfeldt, W.H., Kim, Y., Stols, L., Wu, R., Zhou, M., Burley, S.K., Emtage, J.S., Sauder, J.M., Thompson, D., Bain, K., Luz, J., Gheyi, T., Zhang, F., Atwell, S., Almo, S.C., Bonanno, J.B., Fiser, A., Swaminathan, S., Studier, F.W., Chance, M.R., Sali, A., Acton, T.B., Xiao, R., Zhao, L., Ma, L.C., Hunt, J.F., Tong, L., Cunningham, K., Inouye, M., Anderson, S., Janjua, H., Shastry, R., Ho, C.K., Wang, D., Wang, H., Jiang,

M., Montelione, G.T., Stuart, D.I., Owens, R.J., Daenke, S., Schutz, A., Heinemann, U., Yokoyama, S., Bussow, K., Gunsalus, K.C. (2008) Protein production and purification. *Nat Methods*. **5**(2), 135–146.

Greenbaum, D., Colangelo, C., Williams, K., Gerstein, M. (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biology*. **4**(9).

Greer, E.L., Shi, Y. (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nature reviews. Genetics*. **13**(5), 343–357.

Groth, A., Rocha, W., Verreault, A., Almouzni, G. (2007) Chromatin Challenges during DNA Replication and Repair. *Cell*. **128**(4), 721–733.

Guex, N., Peitsch, M.C., Schwede, T. (2009) Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis*. **30**(SUPPL. 1), S162–S173.

Gunawardena, H., Wedderburn, L.R., Chinoy, H., Betteridge, Z.E., North, J., Ollier, W.E.R., Cooper, R.G., Oddis, C. V., Ramanan, A. V., Davidson, J.E., McHugh, N.J. (2009) Autoantibodies to a 140-kd protein in juvenile dermatomyositis are associated with calcinosis. *Arthritis and Rheumatism*. **60**(6), 1807–1814.

Haberland, M., Montgomery, R.L., Olson, E.N. (2009) The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature reviews. Genetics*. **10**(1), 32–42.

Haggar, F.A., Boushey, R.P. (2009) Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clinics in colon and rectal surgery*. **22**(4), 191–197.

Hanahan, D., Weinberg, R.A. (2011) Hallmarks of cancer: The next generation. *Cell*. **144**(5), 646–674.

Harnett, M.M., Katz, E., Ford, C.A. (2005) Differential signalling during B-cell maturation. *Immunology Letters*. **98**(1), 33–44.

Hauk, G., McKnight, J.N., Nodelman, I.M., Bowman, G.D. (2010) The Chromodomains of the Chd1 Chromatin Remodeler Regulate DNA Access to the ATPase Motor. *Molecular Cell*. **39**(5), 711–723.

Hayden, M.S., Ghosh, S. (2004) Signaling to NF-κB. *Genes & Development* . **18**(18), 2195–2224.

He, F., Umehara, T., Saito, K., Harada, T., Watanabe, S., Yabuki, T., Kigawa, T., Takahashi, M., Kuwasako, K., Tsuda, K., Matsuda, T., Aoki, M., Seki, E., Kobayashi, N., Güntert, P., Yokoyama, S., Muto, Y. (2010) Structural insight into the zinc finger CW domain as a histone modification reader. *Structure*.

Hecker, K.H., Roux, K.H. (1996) High and low annealing temperatures increase both specificity and yield in touchdown and stepdown PCR. *BioTechniques*. **20**(3), 478–485.

Hedhammar, M., Nilvebrant, J., Hober, S. (2014) Zbasic: A Purification Tag for Selective Ion-Exchange Recovery BT - Protein Downstream Processing: Design, Development and Application of High and Low-Resolution Methods. In N. E. Labrou, ed. Totowa, NJ: Humana Press, pp. 197–204.

Henry, N.L., Hayes, D.F. (2012) Cancer biomarkers. *Molecular Oncology*. **6**(2), 140–146.

Herlyn, H. (2016) The phylogenetic system of primates—character evolution in the light of a consolidated tree. *Organisms Diversity & Evolution*. **16**(4), 689–713.

Hirano, T. (2006) At the heart of the chromosome: SMC proteins in action. *Nature Reviews Molecular Cell Biology*. **7**(5), 311–322.

Hoffman, B., Liebermann, D.A. (2008) Apoptotic signaling by c-MYC. *Oncogene*. **27**(50), 6462–6472.

Hoffman, W., Lakkis, F.G., Chalasani, G. (2016) B cells, antibodies, and more. *Clinical Journal of the American Society of Nephrology*. **11**(1), 137–154.

Hohl, M., Kwon, Y., Galván, S.M., Xue, X., Tous, C., Aguilera, A., Sung, P., Petrini, J.H.J. (2011) The Rad50 Coiled Coil Domain is indispensable for Mre11 complex functions. *Nature*

structural & molecular biology. **18**(10), 1124–1131.
Holcomb, J., Spellmon, N., Zhang, Y., Doughan, M., Li, C., Yang, Z. (2017) Protein crystallization: Eluding the bottleneck of X-ray crystallography. *AIMS Biophysics*. **4**(4), 557–575.
van Holde, K., Zlatanova, J. (2007) Chromatin fiber structure: Where is the problem now? *Seminars in Cell and Developmental Biology*. **18**(5), 651–658.
Hong, G., Qiu, H., Wang, C., Jadhav, G., Wang, H., Tickner, J., He, W., Xu, J. (2017) The Emerging Role of MORC Family Proteins in Cancer Development and Bone Homeostasis. *Journal of Cellular Physiology*. **232**(5), 928–934.
Hopfner, K.-P., Gerhold, C., Lakomek, K., Wollmann, P. (2012) Swi2/Snf2 remodelers: hybrid views on hybrid molecular Curr Opin Struct Biol . 2012 April ; 22(2): 225–233. doi:10.1016/j.sbi.2012.02.007. machines Karl-Peter. *Curr Opin Struct Biol*. **22**(2), 225–233.
Hoppmann, V., Thorstensen, T., Kristiansen, P.E., Veiseth, S. V, Rahman, M.A., Finne, K., Aalen, R.B., Aasland, R. (2011) The CW domain, a new histone recognition module in chromatin proteins. *EMBO J*. **30**(10), 1939–1952.
Horchani, H., Ouertani, S., Gargouri, Y., Sayari, A. (2009) The N-terminal His-tag and the recombination process affect the biochemical properties of Staphylococcus aureus lipase produced in Escherichia coli. *Journal of Molecular Catalysis B: Enzymatic*. **61**(3–4), 194–201.
Hota, S.K., Bruneau, B.G. (2016) ATP-dependent chromatin remodeling during mammalian development. *Development (Cambridge)*. **143**(16), 2882–2897.
Hu, X., Machius, M., Yang, W. (2003) Monovalent cation dependence and preference of GHKL ATPases and kinases. *FEBS Letters*. **544**(1–3), 268–273.
Hur, S.-K., Park, E.-J., Han, J.-E., Kim, Y.-A., Kim, J.-D., Kang, D., Kwon, J. (2010) Roles of human INO80 chromatin remodeling enzyme in DNA replication and chromosome segregation suppress genome instability. *Cellular and Molecular Life Sciences*. **67**(13), 2283–2296.
Hyun, K., Jeon, J., Park, K., Kim, J. (2017) Writing, erasing and reading histone lysine methylations. *Experimental and Molecular Medicine*. **49**(4).
Imoberdorf, R.M., Topalidou, I., Strubin, M. (2006) A Role for Gcn5-Mediated Global Histone Acetylation in Transcriptional Regulation. *Molecular and Cellular Biology*. **26**(5), 1610–1616.
Iqbal, J., Naushad, H., Bi, C., Yu, J., Bouska, A., Rohr, J., Chao, W., Fu, K., Chan, W.C., Vose, J.M. (2016) Genomic signatures in B-cell lymphoma: How can these improve precision in diagnosis and inform prognosis? *Blood Reviews*. **30**(2), 73–88.
Iyer, L.M., Abhiman, S., Aravind, L. (2008) MutL homologs in restriction-modification systems and the origin of eukaryotic MORC ATPases. *Biol Direct*. **3**, 8.
Iyer, N.G., Özdag, H., Caldas, C. (2004) p300/CBP and cancer. *Oncogene*. **23**(24), 4225–4231.
Jacob, A.G., Smith, C.W.J. (2017) Intron retention as a component of regulated gene expression programs. *Human Genetics*. **136**(9), 1043–1057.
Jacobson, S., Pillus, L. (2004) Molecular Requirements for Gene Expression Mediated by Targeted Histone Acetyltransferases. *Molecular and Cellular Biology*. **24**(13), 6029–6039.
Jagadeeshan, S., Subramanian, A., Tentu, S., Beesetti, S., Singhal, M., Raghavan, S., Surabhi, R.P., Mavuluri, J., Bhoopalan, H., Biswal, J., Pitani, R.S., Chidambaram, S., Sundaram, S., Malathi, R., Jeyaraman, J., Nair, A.S., Venkatraman, G., Rayala, S.K. (2016) P21-activated kinase 1 (Pak1) signaling influences therapeutic outcome in pancreatic cancer. *Annals of Oncology*. **27**(8), 1546–1556.
Jamieson, C.H., Weissman, I.L., Passegue, E. (2004) Chronic versus acute myelogenous leukemia: a question of self-renewal. *Cancer Cell*. **6**(6), 531–533.
Janas, E., Hofacker, M., Chen, M., Gompf, S., Van der Does, C., Tampé, R. (2003) The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p. *Journal of Biological Chemistry*. **278**(29), 26862–26869.
Jancarik, J., Kim, S.H. (1991) Sparse matrix sampling. A screening method for crystallization of proteins. *Journal of Applied Crystallography*. **24**(pt 4), 409–411.

Jha, J.K., Li, M., Ghirlando, R., Miller Jenkins, L.M., Wlodawer, A., Chattoraj, D. (2017) The DnaK chaperone uses different mechanisms to promote and inhibit replication of vibrio cholerae chromosome 2 G. M. M. Dunny Didier Giraldo, Rafael, ed. *mBio*. **8**(2), e00427-17.

Ji, Y., Tulin, A. V (2010) The roles of PARP1 in gene control and cell differentiation. *Current opinion in genetics & development*. **20**(5), 512–518.

Jia, B., Jeon, C.O. (2016) High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. *Open biology*. **6**(8), 160196.

Jia, B., Zhao, X., Wang, Yao, Wang, J., Wang, Yingying, Yang, Y. (2019) Prognostic roles of MAGE family members in breast cancer based on KM-Plotter Data. *Oncology Letters*. **18**(4), 3501–3516.

Jiang, B.H., Chen, W.Y., Li, H.Y., Chien, Y., Chang, W.C., Hsieh, P.C., Wu, P., Chen, C.Y., Song, H.Y., Chien, C.S., Sung, Y.J., Chiou, S.H. (2015) CHD1L Regulated PARP1-Driven Pluripotency and Chromatin Remodeling during the Early-Stage Cell Reprogramming. *Stem Cells*. **33**(10), 2961–2972.

Jin, J., Cai, Y., Li, B., Conaway, R.C., Workman, J.L., Conaway, J.W., Kusch, T. (2005) In and out: Histone variant exchange in chromatin. *Trends in Biochemical Sciences*. **30**(12), 680–687.

Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *Journal of Molecular Biology*. **292**(2), 195–202.

Kagalwala, M.N., Glaus, B.J., Dang, W., Zofall, M., Bartholomew, B. (2004) Topography of the ISW2-nucleosome complex: insights into nucleosome spacing and chromatin remodeling. *The EMBO journal*. **23**(10), 2092–2104.

Kamel, H.F.M., Al-Amodi, H.S.A.B. (2017) Exploitation of Gene Expression and Cancer Biomarkers in Paving the Path to Era of Personalized Medicine. *Genomics, Proteomics and Bioinformatics*. **15**(4), 220–235.

Kamińska, M., Ciszewski, T., Łopacka-Szatan, K., Miotła, P., Starosławska, E. (2015) Breast cancer risk factors. *Przeglad Menopauzalny*. **14**(3), 196–202.

Kanwal, R., Gupta, S. (2012) Epigenetic modifications in cancer. *Clinical Genetics*. **81**(4), 303–311.

Kapust, R.B., Tözsér, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., Waugh, D.S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Engineering, Design and Selection*. **14**(12), 993–1000.

Kawamura, Y., Arakawa, K., Maeshima, M., Yoshida, S. (2001) ATP analogue binding to the A subunit induces conformational changes in the E subunit that involves a disulfide bond formation in plant V-ATPase. *European Journal of Biochemistry*. **268**(10), 2801–2809.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*. **10**(6), 845–858.

Kelly, B.J., Fraefel, C., Cunningham, A.L., Diefenbach, R.J. (2009) Functional roles of the tegument proteins of herpes simplex virus type 1. *Virus Research*. **145**(2), 173–186.

Keppler, B.R., Archer, T.K. (2008) Chromatin-modifying enzymes as therapeutic targets--Part 1. *Expert opinion on therapeutic targets*. **12**(10), 1301–1312.

Khodadadi, L., Cheng, Q., Radbruch, A., Hiepe, F. (2019) The Maintenance of Memory Plasma Cells . *Frontiers in Immunology* . **10**, 721.

Kiianitsa, K., Solinger, J.A., Heyer, W.D. (2003) NADH-coupled microplate photometric assay for kinetic studies of ATP-hydrolyzing enzymes with low and high specific activities. *Analytical Biochemistry*. **321**(2), 266–271.

Kim, D.E., Chivian, D., Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research*. **32**(WEB SERVER ISS.), 526–531.

Kim, H.J., Yen, L., Wongpalee, S.P., Kirshner, J.A., Mehta, N., Xue, Y., Johnston, J.B., Burlingame, A.L., Kim, J.K., Loparo, J.J., Jacobsen, S.E. (2019) The Gene-Silencing Protein MORC-1 Topologically Entraps DNA and Forms Multimeric Assemblies to Cause DNA Compaction. *Molecular Cell*. **75**(4), 700-710.e6.

Kim, J.-S., DeGiovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R., Kim, S.-H. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proceedings of the National Academy of Sciences of the United States of America*. **102**(9), 3248–3253.

Kim, Y., Quartey, P., Li, H., Volkart, L., Hatzos, C., Chang, C., Nocek, B., Cuff, M., Osipiuk, J., Tan, K., Fan, Y., Bigelow, L., Maltseva, N., Wu, R., Borovilos, M., Duggan, E., Zhou, M., Binkowski, T.A., Zhang, R., Joachimiak, A. (2008) Large-scale evaluation of protein reductive methylation for improving protein crystallization. *Nature methods*. **5**(10), 853–854.

Kimple, M.E., Brill, A.L., Pasker, R.L. (2013) Overview of affinity tags for protein purification. *Current Protocols in Protein Science*. (SUPPL.73), 608–616.

Klein, U., Tu, Y., Stolovitzky, G.A., Keller, J.L., Haddad Jr., J., Miljkovic, V., Cattoretti, G., Califano, A., Dalla-Favera, R. (2003) Transcriptional analysis of the B cell germinal center reaction. *Proc Natl Acad Sci U S A*. **100**(5), 2639–2644.

Ko, H.L., Ren, E.C. (2012) Functional aspects of PARP1 in DNA repair and transcription. *Biomolecules*. **2**(4), 524–548.

Koch, A., Kang, H.G., Steinbrenner, J., Dempsey, D.A., Klessig, D.F., Kogel, K.H. (2017) MORC proteins: Novel players in plant and animal health. *Frontiers in Plant Science*. **8**(October).

Koetsier, G., Cantor, E. (2019) A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers. *New England Biolabs*. **1**(1), 1–8.

Korzus, E., Rosenfeld, M.G., Mayford, M. (2004) CBP histone acetyltransferase activity is a critical component of memory consolidation. *Neuron*. **42**(6), 961–972.

Koschmann, C., Nunez, F.J., Mendez, F., Brosnan-Cashman, J.A., Meeker, A.K., Lowenstein, P.R., Castro, M.G. (2017) Mutated Chromatin Regulatory Factors as Tumor Drivers in Cancer. *Cancer research*. **77**(2), 227–233.

Kozera, B., Rapacz, M. (2013) Reference genes in real-time PCR. *Journal of Applied Genetics*. **54**(4), 391–406.

Kram, K.E., Finkel, S.E. (2015) Rich Medium Composition Affects Escherichia coli Survival, Glycation, and Mutation Frequency during Long-Term Batch Culture. *Applied and environmental microbiology*. **81**(13), 4442–4450.

Kruhlak, M.J., Celeste, A., Dellaire, G., Fernandez-Capetillo, O., Müller, W.G., McNally, J.G., Bazett-Jones, D.P., Nussenzweig, A. (2006) Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *Journal of Cell Biology*. **172**(6), 823–834.

Kufareva, I., Abagyan, R. (2012) Methods of protein structure comparison. *Methods in molecular biology (Clifton, N.J.)*. **857**, 231–257.

Kuipers, E.J., Grady, W.M., Lieberman, D., Seufferlein, T., Sung, J.J., Boelens, P.G., van de Velde, C.J.H., Watanabe, T. (2015) Colorectal cancer. *Nature reviews. Disease primers*. **1**, 15065.

Kukurba, K.R., Montgomery, S.B. (2015) RNA Sequencing and Analysis. *Cold Spring Harbor protocols*. **2015**(11), 951–969.

Kumar, R., Gururaj, A.E., Barnes, C.J. (2006) P21-Activated Kinases in Cancer. *Nature Reviews Cancer*. **6**(6), 459–471.

Kumar, R., Wang, R.-A. (2016) Structure, expression and functions of MTA genes. *Gene*. **582**(2), 112–121.

Kurdistani, S.K., Grunstein, M. (2003) Histone acetylation and deacetylation in yeast. *Nature Reviews Molecular Cell Biology*. **4**(4), 276–284.

Kurnik, R.T., Yu, A.W., Blank, G.S., Burton, A.R., Smith, D., Athalye, A.M., van Reis, R. (1995) Buffer exchange using size exclusion chromatography, countercurrent dialysis, and tangential flow filtration: Models, development, and industrial application. *Biotechnology and Bioengineering*. **45**(2), 149–157.

Kurosaki, T. (2010) B-lymphocyte biology. *Immunological Reviews*. **237**(1), 5–9.

Kwon, T., Chang, J.H., Kwak, E., Lee, C.W., Joachimiak, A., Kim, Y.C., Lee, J.W., Cho, Y. (2003)

Mechanism of histone lysine methyl transfer revealed by the structure of SET7/9-AdoMet. *EMBO Journal*. **22**(2), 292–303.
Lai, A.Y., Wade, P.A. (2011) NuRD: A multi-faceted chromatin remodeling complex in regulating cancer biology. *Nat Rev Cancer*. **11**(8), 588–596.
Laing, S., Unger, M., Koch-Nolte, F., Haag, F. (2011) ADP-ribosylation of arginine. *Amino Acids*. **41**(2), 257–269.
Lam, L.T., Davis, R.E., Pierce, J., Hepperle, M., Xu, Y., Hottelet, M., Nong, Y., Wen, D., Adams, J., Dang, L., Staudt, L.M. (2005) Small molecule inhibitors of IκB kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clinical Cancer Research*. **11**(1), 28–40.
Lam, S.D., Das, S., Sillitoe, I., Orengo, C. (2017) An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. *Acta Crystallographica Section D: Structural Biology*. **73**(8), 628–640.
Langelier, M.F., Servent, K.M., Rogers, E.E., Pascal, J.M. (2008) A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. *J Biol Chem*. **283**(7), 4105–4114.
Längst, G., Manelyte, L. (2015) Chromatin remodelers: From function to dysfunction. *Genes*. **6**(2), 299–324.
Lans, H., Marteijn, J.A., Vermeulen, W. (2012) ATP-dependent chromatin remodeling in the DNA-damage response. *Epigenetics & Chromatin*. **5**, 4.
Laurent, C., Do, C., Gourraud, P.A., De Paiva, G.R., Valmary, S., Brousset, P. (2015) Prevalence of common non-hodgkin lymphomas and subtypes of hodgkin lymphoma by nodal site of involvement. *Medicine (United States)*. **94**(25), e987.
LeBien, T.W., Tedder, T.F. (2008) B lymphocytes: how they develop and function. *Blood*. **112**(5), 1570–1580.
Lebowitz, J., Lewis, M.S., Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science : a publication of the Protein Society*. **11**(9), 2067–2079.
Lee, A.M., Wigle, T.J., Singleton, S.F. (2007) A complementary pair of rapid molecular screening assays for RecA activities. *Analytical biochemistry*. **367**(2), 247–258.
Lehmann, L.C., Hewitt, G., Aibara, S., Leitner, A., Marklund, E., Maslen, S.L., Maturi, V., Chen, Y., van der Spoel, D., Skehel, J.M., Moustakas, A., Boulton, S.J., Deindl, S. (2017) Mechanistic Insights into Autoinhibition of the Oncogenic Chromatin Remodeler ALC1. *Molecular Cell*. **68**(5), 847-859.e7.
Lemmers, R.J.L.F., van der Stoep, N., Vliet, P.J. van der, Moore, S.A., San Leon Granado, D., Johnson, K., Topf, A., Straub, V., Evangelista, T., Mozaffar, T., Kimonis, V., Shaw, N.D., Selvatici, R., Ferlini, A., Voermans, N., van Engelen, B., Sacconi, S., Tawil, R., Lamers, M., van der Maarel, S.M. (2019) 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. *Journal of medical genetics*. **56**(10), 693–700.
Lenz, G., Wright, G.W., Emre, N.C.T., Kohlhammer, H., Dave, S.S., Davis, R.E., Carty, S., Lam, L.T., Shaffer, A.L., Xiao, W., Powell, J., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Rimsza, L.M., Fisher, R.I., Weisenburger, D.D., Chan, W.C., Staudt, L.M. (2008) Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proceedings of the National Academy of Sciences*. **105**(36), 13520 LP – 13525.
Leong, H.S., Chen, K., Hu, Y., Lee, S., Corbin, J., Pakusch, M., Murphy, J.M., Majewski, I.J., Smyth, G.K., Alexander, W.S., Hilton, D.J., Blewitt, M.E. (2013) Epigenetic regulator Smchd1 functions as a tumor suppressor. *Cancer Res*. **73**(5), 1591–1599.
Leroy, K., Haioun, C., Lepage, E., Le Métayer, N., Berger, F., Labouyrie, E., Meignin, V., Petit, B., Bastard, C., Salles, G., Gisselbrecht, C., Reyes, F., Gaulard, P. (2002) p53 gene mutations are

associated with poor survival in low and low-intermediate risk diffuse large B-cell lymphomas. *Annals of Oncology*. **13**(7), 1108–1115.
Letunic, I., Bork, P. (2018) 20 years of the SMART protein domain annotation resource. *Nucleic Acids Research*. **46**(D1), D493–D496.
Letunic, I., Doerks, T., Bork, P. (2015) SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Research*. **43**(D1), D257–D260.
De Leval, L., Harris, N.L. (2010) Diffuse large B cell lymphomas. *The Lymphoid Neoplasms 3ed*. **22**(5), 560–586.
Li, B., Carey, M., Workman, J.L. (2007) The Role of Chromatin during Transcription. *Cell*. **128**(4), 707–719.
Li, D.Q., Nair, S.S., Kumar, R. (2013) The MORC family: new epigenetic regulators of transcription and DNA damage response. *Epigenetics*. **8**(7), 685–693.
Li, D Q, Nair, S.S., Ohshiro, K., Kumar, A., Nair, V.S., Pakala, S.B., Reddy, S.D., Gajula, R.P., Eswaran, J., Aravind, L., Kumar, R. (2012) MORC2 signaling integrates phosphorylation-dependent, ATPase-coupled chromatin remodeling during the DNA damage response. *Cell Rep*. **2**(6), 1657–1669.
Li, J., Thompson, T.D., Miller, J.W., Pollack, L.A., Stewart, S.L. (2008) Cancer Incidence Among Children and Adolescents in the United States, 2001–2003. *Pediatrics*. **121**(6), e1470 LP-e1477.
Li, M., Huang, T., Li, M.-J., Zhang, C.-X., Yu, X.-C., Yin, Y.-Y., Liu, C., Wang, X., Feng, H.-W., Zhang, T., Liu, M.-F., Han, C.-S., Lu, G., Li, W., Ma, J.-L., Chen, Z.-J., Liu, H.-B., Liu, K. (2019) The histone modification reader ZCWPW1 is required for meiosis prophase I in male but not in female mice. *Science Advances*. **5**(8), eaax1101.
Li, Mengyuan, Sun, Q., Wang, X. (2017) Transcriptional landscape of human cancers. *Oncotarget*. **8**(21), 34534–34551.
Li, N., Lam, W.H., Zhai, Y., Cheng, J., Cheng, E., Zhao, Y., Gao, N., Tye, B.K. (2018) Structure of the origin recognition complex bound to DNA replication origin. *Nature*. **559**(7713), 217–222.
Li, Qiu, Li, X., Tang, H., Jiang, B., Dou, Y., Gorospe, M., Wang, W. (2017) NSUN2-Mediated m5C Methylation and METTL3/METTL14-Mediated m6A Methylation Cooperatively Enhance p21 Translation. *Journal of cellular biochemistry*. **118**(9), 2587–2598.
Li, S., Yen, L., Pastor, W.A., Johnston, J.B., Du, J., Shew, C.J., Liu, W., Ho, J., Stender, B., Clark, A.T., Burlingame, A.L., Daxinger, L., Patel, D.J., Jacobsen, S.E. (2016) Mouse MORC3 is a GHKL ATPase that localizes to H3K4me3 marked chromatin. *Proc Natl Acad Sci U S A*. **113**(35), E5108-16.
Liao, X.H., Zhang, Y., Dong, W.J., Shao, Z.M., Li, D.Q. (2017) Chromatin remodeling protein MORC2 promotes breast cancer invasion and metastasis through a PRD domain-mediated interaction with CTNND1. *Oncotarget*. **8**(58), 97941–97954.
Lichtinghagen, R., Musholt, P.B., Lein, M., Römer, A., Rudolph, B., Kristiansen, G., Hauptmann, S., Schnorr, D., Loening, S.A., Jung, K. (2002) Different mRNA and protein expression of matrix metalloproteinases 2 and 9 and tissue inhibitor of metalloproteinases 1 in benign and malignant prostate tissue. *European Urology*. **42**(4), 398–406.
Liggins, A P, Cooper, C.D., Lawrie, C.H., Brown, P.J., Collins, G.P., Hatton, C.S., Pulford, K., Banham, A.H. (2007) MORC4, a novel member of the MORC family, is highly expressed in a subset of diffuse large B-cell lymphomas. *Br J Haematol*. **138**(4), 479–486.
Linding, R., Russell, R.B., Neduva, V., Gibson, T.J. (2003) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Research*. **31**(13), 3701–3708.
Lindley, P.F. (1995) The use of synchrotron radiation in protein crystallography. *Radiation Physics and Chemistry*. **45**(3), 367–383.
Liu, B., K.H Yip, R., Zhou, Z. (2012) Chromatin Remodeling, DNA Damage Repair and Aging. *Current Genomics*. **13**(7), 533–547.

Liu, N., Balliano, A., Hayes, J.J. (2011) Mechanism(s) of SWI/SNF-Induced Nucleosome Mobilization. *ChemBioChem*. **12**(2), 196–204.
Liu, Y., Barta, S.K. (2019) Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *American Journal of Hematology*. **94**(5), 604–616.
Liu, Yansheng, Beyer, A., Aebersold, R. (2016) On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*. **165**(3), 535–550.
Liu, Yanli, Tempel, W., Zhang, Q., Liang, X., Loppnau, P., Qin, S., Min, J. (2016) Family-wide characterization of histone binding abilities of human CW domain-containing proteins. *Journal of Biological Chemistry*.
Loll, P.J., Xu, P., Schmidt, J.T., Melideo, S.L. (2014) Enhancing ubiquitin crystallization through surface-entropy reduction. *Acta Crystallographica Section F:Structural Biology Communications*. **70**, 1434–1442.
London, W.B., Castleberry, R.P., Matthay, K.K., Look, A.T., Seeger, R.C., Shimada, H., Thorner, P., Brodeur, G., Maris, J.M., Reynolds, C.P., Cohn, S.L. (2005) Evidence for an age cutoff greater than 365 days for neuroblastoma risk group stratification in the Children's Oncology Group. *Journal of Clinical Oncology*. **23**(27), 6459–6465.
Lorch, Y., Maier-Davis, B., Kornberg, R.D. (2010) Mechanism of chromatin remodeling. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(8), 3458–3462.
Lorkovic, Z.J. (2012) MORC proteins and epigenetic regulation. *Plant Signal Behav*. **7**(12), 1561–1565.
Lossos, I.S., Czerwinski, D.K., Wechser, M.A., Levy, R. (2003) Optimization of quantitative real-time RT-PCR parameters for the study of lymphoid malignancies. *Leukemia*. **17**(4), 789–795.
Love, I.M., Sekaric, P., Shi, D., Grossman, S.R., Androphy, E.J. (2012) The histone acetyltransferase PCAF regulates p21 transcription through stress-induced acetylation of histone H3 © 2012 Landes Bioscience . Do not distribute . © 2012 Landes Bioscience . Do not distribute . . **11**(13), 2458–2466.
Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A. (2020) CDD/SPARCLE: the conserved domain database in 2020. *Nucleic acids research*. **48**(D1), D265–D268.
Lubelsky, Y., Prinz, J.A., DeNapoli, L., Li, Y., Belsky, J.A., MacAlpine, D.M. (2014) DNA replication and transcription programs respond to the same chromatin cues. *Genome Research*. **24**(7), 1102–1114.
Luger, K., Dechassa, M.L., Tremethick, D.J. (2012) New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nature reviews. Molecular cell biology*. **13**(7), 436–447.
Lyu, C., Wang, L., Zhang, J. (2018) Deep learning for DNase i hypersensitive sites identification. *BMC Genomics*. **19**(Suppl 10).
MacPherson, S., Larochelle, M., Turcotte, B. (2006) A Fungal Family of Transcriptional Regulators: the Zinc Cluster Proteins. *Microbiology and Molecular Biology Reviews*. **70**(3), 583 LP – 604.
Maeshima, K., Hihara, S., Eltsov, M. (2010) Chromatin structure: Does the 30-nm fibre exist in vivo? *Current Opinion in Cell Biology*. **22**(3), 291–297.
Makino, Y., Kanemaki, M., Koga, A., Osano, K., Matsu-ura, T., Kurokawa, Y., Kishimoto, T., Tamura, T.-A. (2000) Chromosome Mapping and Expression of Human Tip49 Family Genes. *DNA Sequence*. **11**(1–2), 145–148.
Malumbres, M., Barbacid, M. (2005) Mammalian cyclin-dependent kinases. *Trends in Biochemical Sciences*. **30**(11), 630–641.
Manning, B.D., Cantley, L.C. (2003) Rheb fills a GAP between TSC and TOR. *Trends Biochem Sci*. **28**(11), 573–576.
El Marabti, E., Younis, I. (2018) The cancer spliceome: Reprograming of alternative splicing in cancer.

Frontiers in Molecular Biosciences. **5**(SEP), 1–11.
Mariño-Ramírez, L., Kann, M.G., Shoemaker, B.A., Landsman, D. (2005) Histone structure and nucleosome stability. *Expert Review of Proteomics*. **2**(5), 719–729.
Maris, J.M. (2010) Recent advances in neuroblastoma. *New England Journal of Medicine*. **362**(23), 2202.
Marley, A.R., Nan, H. (2016) Epidemiology of colorectal cancer. *International journal of molecular epidemiology and genetics*. **7**(3), 105–114.
Marmorstein, R., Berger, S.L. (2001) Structure and function of bromodomains in chromatin-regulating complexes. *Gene*. **272**(1–2), 1–9.
Marsh, D.J., Shah, J.S., Cole, A.J. (2014) Histones and their modifications in ovarian cancer - drivers of disease and therapeutic targets. *Frontiers in Oncology*. **4 JUN**(June), 1–12.
Martelli, M., Ferreri, A.J.M., Agostinelli, C., Di Rocco, A., Pfreundschuh, M., Pileri, S.A. (2013) Diffuse large B-cell lymphoma. *Critical Reviews in Oncology / Hematology*. **87**, 146–171.
Martin, N., Li, C.K., Wedderburn, L.R. (2012) Juvenile dermatomyositis: New insights and new treatment strategies. *Therapeutic Advances in Musculoskeletal Disease*. **4**(1), 41–50.
Martínez-Reyes, I., Chandel, N.S. (2018) Acetyl-coa-directed gene transcription in cancer cells. *Genes and Development*. **32**(7–8), 463–465.
Mason, J.M., Arndt, K.M. (2004) Coiled coil domains: stability, specificity, and biological implications. *Chembiochem*. **5**(2), 170–176.
Mauri, G., Sartore-Bianchi, A., Russo, A.-G., Marsoni, S., Bardelli, A., Siena, S. (2019) Early-onset colorectal cancer in young individuals. *Molecular oncology*. **13**(2), 109–131.
Mayeux, R. (2004) Biomarkers: potential uses and limitations. *NeuroRx : the journal of the American Society for Experimental NeuroTherapeutics*. **1**(2), 182–188.
McInerney, P., Adams, P., Hadi, M.Z. (2014) Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase A. Desideri, ed. *Molecular Biology International*. **2014**, 287430.
Mehrgou, A., Akouchekian, M. (2016) The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. *Medical journal of the Islamic Republic of Iran*. **30**, 369.
Menafra, R., Stunnenberg, H.G. (2014) MBD2 and MBD3: elusive functions and mechanisms. *Frontiers in genetics*. **5**, 428.
Mendiratta, S., Bhatia, S., Jain, S., Kaur, T., Brahmachari, V. (2016) Interaction of the chromatin remodeling protein hINO80 with DNA. *PLoS ONE*. **11**(7), 1–21.
Messner, S., Altmeyer, M., Zhao, H., Pozivil, A., Roschitzki, B., Gehrig, P., Rutishauser, D., Huang, D., Caflisch, A., Hottiger, M.O. (2010) PARP1 ADP-ribosylates lysine residues of the core histone tails. *Nucleic acids research*. **38**(19), 6350–6362.
Messner, S., Hottiger, M.O. (2011) Histone ADP-ribosylation in DNA repair, replication and transcription. *Trends in Cell Biology*. **21**(9), 534–542.
Mimura, Y., Takahashi, K., Kawata, K., Akazawa, T., Inoue, N. (2010) Two-step colocalization of MORC3 with PML nuclear bodies. *Journal of Cell Science*. **123**(12), 2014–2024.
Miotto, B., Ji, Z., Struhl, K. (2016) Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proceedings of the National Academy of Sciences of the United States of America*. **113**(33), E4810–E4819.
Mirabella, A.C., Foster, B.M., Bartke, T. (2016) Chromatin deregulation in disease. *Chromosoma*. **125**(1), 75–93.
Mohd-Sarip, A., Teeuwssen, M., Bot, A.G., De Herdt, M.J., Willems, S.M., Baatenburg de Jong, R.J., Looijenga, L.H.J., Zatreanu, D., Bezstarosti, K., van Riet, J., Oole, E., van Ijcken, W.F.J., van de Werken, H.J.G., Demmers, J.A., Fodde, R., Verrijzer, C.P. (2017) DOC1-Dependent Recruitment of NURD Reveals Antagonism with SWI/SNF during Epithelial-Mesenchymal Transition in Oral Cancer Cells. *Cell Reports*. **20**(1), 61–75.
Moissiard, G., Bischof, S., Husmann, D., Pastor, W.A., Hale, C.J., Yen, L., Stroud, H., Papikian, A., Vashisht, A.A., Wohlschlegel, J.A., Jacobsen, S.E. (2014) Transcriptional gene silencing by

Arabidopsis microrchidia homologues involves the formation of heteromers. *Proceedings of the National Academy of Sciences of the United States of America*. **111**(20), 7474–7479.
Moissiard, G., Cokus, S.J., Cary, J., Feng, S., Billi, A.C., Stroud, H., Husmann, D., Zhan, Y., Lajoie, B.R., McCord, R.P., Hale, C.J., Feng, W., Michaels, S.D., Frand, A.R., Pellegrini, M., Dekker, J., Kim, J.K., Jacobsen, S. (2012) MORC Family ATPases Required for Heterochromatin Condensation and Gene Silencing. *Science (New York, N.Y.)*. **336**(6087), 1448–1451.
Mooij, W.T.M., Mitsiki, E., Perrakis, A. (2009) ProteinCCD: enabling the design of protein truncation constructs for expression and crystallization experiments. *Nucleic Acids Research*. **37**(suppl_2), W402–W405.
Moore, L.D., Le, T., Fan, G. (2013) DNA methylation and its basic function. *Neuropsychopharmacology*. **38**(1), 23–38.
Morrison, A.J. (2020) Chromatin-remodeling links metabolic signaling to gene expression. *Molecular Metabolism*. **38**(March), 100973.
Mortlock, D., Keller, E.B., Ziegra, C.J., Suter, M.M. (1993) High efficiency transfection of monkey kidney COS-1 cells. *Journal of tissue culture methods: Tissue Culture Association manual of cell, tissue, and organ culture procedures*. **15**(4), 176–180.
Mosse, Y.P., Laudenslager, M., Khazi, D., Carlisle, A.J., Winter, C.L., Rappaport, E., Maris, J.M. (2004) Germline PHOX2B mutation in hereditary neuroblastoma. *American journal of human genetics*. **75**(4), 727–730.
Mossé, Y.P., Laudenslager, M., Longo, L., Cole, K.A., Wood, A., Attiyeh, E.F., Laquaglia, M.J., Sennett, R., Lynch, J.E., Perri, P., Laureys, G., Speleman, F., Kim, C., Hou, C., Hakonarson, H., Torkamani, A., Schork, N.J., Brodeur, G.M., Tonini, G.P., Rappaport, E., Devoto, M., Maris, J.M. (2008) Identification of ALK as a major familial neuroblastoma predisposition gene. *Nature*. **455**(7215), 930–935.
Moynagh, P.N. (2005) The NF-κB pathway. *Journal of Cell Science*. **118**(20), 4589–4592.
Mozaheb, Z., Aledavood, A., Farzad, F. (2012) Diet and non-Hodgkin's lymphoma risk. *The Pan African medical journal*. **12**, 53.
Mugnaini, E.N., Ghosh, N. (2016) Lymphoma. *Primary Care: Clinics in Office Practice*. **43**(4), 661–675.
Mullassery, D., Losty, P.D. (2016) Neuroblastoma. *Paediatrics and Child Health (United Kingdom)*. **26**(2), 68–72.
Müller, I. (2017) Guidelines for the successful generation of protein-ligand complex crystals. *Acta Crystallographica Section D: Structural Biology*. **73**(2), 79–92.
Mundorf, A., Schmitz, J., Güntürkün, O., Freund, N., Ocklenburg, S. (2018) Methylation of MORC1: A possible biomarker for depression? *Journal of Psychiatric Research*. **103**(May), 208–211.
Murawska, M., Brehm, A. (2011) CHD chromatin remodelers and the transcription cycle. *Transcription*. **2**(6), 244–253.
Musselman, C.A., Lalonde, M.-E., Côté, J., Kutateladze, T.G. (2012) Perceiving the epigenetic landscape through histone readers. *Nature structural & molecular biology*. **19**(12), 1218–1227.
Nachtergaele, S., He, C. (2017) The emerging biology of RNA post-transcriptional modifications. *RNA Biology*. **14**(2), 156–163.
Nagy, Z., Tora, L. (2007) Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation. *Oncogene*. **26**(37), 5341–5357.
Nakano, T., Miyamoto-Matsubara, M., Shoulkamy, M.I., Salem, A.M.H., Pack, S.P., Ishimi, Y., Ide, H. (2013) Translocation and stability of replicative DNA helicases upon encountering DNA-protein cross-links. *The Journal of biological chemistry*. **288**(7), 4649–4658.
Nathan, D., Sterner, D.E., Berger, S.L. (2003) Histone modifications: Now summoning sumoylation. *Proceedings of the National Academy of Sciences of the United States of America*. **100**(23), 13118–13120.

Neagu, A., Neagu, M., Dér, A. (2001) Fluctuations and the Hofmeister effect. *Biophysical Journal*. **81**(3), 1285–1294.
Neigeborn, L., Carlson, M. (1984) Genes affecting the regulation of SUC2 gene expression by glucose repression in Saccharomyces cerevisiae. *Genetics*. **108**(4), 845–858.
Newman, E.A., Abdessalam, S., Aldrink, J.H., Austin, M., Heaton, T.E., Bruny, J., Ehrlich, P., Dasgupta, R., Baertschiger, R.M., Lautz, T.B., Rhee, D.S., Langham, M.R., Malek, M.M., Meyers, R.L., Nathan, J.D., Weil, B.R., Polites, S., Madonna, M.B. (2019) Update on neuroblastoma. *Journal of Pediatric Surgery*. **54**(3), 383–389.
Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Jelloul, M. Ben, Sussman, J.L., Stuart, D.I., Perrakis, A. (2005) Towards rationalization of crystallization screening for small- To medium-sized academic laboratories: The PACT/JCSG+ strategy. *Acta Crystallographica Section D: Biological Crystallography*. **61**(10), 1426–1431.
Nie, X., Wei, J., Hao, Y., Tao, J., Li, Y., Liu, M., Xu, B., Li, B. (2019) Consistent biomarkers and related pathogenesis underlying asthma revealed by systems biology approach. *International Journal of Molecular Sciences*. **20**(16).
Nikfarjam, L., Farzaneh, P. (2012) Prevention and detection of mycoplasma contamination in cell culture. *Cell Journal*. **13**(4), 203–212.
Ninkovic, S., Lambert, J. (2017) Non-Hodgkin lymphoma. *Medicine (United Kingdom)*. **45**(5), 297–304.
Nishihara, H., Maruyama, S., Okada, N. (2009) Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proceedings of the National Academy of Sciences*. **106**(13), 5235 LP – 5240.
North, J.A., Šimon, M., Ferdinand, M.B., Shoffner, M.A., Picking, J.W., Howard, C.J., Mooney, A.M., Van Noort, J., Poirier, M.G., Ottesen, J.J. (2014) Histone H3 phosphorylation near the nucleosome dyad alters chromatin structure. *Nucleic Acids Research*. **42**(8), 4922–4933.
Nowakowski, A.B., Wobig, W.J., Petering, D.H. (2014) Native SDS-PAGE: high resolution electrophoretic separation of proteins with retention of native properties including bound metal ions. *Metallomics : integrated biometal science*. **6**(5), 1068–1078.
Ó'Fágáin, C., Cummins, P.M., O'Connor, B.F. (2017) Gel-Filtration Chromatography. *Methods in molecular biology (Clifton, N.J.)*. **1485**, 15–25.
Ogryzko, V. V., Kotani, T., Zhang, X., Schiltz, R.L., Howard, T., Yang, X.J., Howard, B.H., Qin, J., Nakatani, Y. (1998) Histone-like TAFS within the PCAF histone acetylase complex. *Cell*. **94**(1), 35–44.
Olave, I.A., Reck-Peterson, S.L., Crabtree, G.R. (2002) Nuclear Actin and Actin-Related Proteins in Chromatin Remodeling. *Annual Review of Biochemistry*. **71**(1), 755–781.
Ørntoft, T.F., Thykjaer, T., Waldman, F.M., Wolf, H., Celis, J.E. (2002) Genome-wide study of gene copy numbers, transcripts, and protein levels in pairs of non-invasive and invasive human transitional cell carcinomas. *Molecular & cellular proteomics : MCP*. **1**(1), 37–45.
Ossovskaya, V., Koo, I.C., Kaldjian, E.P., Alvares, C., Sherman, B.M. (2010) Upregulation of poly (ADP-Ribose) polymerase-1 (PARP1) in triple-negative breast cancer and other primary human tumor types. *Genes and Cancer*. **1**(8), 812–821.
Padrón, R., Huxley, H.E. (1984) The effect of the ATP analogue AMPPNP on the structure of crossbridges in vertebrate skeletal muscles: X-ray diffraction and mechanical studies. *Journal of Muscle Research and Cell Motility*. **5**(6), 613–655.
Page, R., Grzechnik, S.K., Canaves, J.M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R.C., Lesley, S.A. (2003) Shotgun crystallization strategy for structural genomics: An optimized two-tiered crystallization screen against the Thermotoga maritima proteome. *Acta Crystallographica - Section D Biological Crystallography*. **59**(6), 1028–1037.
Pampal, A. (2010) CHARGE: An association or a syndrome? *International Journal of Pediatric Otorhinolaryngology*. **74**(7), 719–722.

Pan, Z., Ding, Q., Guo, Q., Guo, Y., Wu, Lianlian, Wu, Lu, Tang, M., Yu, H., Zhou, F. (2018) MORC2, a novel oncogene, is upregulated in liver cancer and contributes to proliferation, metastasis and chemoresistance. *International Journal of Oncology*. **53**(1), 59–72.

Papaioannou, G., McHugh, K. (2005) Neuroblastoma in childhood: Review and radiological findings. *Cancer Imaging*. **5**(1), 116–127.

Park, S.J., Yoon, B.H., Kim, S.K., Kim, S.Y. (2019) GENT2: An updated gene expression database for normal and tumor tissues. *BMC Medical Genomics*. **12**(Suppl 5), 1–8.

Pastor, W.A., Stroud, H., Nee, K., Liu, W., Pezic, D., Manakov, S., Lee, S.A., Moissiard, G., Zamudio, N., Bourc'his, D., Aravin, A.A., Clark, A.T., Jacobsen, S.E. (2014) MORC1 represses transposable elements in the mouse male germline. *Nat Commun*. **5**, 5795.

Paul, P., Malakar, A.K., Chakraborty, S. (2019) The significance of gene mutations across eight major cancer types. *Mutation Research - Reviews in Mutation Research*. **781**(April), 88–99.

Pazzaglia, S., Pioli, C. (2019) Multifaceted Role of PARP-1 in DNA Repair and Inflammation: Pathological and Therapeutic Implications in Cancer and Non-Cancer Diseases. *Cells*. **9**(1), 41.

Pedersen, L.C., Inoue, K., Kim, S., Perera, L., Shaw, N.D. (2019) A ubiquitin-like domain is required for stabilizing the N-terminal ATPase module of human SMCHD1. *Communications Biology*. **2**(1), 1–10.

Peng, C., Zeng, W., Su, J., Kuang, Y., He, Y., Zhao, S., Zhang, J., Ma, W., Bode, A.M., Dong, Z., Chen, X. (2016) Cyclin-dependent kinase 2 (CDK2) is a key mediator for EGF-induced cell transformation mediated through the ELK4/c-Fos signaling pathway. *Oncogene*. **35**(9), 1170–1179.

Perera, J., Huang, H. (2015) The development and function of thymic B cells. *Cellular and molecular life sciences : CMLS*. **72**(14), 2657–2663.

Perry, J., Zhao, Y. (2003) The CW domain, a structural module shared amongst vertebrates, vertebrate-infecting parasites and higher plants. *Trends in Biochemical Sciences*. **28**(11), 576–580.

Petesch, S.J., Lis, J.T. (2012) Overcoming the nucleosome barrier during transcript elongation. *Trends in genetics : TIG*. **28**(6), 285–294.

Petryszak, R., Fonseca, N.A., Füllgrabe, A., Huerta, L., Keays, M., Tang, Y.A., Brazma, A. (2017) The RNASeq-er API-A gateway to systematically updated analysis of public RNA-seq data. *Bioinformatics*. **33**(14), 2218–2220.

Peyressatre, M., Prével, C., Pellerano, M., Morris, M.C. (2015) Targeting cyclin-dependent kinases in human cancers: From small molecules to peptide inhibitors. *Cancers*. **7**(1), 179–237.

Pfeifer, M., Grau, M., Lenze, D., Wenzel, S.-S., Wolf, A., Wollert-Wulf, B., Dietze, K., Nogai, H., Storek, B., Madle, H., Dörken, B., Janz, M., Dirnhofer, S., Lenz, P., Hummel, M., Tzankov, A., Lenz, G. (2013) PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*. **110**(30), 12420–12425.

Pieper, K., Grimbacher, B., Eibel, H. (2013) B-cell biology and development. *Journal of Allergy Clinical Immunology*. **131**(4), 959–971.

Prager, G.W., Zielinski, C.C. (2013) Angiogenesis in cancer. *Biochemical Basis and Therapeutic Implications of Angiogenesis*. **2**(3), 335–356.

Prendergast, A., Raible, D.W. (2014) Chapter 13 - Neural Crest Cells and Peripheral Nervous System Development. In P. A. B. T.-N. C. C. Trainor, ed. Boston: Academic Press, pp. 255–286.

Price, B.D. (2013) Chromatin remodeling at DNA double-strand breaks. *Cell*. **152**(6), 1344–1354.

Prilusky, J., Felder, C.E., Zeev-Ben-Mordehai, T., Rydberg, E.H., Man, O., Beckmann, J.S., Silman, I., Sussman, J.L. (2005) FoldIndex©: A simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. **21**(16), 3435–3438.

Quan, T.K., Hartzog, G.A. (2010) Histone H3K4 and K36 methylation, Chd1 and Rpd3S oppose the functions of Saccharomyces cerevisiae Spt4-Spt5 in transcription. *Genetics*. **184**(2), 321–334.

Quina, A.S., Buschbeck, M., Di Croce, L. (2006) Chromatin structure and epigenetics. *Biochemical Pharmacology*. **72**(11), 1563–1569.
Rack, J.G.M., Ariza, A., Drown, B.S., Henfrey, C., Bartlett, E., Shirai, T., Hergenrother, P.J., Ahel, I. (2018) (ADP-ribosyl)hydrolases: Structural Basis for Differential Substrate Recognition and Inhibition. *Cell Chemical Biology*. **25**(12), 1533-1546.e12.
Radu, M., Semenova, G., Kosoff, R., Chernoff, J. (2014) Pak Signaling in the Development and Progression of Cancer. *Nature reviews. Cancer*. **14**(1), 13–25.
Raghunathan, K., Harris, P.T., Arvidson, D.N. (2010) Trial by fire: Are the crystals macromolecules? *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*. **66**(5), 615–620.
Rai, V., Gaur, M., Shukla, Sudhanshu, Shukla, Suneet, Ambudkar, S. V, Komath, S.S., Prasad, R. (2006) Conserved Asp327 of walker B motif in the N-terminal nucleotide binding domain (NBD-1) of Cdr1p of Candida albicans has acquired a new role in ATP hydrolysis. *Biochemistry*. **45**(49), 14726–14739.
Ramakrishnan, C., Dani, V.S., Ramasarma, T. (2002) A conformational analysis of Walker motif A [GXXXXGKT (S)] in nucleotide-binding and other proteins. *Protein Engineering, Design and Selection*. **15**(10), 783–798.
Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B.-H., Das, R., Grishin, N. V, Baker, D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins*. **77 Suppl 9**(0 9), 89–99.
Ramezani-Rad, P., Rickert, R.C. (2017) Murine models of germinal center derived-lymphomas. *Curr Opin Immunol*. **45**, 31–36.
Rao, X., Huang, X., Zhou, Z., Lin, X. (2013) An improvement of the 2ˆ(-delta delta CT) method for quantitative real-time polymerase chain reaction data analysis. *Biostatistics, bioinformatics and biomathematics*. **3**(3), 71–85.
Rappsilber, J. (2011) The beginning of a beautiful friendship: cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. *Journal of structural biology*. **173**(3), 530–540.
Ray Chaudhuri, A., Nussenzweig, A. (2017) The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nature reviews. Molecular cell biology*. **18**(10), 610–621.
Razin, S. V., Gavrilov, A.A. (2014) Chromatin without the 30-nm fiber constrained disorder instead of hierarchical folding. *Epigenetics*. **9**(5), 653–657.
Rial, D. V, Ceccarelli, E.A. (2002) Removal of DnaK contamination during fusion protein purifications. *Protein Expression and Purification*. **25**(3), 503–507.
Rice, J.C., Allis, C.D. (2001) Histone methylation versus histone acetylation: New insights into epigenetic regulation. *Current Opinion in Cell Biology*. **13**(3), 263–273.
Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., Längst, G. (2007) DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences of the United States of America*. **104**(40), 15635–15640.
Rivlin, N., Brosh, R., Oren, M., Rotter, V. (2011) Mutations in the p53 tumor suppressor gene: Important milestones at the various steps of tumorigenesis. *Genes and Cancer*. **2**(4), 466–474.
Roberts, C.W.M., Orkin, S.H. (2004) The SWI/SNF complex — chromatin and cancer. *Nature Reviews Cancer*. **4**(2), 133–142.
Robinson, J.D. (1980) Binding to the high-affinity substrate site of the (Na+ + K+)-dependent ATPase. *Journal of Bioenergetics and Biomembranes*. **12**(3–4), 165–174.
Roca, J. (2011) Transcriptional inhibition by DNA torsional stress. *Transcription*. **2**(2), 82–85.
Rohde, J.R., Cardenas, M.E. (2003) The Tor Pathway Regulates Gene Expression by Linking Nutrient Sensing to Histone Acetylation. *Molecular and Cellular Biology*. **23**(2), 629–635.

Romero Romero, M.L., Yang, F., Lin, Y.R., Toth-Petroczy, A., Berezovsky, I.N., Goncearenco, A., Yang, W., Wellner, A., Kumar-Deshmukh, F., Sharon, M., Baker, D., Varani, G., Tawfik, D.S. (2018) Simple yet functional phosphate-loop proteins. *Proceedings of the National Academy of Sciences of the United States of America*. **115**(51), E11943–E11950.

Rosano, G.L., Ceccarelli, E.A. (2014) Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology*. **5**(APR), 1–17.

Rosendorff, A., Sakakibara, S., Lu, S., Kieff, E., Xuan, Y., DiBacco, A., Shi, Yujiang, Shi, Yang, Gill, G. (2006) NXP-2 association with SUMO-2 depends on lysines required for transcriptional repression. *Proceedings of the National Academy of Sciences*. **103**(14), 5308–5313.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., López-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T., Staudt, L.M. (2002) The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-Cell Lymphoma. *New England Journal of Medicine*. **346**(25), 1937–1947.

Rosenwald, A., Wright, G., Leroy, K., Yu, X., Gaulard, P., Gascoyne, R.D., Chan, W.C., Zhao, T., Haioun, C., Greiner, T.C., Weisenburger, D.D., Lynch, J.C., Vose, J., Armitage, J.O., Smeland, E.B., Kvaloy, S., Holte, H., Delabie, J., Campo, E., Montserrat, E., Lopez-Guillermo, A., Ott, G., Muller-Hermelink, H.K., Connors, J.M., Braziel, R., Grogan, T.M., Fisher, R.I., Miller, T.P., LeBlanc, M., Chiorazzi, M., Zhao, H., Yang, L., Powell, J., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Staudt, L.M. (2003) Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *The Journal of experimental medicine*. **198**(6), 851–862.

Rossetto, D., Avvakumov, N., Côté, J. (2012) Histone phosphorylation: A chromatin modification involved in diverse nuclear events. *Epigenetics*. **7**(10), 1098–1108.

Rozeboom, B., Dey, N., De, P. (2019) ER+ metastatic breast cancer: past, present, and a prescription for an apoptosis-targeted future. *American journal of cancer research*. **9**(12), 2821–2831.

Rule, C.S., Patrick, M., Sandkvist, M. (2016) Measuring in vitro ATPase activity for enzymatic characterization. *Journal of Visualized Experiments*. **2016**(114), 3–7.

Russell, R., Jordan, R., McMacken, R. (1998) Kinetic Characterization of the ATPase Cycle of the DnaK Molecular Chaperone. *Biochemistry*. **37**(2), 596–607.

Ryan, D.P., Owen-Hughes, T. (2011) Snf2-family proteins: chromatin remodellers for any occasion. *Curr Opin Chem Biol*. **15**(5), 649–656.

Saini, N., Roberts, S.A., Klimczak, L.J., Chan, K., Grimm, S.A., Dai, S., Fargo, D.C., Boyer, J.C., Kaufmann, W.K., Taylor, J.A., Lee, E., Cortes-Ciriano, I., Park, P.J., Schurman, S.H., Malc, E.P., Mieczkowski, P.A., Gordenin, D.A. (2016) The Impact of Environmental and Endogenous Damage on Somatic Mutation Load in Human Skin Fibroblasts. *PLoS Genetics*. **12**(10), 1–25.

Sakabe, K., Wang, Z., Hart, G.W. (2010) β-N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America*. **107**(46), 19915–19920.

Sakakibara, Y., Nagao, K., Blewitt, M., Sasaki, H., Obuse, C., Sado, T. (2018) Role of smcHD1 in establishment of epigenetic states required for the maintenance of the x-inactivated state in mice. *Development (Cambridge)*. **145**(18).

Sanchez-Solana, B., Li, D.Q., Kumar, R. (2014) Cytosolic functions of MORC2 in lipogenesis and adipogenesis. *Biochim Biophys Acta*. **1843**(2), 316–326.

Sanchez, R., Zhou, M.M. (2011) The PHD finger: A versatile epigenome reader. *Trends in Biochemical Sciences*. **36**(7), 364–372.

Santos-rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Cote, J., Kouzarides, T. (2012) Histone H3 tail clipping regulates gene expression. . **16**(1), 17–22.

Saravanan, M., Wuerges, J., Bose, D., McCormack, E.A., Cook, N.J., Zhang, X., Wigley, D.B. (2012) Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(51), 20883–20888.

Savage, K.J. (2006) Primary Mediastinal Large B-Cell Lymphoma. *The Oncologist* . **11**(5), 488–495.

Savage, K.J., Monti, S., Kutok, J.L., Cattoretti, G., Neuberg, D., de Leval, L., Kurtin, P., Cin, P.D., Ladd, C., Feuerhake, F., Aguiar, R.C.T., Li, S., Salles, G., Berger, F., Jing, W., Pinkus, G.S., Habermann, T., Dalla-Favera, R., Harris, N.L., Aster, J.C., Golub, T.R., Shipp, M.A. (2003) The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood*. **102**(12), 3871–3879.

Savitsky, P., Bray, J., Cooper, C.D., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., Gileadi, O. (2010) High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol*. **172**(1), 3–13.

Sawicka, A., Seiser, C. (2014) Sensing core histone phosphorylation - A matter of perfect timing. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. **1839**(8), 711–718.

Sawyer, S.A., Parsch, J., Zhang, Z., Hartl, D.L. (2007) Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*. **104**(16), 6504–6510.

Scacchetti, A., Brueckner, L., Jain, D., Schauer, T., Zhang, X., Schnorrer, F., van Steensel, B., Straub, T., Becker, P.B. (2018) CHRAC/ACF contribute to the repressive ground state of chromatin. *Life Science Alliance*. **1**(1), e201800024.

Schmidt, M., Brandwein, C., Luoni, A., Sandrini, P., Calzoni, T., Deuschle, M., Cirulli, F., Riva, M.A., Gass, P. (2016) Morc1 knockout evokes a depression-like phenotype in mice. *Behavioural Brain Research*. **296**, 7–14.

Scott, D.W., Gascoyne, R.D. (2014) The tumour microenvironment in B cell lymphomas. *Nat Rev Cancer*. **14**(8), 517–534.

Sebesta, M., Cooper, C.D.O., Ariza, A., Carnie, C.J., Ahel, D. (2017) Structural insights into the function of ZRANB3 in replication stress response. *Nature Communications*. **8**(May), 1–16.

Seifert, M., Küppers, R. (2016) Human memory B cells. *Leukemia*. **30**(12), 2283–2292.

van Seijen, M., Lips, E.H., Thompson, A.M., Nik-Zainal, S., Futreal, A., Hwang, E.S., Verschuur, E., Lane, J., Jonkers, J., Rea, D.W., Wesseling, J. (2019) Ductal carcinoma in situ: to treat or not to treat, that is the question. *British Journal of Cancer*. **121**(4), 285–292.

Sena, J.A., Wang, L., Hu, C.-J. (2013) BRG1 and BRM Chromatin-Remodeling Complexes Regulate the Hypoxia Response by Acting as Coactivators for a Subset of Hypoxia-Inducible Transcription Factor Target Genes. *Molecular and Cellular Biology*. **33**(19), 3849–3863.

Seto, E., Yoshida, M. (2014) Erasers of histone acetylation: The histone deacetylase enzymes. *Cold Spring Harbor Perspectives in Biology*. **6**(4), 1–26.

Sezonov, G., Joseleau-Petit, D., D'Ari, R. (2007) Escherichia coli physiology in Luria-Bertani broth. *Journal of Bacteriology*. **189**(23), 8746–8749.

Shah, S.P., Morin, R.D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R.A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G.A., Teschendorff, A.E., Tse, K., Turashvili, G., Varhol, R., Warren, R.L., Watson, P., Zhao, Y., Caldas, C., Huntsman, D., Hirst, M., Marra, M.A., Aparicio, S. (2009) Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. **461**(7265), 809–813.

Shankland, K.R., Armitage, J.O., Hancock, B.W. (2012) Non-Hodgkin lymphoma. *The Lancet*. **380**(9844), 848–857.

Sharma, G.N., Dave, R., Sanadya, J., Sharma, P., Sharma, K.K. (2010) Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology & research*. **1**(2), 109–126.

Sharma, Rohini, Cunningham, D., Smith, P., Robertson, G., Dent, O., Clarke, S.J. (2009) Inflammatory (B) symptoms are independent predictors of myelosuppression from chemotherapy in Non-Hodgkin Lymphoma (NHL) patients – analysis of data from a British National Lymphoma Investigation phase III trial comparing CHOP to PMitCEBO. *BMC Cancer*. **9**(1), 153.

Sharma, Soniya, Salehi, F., Scheithauer, B.W., Rotondo, F., Syro, L. V., Kovacs, K. (2009) Role of MGMT in tumor development, progression, diagnosis, treatment and prognosis. *Anticancer Research*. **29**(10), 3759–3768.

Shaw, N.D., Brand, H., Kupchinsky, Z.A., Bengani, H., Plummer, L., Jones, T.I., Erdin, S., Williamson, K.A., Rainger, J., Stortchevoi, A., Samocha, K., Currall, B.B., Dunican, D.S., Collins, R.L., Willer, J.R., Lek, A., Lek, M., Nassan, M., Pereira, S., Kammin, T., Lucente, D., Silva, A., Seabra, C.M., Chiang, C., An, Y., Ansari, M., Rainger, J.K., Joss, S., Smith, J.C., Lippincott, M.F., Singh, S.S., Patel, N., Jing, J.W., Law, J.R., Ferraro, N., Verloes, A., Rauch, A., Steindl, K., Zweier, M., Scheer, I., Sato, D., Okamoto, N., Jacobsen, C., Tryggestad, J., Chernausek, S., Schimmenti, L.A., Brasseur, B., Cesaretti, C., García-Ortiz, J.E., Buitrago, T.P., Silva, O.P., Hoffman, J.D., Mühlbauer, W., Ruprecht, K.W., Loeys, B.L., Shino, M., Kaindl, A.M., Cho, C.-H., Morton, C.C., Meehan, R.R., van Heyningen, V., Liao, E.C., Balasubramanian, R., Hall, J.E., Seminara, S.B., Macarthur, D., Moore, S.A., Yoshiura, K.-I., Gusella, J.F., Marsh, J.A., Graham Jr, J.M., Lin, A.E., Katsanis, N., Jones, P.L., Crowley Jr, W.F., Davis, E.E., FitzPatrick, D.R., Talkowski, M.E. (2017) SMCHD1 mutations associated with a rare muscular dystrophy can also cause isolated arhinia and Bosma arhinia microphthalmia syndrome. *Nature genetics*. **49**(2), 238–248.

Shen, X., Liu, Y., Hsu, Y.-J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.-C., Orkin, S.H. (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Molecular cell*. **32**(4), 491–502.

Simon, K. (2016) Colorectal cancer development and advances in screening. *Clinical interventions in aging*. **11**, 967–976.

Skarzynski, T. (2013) Collecting data in the home laboratory: Evolution of X-ray sources, detectors and working practices. *Acta Crystallographica Section D: Biological Crystallography*. **69**(7), 1283–1288.

Sloan, E., Orr, A., Everett, R.D. (2016) MORC3, a Component of PML Nuclear Bodies, Has a Role in Restricting Herpes Simplex Virus 1 and Human Cytomegalovirus. *J Virol*. **90**(19), 8621–8633.

Sloan, E., Tatham, M.H., Groslambert, M., Glass, M., Orr, A., Hay, R.T., Everett, R.D. (2015) Analysis of the SUMO2 Proteome during HSV-1 Infection. *PLoS Pathogens*. **11**(7), 1–34.

Smith, M.C., Boutell, C., Davido, D.J. (2011) HSV-1 ICP0: paving the way for viral replication. *Future virology*. **6**(4), 421–429.

Smith, R., Sellou, H., Chapuis, C., Huet, S., Timinszky, G. (2018) CHD3 and CHD4 recruitment and chromatin remodeling activity at DNA breaks is promoted by early poly(ADP-ribose)-dependent chromatin relaxation. *Nucleic Acids Research*. **46**(12), 6087–6098.

Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., Kobe, B. (2003) Crystal structures of fusion proteins with large-affinity tags. *Protein science : a publication of the Protein Society*. **12**(7), 1313–1322.

Smyth, M.S., Martin, J.H.J. (2000) x Ray crystallography. *Journal of Clinical Pathology - Molecular Pathology*. **53**(1), 8–14.

Snell, E.H., Nagel, R.M., Wojtaszcyk, A., O'Neill, H., Wolfley, J.L., Luft, J.R. (2008) The application and use of chemical space mapping to interpret crystallization screening results. *Acta Crystallographica Section D: Biological Crystallography*. **64**(12), 1240–1249.

Soderman, J., Noren, E., Christiansson, M., Bragde, H., Thiebaut, R., Hugot, J.P., Tysk, C., O'Morain, C.A., Gassull, M., Finkel, Y., Colombel, J.F., Lemann, M., Almer, S. (2013) Analysis of single nucleotide polymorphisms in the region of CLDN2-MORC4 in relation to inflammatory bowel disease. *World J Gastroenterol*. **19**(30), 4935–4943.

Sokpor, G., Xie, Y., Rosenbusch, J., Tuoc, T. (2017) Chromatin remodeling BAF (SWI/SNF) complexes in neural development and disorders. *Frontiers in Molecular Neuroscience*. **10**(August), 1–22.

Some, D., Amartely, H., Tsadok, A., Lebendiker, M. (2019) Characterization of proteins by size-exclusion chromatography coupled to multi-angle light scattering (Sec-mals). *Journal of Visualized Experiments*. **2019**(148), 1–9.

Song, Y., Dimaio, F., Wang, R.Y.R., Kim, D., Miles, C., Brunette, T., Thompson, J., Baker, D. (2013) High-resolution comparative modeling with RosettaCM. *Structure*. **21**(10), 1735–1742.

Song, Y., Peisach, D., Pioszak, A.A., Xu, Z., Ninfa, A.J. (2004) Crystal Structure of the C-terminal Domain of the Two-Component System Transmitter Protein Nitrogen Regulator II (NRII; NtrB), Regulator of Nitrogen Assimilation in Escherichia coli,. *Biochemistry*. **43**(21), 6670–6678.

Śpibida, M., Krawczyk, B., Olszewski, M., Kur, J. (2017) Modified DNA polymerases for PCR troubleshooting. *Journal of Applied Genetics*. **58**(1), 133–142.

Starkman, B.G., Sakharkar, A.J., Pandey, S.C. (2012) Epigenetics-beyond the genome in alcoholism. *Alcohol research : current reviews*. **34**(3), 293–305.

Statland, J., Tawil, R. (2014) Facioscapulohumeral Muscular Dystrophy. *Neurologic Clinics*. **32**(3), 721–728.

Staudt, L.M., Dave, S. (2005) The biology of human lymphoid malignancies revealed by gene expression profiling. *Advances in immunology*. **87**, 163–208.

Stewart, M.D., Li, J., Wong, J. (2005) Relationship between Histone H3 Lysine 9 Methylation, Transcription Repression, and Heterochromatin Protein 1 Recruitment. *Molecular and Cellular Biology*. **25**(7), 2525–2538.

Stidham, R.W., Higgins, P.D.R. (2018) Colorectal Cancer in Inflammatory Bowel Disease. *Clinics in colon and rectal surgery*. **31**(3), 168–178.

Stojanovic, N., Hassan, Z., Wirth, M., Wenzel, P., Beyer, M., Schäfer, C., Brand, P., Kroemer, A., Stauber, R.H., Schmid, R.M., Arlt, A., Sellmer, A., Mahboobi, S., Rad, R., Reichert, M., Saur, D., Krämer, O.H., Schneider, G. (2017) HDAC1 and HDAC2 integrate the expression of p53 mutants in pancreatic cancer. *Oncogene*. **36**(13), 1804–1815.

Strahl, B.D., Allis, C.D. (2000) The language of covalent histone modifications. *Nature*. **403**(6765), 41–45.

Strimbu, K., Tavel, J.A. (2010) What are biomarkers? *Current opinion in HIV and AIDS*. **5**(6), 463–466.

Strzalka, W., Ziemienowicz, A. (2011) Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. *Annals of botany*. **107**(7), 1127–1140.

Stucki, M., Clapperton, J.A., Mohammad, D., Yaffe, M.B., Smerdon, S.J., Jackson, S.P. (2005) MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. *Cell*. **123**(7), 1213–1226.

Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, R., Haas, J., Schwede, T. (2020) QMEANDisCo-distance constraints applied on model quality estimation. *Bioinformatics (Oxford, England)*. **36**(6), 1765–1771.

Sullivan, C.A.W., Chung, G.G. (2008) Biomarker Validation: In Situ Analysis of Protein Expression Using Semiquantitative Immunohistochemistry-Based Techniques. *Clinical Colorectal Cancer*. **7**(3), 172–177.

Sun, Y.S., Zhao, Z., Yang, Z.N., Xu, F., Lu, H.J., Zhu, Z.Y., Shi, W., Jiang, J., Yao, P.P., Zhu, H.P. (2017) Risk factors and preventions of breast cancer. *International Journal of Biological*

Sciences. **13**(11), 1387–1397.
Sundaramoorthy, R., Hughes, A.L., El-Mkami, H., Norman, D.G., Ferreira, H., Owen-Hughes, T. (2018) Structure of the chromatin remodelling enzyme Chd1 bound to a ubiquitinylated nucleosome. *eLife*. **7**, 1–28.
Suryanarayanan, V., Rajavel, T., Devi, K.P., Singh, S.K. (2018) Structure based identification and biological evaluation of novel and potent inhibitors of PCAF catalytic domain. *International journal of biological macromolecules*. **120**(Pt A), 823–834.
Szerlong, H., Hinata, K., Viswanathan, R., Erdjument-Bromage, H., Tempst, P., Cairns, B.R. (2008) The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nature Structural & Molecular Biology*. **15**(5), 469–476.
Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., Shinkai, Y. (2008) G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *The EMBO Journal*. **27**(20), 2681–2690.
Tácio Quadros Santos Monteiro Fonseca, A., Zanoteli, E. (2018) Charcot-Marie-Tooth disease Enfermedad de Charcot-Marie-Tooth. *Revista Clínica Las Condes*. **29**(5), 521–529.
Takahashi, H., McCaffery, J.M., Irizarry, R.A., Boeke, J.D. (2006) Nucleocytosolic Acetyl-Coenzyme A Synthetase Is Required for Histone Acetylation and Global Transcription. *Molecular Cell*. **23**(2), 207–217.
Tang, L., Nogales, E., Ciferri, C. (2010) Structure and Function of SWI/SNF Chromatin Remodeling Complexes and Mechanistic Implications for Transcription. *Progress in biophysics and molecular biology*. **102**(2–3), 122–128.
Tang, Z., Li, Chenwei, Kang, B., Gao, G., Li, Cheng, Zhang, Z. (2017) GEPIA: A web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Research*. **45**(W1), W98–W102.
Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C. (2018) Europe PMC Funders Group Hyper-activation of HUSH complex function by Charcot-Marie- Tooth disease mutation in MORC2. . **49**(7), 1035–1044.
Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., Lehner, P.J. (2015) Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science*. **348**(6242), 1481–1485.
Thomas, M., Coope, A., Falkenberg, C., Dunlop, B.W., Czamara, D., Provencal, N., Craighead, W.E., Mayberg, H.S., Nemeroff, C.B., Binder, E.B., Nieratschker, V. (2020) Investigation of MORC1 DNA methylation as biomarker of early life stress and depressive symptoms. *Journal of Psychiatric Research*. **120**(October 2019), 154–162.
Tomura, M., Sakaue-Sawano, A., Mori, Y., Takase-Utsugi, M., Hata, A., Ohtawa, K., Kanagawa, O., Miyawaki, A. (2013) Contrasting Quiescent G0 Phase with Mitotic Cell Cycling in the Mouse Immune System. *PLoS ONE*. **8**(9), 1–10.
Torchy, M.P., Hamiche, A., Klaholz, B.P. (2015) Structure and function insights into the NuRD chromatin remodeling complex. *Cellular and Molecular Life Sciences*. **72**(13), 2491–2507.
Toriola, A.T., Colditz, G.A. (2013) Trends in breast cancer incidence and mortality in the United States: implications for prevention. *Breast Cancer Research and Treatment*. **138**(3), 665–673.
Travis, R.C., Key, T.J. (2003) Oestrogen exposure and breast cancer risk. *Breast Cancer Research*. **5**(5), 239–247.
Trochet, D., Bourdeaut, F., Janoueix-Lerosey, I., Deville, A., de Pontual, L., Schleiermacher, G., Coze, C., Philip, N., Frébourg, T., Munnich, A., Lyonnet, S., Delattre, O., Amiel, J. (2004) Germline mutations of the paired-like homeobox 2B (PHOX2B) gene in neuroblastoma. *American journal of human genetics*. **74**(4), 761–764.
Truebestein, L., Leonard, T.A. (2016) Coiled-coils: The long and short of it. *BioEssays*. **38**(9), 903–916.
Tsang, C.K., Bertram, P.G., Ai, W., Drenan, R., Zheng, X.F.S. (2003) Chromatin-mediated regulation

of nucleolar structure and RNA Pol I localization by TOR. *EMBO Journal*. **22**(22), 6045–6056.
Tsuda, M., Cho, K., Ooka, M., Shimizu, N., Watanabe, R., Yasui, A., Nakazawa, Y., Ogi, T., Harada, H., Agama, K., Nakamura, J., Asada, R., Fujiike, H., Sakuma, T., Yamamoto, T., Murai, J., Hiraoka, M., Koike, K., Pommier, Y., Takeda, S., Hirota, K. (2017) ALC1/CHD1L, a chromatin-remodeling enzyme, is required for efficient base excision repair. *PLOS ONE*. **12**(11), e0188320.
Turegun, B., Baker, R.W., Leschziner, A.E., Dominguez, R. (2018) Actin-related proteins regulate the RSC chromatin remodeler by weakening intramolecular interactions of the Sth1 ATPase. *Communications Biology*. **1**(1), 1–11.
Tyagi, M., Imam, N., Verma, K., Patel, A.K. (2016) Chromatin remodelers: We are the drivers!! *Nucleus*. **7**(4), 388–404.
Ud-Din, A.I.M.S., Tikhomirova, A., Roujeinikova, A. (2016) Structure and functional diversity of GCN5-related n-acetyltransferases (GNAT). *International Journal of Molecular Sciences*. **17**(7).
Umapathy, G., Mendoza-Garcia, P., Hallberg, B., Palmer, R.H. (2019) Targeting anaplastic lymphoma kinase in neuroblastoma. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. **127**(5), 288–302.
Utley, R.T., Côté, J. (2003) The MYST family of histone acetyltransferases. In J. L. Workman, ed. *Current Topics in Microbiology and Immunology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 203–236.
Vadlamudi, R.K., Kumar, R. (2003) p21-activated kinases in human cancer. *Cancer and Metastasis Reviews*. **22**(4), 385–393.
Varga-Weisz, P. (2001) ATP-dependent chromatin remodeling factors: Nucleosome shufflers with many missions. *Oncogene*. **20**(24 REV. ISS. 3), 3076–3085.
Vedadi, M., Arrowsmith, C.H., Allali-Hassani, A., Senisterra, G., Wasney, G.A. (2010) Biophysical characterization of recombinant proteins: A key to higher structural genomics success. *Journal of Structural Biology*. **172**(1), 107–119.
Verdone, L., Agricola, E., Caserta, M., Di Mauro, E. (2006) Histone acetylation in gene regulation. *Briefings in Functional Genomics and Proteomics*. **5**(3), 209–221.
Verdone, L., La Fortezza, M., Ciccarone, F., Caiafa, P., Zampieri, M., Caserta, M. (2015) Poly(ADP-ribosyl)ation affects histone acetylation and transcription. *PLoS ONE*. **10**(12), 1–14.
Verschueren, K.H.G., Blanchet, C., Felix, J., Dansercoer, A., De Vos, D., Bloch, Y., Van Beeumen, J., Svergun, D., Gutsche, I., Savvides, S.N., Verstraete, K. (2019) Structure of ATP citrate lyase and the origin of citrate synthase in the Krebs cycle. *Nature*. **568**(7753), 571–575.
Victora, G.D., Nussenzweig, M.C. (2012) Germinal Centers. *Annual Review of Immunology*. **30**(1), 429–457.
Vidanes, G.M., Bonilla, C.Y., Toczyski, D.P. (2005) Complicated tails: Histone modifications and the DNA damage response. *Cell*. **121**(7), 973–976.
Vignali, M., Hassan, A.H., Neely, K.E., Workman, J.L. (2000) ATP-dependent chromatin-remodeling complexes. *Molecular and cellular biology*. **20**(6), 1899–910.
Vikstrom, I., Tarlinton, D.M. (2011) B cell memory and the role of apoptosis in its formation. *Molecular Immunology*. **48**(11), 1301–1306.
Vocadlo, D.J. (2012) O-GlcNAc processing enzymes: Catalytic mechanisms, substrate specificity, and enzyme regulation. *Current Opinion in Chemical Biology*. **16**(5–6), 488–497.
Wahba, H.A., El-Hadaad, H.A. (2015) Current approaches in treatment of triple-negative breast cancer. *Cancer Biology and Medicine*. **12**(2), 106–116.
Walls, D., Loughran, S.T. (2011) Purification of proteins fused to glutathione S-tranferase. *Methods in molecular biology (Clifton, N.J.)*. **681**, 151–175.
Wang, Chen-Yu, Brand, H., Shaw, N.D., Talkowski, M.E., Lee, J.T. (2019) Role of the Chromosome Architectural Factor SMCHD1 in X-Chromosome Inactivation, Gene Regulation, and Disease

in Humans. *Genetics*. **213**(2), 685 LP – 703.
Wang, Li Hui, Wu, C.F., Rajasekaran, N., Shin, Y.K. (2019) Loss of tumor suppressor gene function in human cancer: An overview. *Cellular Physiology and Biochemistry*. **51**(6), 2647–2693.
Wang, Yulin, Zhao, W., Liu, X., Guan, G., Zhuang, M. (2019) ARL3 is downregulated and acts as a prognostic biomarker in glioma. *Journal of Translational Medicine*. **17**(1), 1–15.
Wapenaar, H., Dekker, F.J. (2016) Histone acetyltransferases : challenges in targeting bi-substrate enzymes. *Clinical Epigenetics*, 1–11.
Watanabe, M., Yanagi, Y., Masuhiro, Y., Yano, T., Yoshikawa, H., Yanagisawa, J., Kato, S. (1998) A Putative Tumor Suppressor, TSG101, Acts as a Transcriptional Suppressor through Its Coiled-Coil Domain. *Biochemical and Biophysical Research Communications*. **245**(3), 900–905.
Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. (2018) SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*. **46**(W1), W296–W303.
Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J. (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **25**(9), 1189–1191.
Watson, M.L., Zinn, A.R., Inoue, N., Hess, K.D., Cobb, J., Handel, M.A., Halaban, R., Duchene, C.C., Albright, G.M., Moreadith, R.W. (1998) Identification of morc (microrchidia), a mutation that results in arrest of spermatogenesis at an early meiotic stage in the mouse. *Proceedings of the National Academy of Sciences of the United States of America*. **95**(24), 14361–14366.
Waugh, D.S. (2011) An overview of enzymatic reagents for the removal of affinity tags. *Protein Expression and Purification*. **80**(2), 283–293.
Weigert, O., Weinstock, D.M. (2012) The evolving contribution of hematopoietic progenitor cells to lymphomagenesis. *Blood*. **120**(13), 2553–2561.
Wellen, K.E., Hatzivassiliou, G., Sachdeva, U.M., Bui, T. V, Cross, J.R., Thompson, C.B. (2009) ATP-Citrate Lyase Links Cellular Metabolism to Histone Acetylation. *Science*. **324**(5930), 1076 LP – 1080.
Wiegand, T., Cadalbert, R., Lacabanne, D., Timmins, J., Terradot, L., Böckmann, A., Meier, B.H. (2019) The conformational changes coupling ATP hydrolysis and translocation in a bacterial DnaB helicase. *Nature Communications*. **10**(1), 1–11.
Wiese, C., Hinz, J.M., Tebbs, R.S., Nham, P.B., Urbin, S.S., Collins, D.W., Thompson, L.H., Schild, D. (2006) Disparate requirements for the Walker A and B ATPase motifs of human RAD51D in homologous recombination. *Nucleic Acids Research*. **34**(9), 2833–2843.
Williams, C., Lin, C.Y. (2013) Oestrogen receptors in breast cancer: Basic mechanisms and clinical implications. *Ecancermedicalscience*. **7**(1), 1–12.
Williams, G.H., Stoeber, K. (2012) The cell cycle and cancer. *Journal of Pathology*. **226**(2), 352–364.
Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*. **275**(1), 1–21.
Wood, W.N., Smith, K.D., Ream, J.A., Kevin Lewis, L. (2017) Enhancing yields of low and single copy number plasmid DNAs from Escherichia coli cells. *Journal of Microbiological Methods*. **133**, 46–51.
Wright, G., Tan, B., Rosenwald, A., Hurt, E.H., Wiestner, A., Staudt, L.M. (2003) A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America*. **100**(17), 9991–9996.
Wu, C. (1997) Chromatin remodeling and the control of gene expression. *Journal of Biological Chemistry*. **272**(45), 28171–28174.

Wu, X., Wu, D., Lu, Z., Chen, W., Hu, X., Ding, Y. (2009) A novel method for high-level production of TEV protease by superfolder GFP tag. *Journal of biomedicine & biotechnology*. **2009**, 591923.

Xenaki, G., Ontikatze, T., Rajendran, R., Stratford, I.J., Dive, C., Krstic-Demonacos, M., Demonacos, C. (2008) PCAF is an HIF-1α cofactor that regulates p53 transcriptional activity in hypoxia. *Oncogene*. **27**(44), 5785–5796.

Xi, Z., Gao, Y., Sirinakis, G., Guo, H., Zhang, Y. (2012) Single-molecule observation of helix staggering, sliding, and coiled coil misfolding. *Proceedings of the National Academy of Sciences of the United States of America*. **109**(15), 5711–5716.

Xiang, C., Zhang, S., Dong, X., Ma, S., Cong, S. (2018) Transcriptional dysregulation and post-translational modifications in polyglutamine diseases: From pathogenesis to potential therapeutic strategies. *Frontiers in Molecular Neuroscience*. **11**(May).

Xiao, H., Sandaltzopoulos, R., Wang, H.M., Hamiche, A., Ranallo, R., Lee, K.M., Fu, D., Wu, C. (2001) Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Molecular Cell*. **8**(3), 531–543.

Xie, H.Y., Zhang, T.M., Hu, S.Y., Shao, Z.M., Li, D.Q. (2019) Dimerization of MORC2 through its C-terminal coiled-coil domain enhances chromatin dynamics and promotes DNA repair. *Cell Communication and Signaling*. **17**(1), 1–11.

Xu, G., Jaffrey, S.R. (2013) Proteomic identification of protein ubiquitination events. *Biotechnology and Genetic Engineering Reviews*. **29**(1), 73–109.

Xu, Y., McKenna, R.W., Hoang, M.P., Collins, R.H., Kroft, S.H. (2002) Composite Angioimmunoblastic T-Cell Lymphoma and Diffuse Large B-Cell LymphomaA Case Report and Review of the Literature. *American Journal of Clinical Pathology*. **118**(6), 848–854.

Xue-Franzén, Y., Henriksson, J., Bürglin, T.R., Wright, A.P.H. (2013) Distinct roles of the Gcn5 histone acetyltransferase revealed during transient stress-induced reprogramming of the genome. *BMC Genomics*. **14**(1).

Yamada, T., Mizuno, K.I., Hirota, K., Kon, N., Wahls, W.P., Hartsuiker, E., Murofushi, H., Shibata, T., Ohta, K. (2004) Roles of histone acetylation and chromatin remodeling factor in a meiotic recombination hotspot. *EMBO Journal*. **23**(8), 1792–1803.

Yamane, K., Toumazou, C., Tsukada, Y. ichi, Erdjument-Bromage, H., Tempst, P., Wong, J., Zhang, Y. (2006) JHDM2A, a JmjC-Containing H3K9 Demethylase, Facilitates Transcription Activation by Androgen Receptor. *Cell*. **125**(3), 483–495.

Yang, X.-J. (2004) Lysine acetylation and the bromodomain: a new partnership for signaling. *BioEssays*. **26**(10), 1076–1087.

Yang, Z., Zhuang, Q., Hu, G., Geng, S. (2019) MORC4 is a novel breast cancer oncogene regulated by miR-193b-3p. *Journal of Cellular Biochemistry*. **120**(3), 4634–4643.

Yao, W., Beckwith, S.L., Zheng, T., Young, T., Dinh, T. Van, Ranjan, A., Morrison, A.J. (2015) Assembly of the Arp5 (actin-related protein) subunit involved in distinct ino80 chromatin remodeling activities. *Journal of Biological Chemistry*. **290**(42), 25700–25709.

Yi, S.J., Kim, K. (2018) Histone tail cleavage as a novel epigenetic regulatory mechanism for gene expression. *BMB Reports*. **51**(5), 211–218.

Yu, J., Qin, B., Lou, Z. (2020) Ubiquitin and ubiquitin-like molecules in DNA double strand break repair. *Cell and Bioscience*. **10**(1), 13.

van Zelm, M.C., Szczepanski, T., van der Burg, M., van Dongen, J.J.M. (2007) Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion. *The Journal of experimental medicine*. **204**(3), 645–655.

Zentner, G.E., Hurd, E.A., Schnetz, M.P., Handoko, L., Wang, C., Wang, Z., Wei, C., Tesar, P.J., Hatzoglou, M., Martin, D.M., Scacheri, P.C. (2010) CHD7 functions in the nucleolus as a positive regulator of ribosomal RNA biogenesis. *Human Molecular Genetics*. **19**(18), 3491–3501.

Zhang, Fang Lin, Cao, J.L., Xie, H.Y., Sun, R., Yang, L.F., Shao, Z.M., Li, D.Q. (2018) Cancer-associated MORC2-mutant M276I regulates an hnRNPM-mediated CD44 splicing switch to promote invasion and metastasis in triple-negative breast cancer. *Cancer Research*. **78**(20), 5780–5792.

Zhang, L., Li, D.-Q. (2019) MORC2 regulates DNA damage response through a PARP1-dependent pathway. *Nucleic Acids Research*. **47**(16), 8502–8520.

Zhang, P., Torres, K., Liu, X., Liu, C., E. Pollock, R. (2016) An Overview of Chromatin-Regulating Proteins in Cells. *Current Protein & Peptide Science*. **17**(5), 401–410.

Zhang, R., Erler, J., Langowski, J. (2017) Histone Acetylation Regulates Chromatin Accessibility: Role of H4K16 in Inter-nucleosome Interaction. *Biophysical Journal*. **112**(3), 450–459.

Zhang, Y., Klein, B.J., Cox, K.L., Bertulat, B., Tencer, A.H., Holden, M.R., Wright, G.M., Black, J., Cardoso, M.C., Poirier, M.G., Kutateladze, T.G. (2019) Mechanism for autoinhibition and activation of the MORC3 ATPase. *Proceedings of the National Academy of Sciences of the United States of America*. **116**(13), 6111–6119.

Zhang, Z., Sauter, N.K., Van Den Bedem, H., Snell, G., Deacon, A.M. (2006) Automated diffraction image analysis and spot searching for high-throughput crystal screening. *Journal of Applied Crystallography*. **39**(1), 112–119.

Zhang, Zhihui, Wang, X., Xin, J., Ding, Z., Liu, S., Fang, Q., Yang, N., Xu, R. min, Cai, G. (2018) Architecture of SWI/SNF chromatin remodeling complex. *Protein and Cell*. **9**(12), 1045–1049.

Zhao, Y., Dunker, W., Yu, Y.T., Karijolich, J. (2018) The role of noncoding RNA pseudouridylation in nuclear gene expression events. *Frontiers in Bioengineering and Biotechnology*. **6**(FEB).

Ziv, Y., Bielopolski, D., Galanty, Y., Lukas, C., Taya, Y., Schultz, D.C., Lukas, J., Bekker-Jensen, S., Bartek, J., Shiloh, Y. (2006) Chromatin relaxation in response to DNA double-strand breaks is modulated by a novel ATM- and KAP-1 dependent pathway. *Nature Cell Biology*. **8**(8), 870–876.

Zofall, M., Persinger, J., Bartholomew, B. (2004) Functional role of extranucleosomal DNA and the entry site of the nucleosome in chromatin remodeling by ISW2. *Molecular and cellular biology*. **24**(22), 10047–10057.

Zoncu, R., Efeyan, A., Sabatini, D.M. (2011) MTOR: From growth signal integration to cancer, diabetes and ageing. *Nature Reviews Molecular Cell Biology*. **12**(1), 21–35.

Zühlke, C., Hellenbroich, Y., Dalski, A., Hagenah, J., Vieregge, P., Riess, O., Klein, C., Schwinger, E. (2001) Different types of repeat expansion in the TATA-binding protein gene are associated with a new form of inherited ataxia. *European Journal of Human Genetics*. **9**(3), 160–164.

Chapter 8 Appendices

Appendix 8.1: MORC4 full length sequence indicating the location of the mutations.

Using FL-MORC4 (NN061) in pcDNA3-N-FLAG-LIC as template

ATGCTCCTGTACCGAGGGGCCCCCGCCGGCCCTGGCGCGCCGGGCTGCGGGCTGGCCCGGCC
GGCGGCGGCCCGCAGGCCTTCGGGATCCGCCTGAGCACGATGAGCCCCCGCTACCTCCAGAG
CAACTCCAGCAGCCACACGCGACCCTTCAGTGCCATCGCGGAGCTG

N60A (N39A in MORC2)

CTAGAT**AAT**GCTGTAGATCCAGATGTATCTGCCAGGACGGTCTTTATAGATGTTGAG

D88A (D68A in MORC2)

GAGGTCAAGAATAAATCTTGTTTGACCTTTACC**GAT**GATGGATGTGGGATGACACCTCATAA
ACTACACCGAATGCTCAGCTTTGGCTTTACAGATAAAGTAATAAAGAAGAGC

K126R

CAGTGTCCCATTGGGGTCTTTGGTAATGGTTTC**AAG**TCAGGCTCCATGCGGCTAGGAAAGGA
CGCCCTTGTCTTCACCAAGAATGGGGGTACTCTCACTGTTGGACTTCTATCACAGACCTATC
TGGAATGTGTCCAGGCCCAGGCAGTTATTGTACCAATTGTTCCATTCAACCAGCAAAACAAA
AAAATGATTATTACCGAGGATTCATTGCCCAGCCTAGAAGCCATCTTGAACTATTCCATTTT
CAACCGTGAAAATGACCTGCTGGCCCAGTTTGATGCCATCCCAGGCAAAAAAGGCACTCGTG
TTCTCATTTGGAACATCCGCAGAAATAAAAATGGAAAATCTGAGTTGGACTTTGATACAGAT
CAATATGACATCCTGGTATCAGACTTTGACACAGAAGAAAAAATGACTGGCGGTGTTACCTC
TGAGCTACCAGAAACAGAATATTCTTTAAGGGCATTTTGTGGTATTCTATACATGAAGCCAC
GCATGAAAATTTTTCTGCGTCAAAAGAAGGTGACTACCCAGATGATTGCCAAGAGCCTGGCC
AATGTAGAATATGATACATATAAACCTACCTTCACAAATAAGCAGGTGAGAATCACCTTTGG
GTTCTCTTGCAAGAATAGTAACCAGTTTGGAATAATGATGTATCATAACAACCGACTCATAA
AATCTTTTGAGAAGGTGGGGTGCCAGGTGAAGCCAACTCGTGGAGAAGGTGTAGGAGTAATT
GGAGTCATTGAGTGCAATTTCCTAAAACCTGCCTACAACAAACAAGACTTTGAGTATACCAA
GGAGTACCGGCTAACAATAAATGCCCTTGCCCAGAAGCTCAATGCTTACTGGAAGGAAAAAA
CATCTCAAGATAATTTTGAGACCTCAACTGTAGCCAGGCCAATA

W435K (W419K in MORC3)

CCGAAGGTTCCTGACCAGACATGGGTTCAGTGTGATGAGTGTCTTAAA**TGG**AGAAAGCTTCC
TGGGAAGATTGATCCATCCATGTTACCTGCAAGATGGTTTTGTTATTATAATTCCCATCCAA
AGTACAGGAGATGCTCTGTTCCAGAGGAACAAGAACTCACTGATGAAGACCTGTGCTTGAGC
AAAGCTAAGAAACAAGAACAAACTGTTGAGGAGAAGAAGAAGATGCCTATGGAAAATGAGAA
CCACCAGGTATTCAGTAATCCACCAAAGATCCTTACTGTTCAAGAAATGGCTGAATTGAATA
ACAAGACAATTGGATATGAGGGAATTCATAGCCCTAGTGTGCTTCCTTCTGGTGGAGAAGAA
AGCAGATCACCATCTCTTCAACTTAAGCCTCTGGATTCCAGTGTTTTACAGTTTTCCAGTAA
GTACAAATGGATCCTAGGTGAAGAACCGGTGGAGAAACGAAGAAGGCTCCAGAATGAGATGA
CAACACCTTCTCTAGATTATTCCATGCCTGCTCCTTACAGGAGGGTAGAAGCACCTGTTGCC
TACCCAGAAGGGGAGAACAGCCATGATAAATCGAGTTCTGAGAGAAGTACACCACCATACCT
TTTCCCAGAATACCCAGAAGCAAGCAAGAATACAGGTCAGAATAGGGAGGTTTCAATTCTGT
ATCCAGGGGCCAAAGACCAACGCCAGGGGTCCCTGCTTCCTGAAGAATTAGAAGATCAGATG
CCAAGATTGGTGGCAGAAGAATCTAACAGAGGTAGCACAACCATAAACAAAGAAGAAGTCAA
CAAGGGACCTTTTGTAGCTGTTGTGGGTGTTGCCAAAGGTGTTAGAGATTCAGGAGCTCCCA
TTCAGCTGATCCCTTTTAACAGAGAGGAGCTTGCTGAGAGACGAAAAGCAGTTGAATCCTGG
AACCCAGTGCCTTATTCTGTGGCCTCTGCTGCAATCCCTGCTGCAGCCATTGGGGAGAAAGC
AAGAGGCTATGAGGAGAGCGAAGGTCATAATACACCAAAGTTGAAGAACCAGAGAGAGCTGG
AAGAATTGAAGAGAACCACAGAAAAATTGGAACGTGTTTTGGCTGAAAGGAATTTGTTCCAG
CAAAAGGTGGAGGAGCTGGAACAGGAGAGGAATCACTGGCAGTCTGAATTCAAGAAAGTCCA

ACATGAATTGGTGATCTACAGTACCCAGGAGGCGGAAGGCTTGTACTGGAGCAAGAAACACA
TGGGTTATCGCCAAGCTGAATTCCAGATTCTGAAAGCTGAGCTGGAAAGAACCAAAGAGGAA
AAGCAAGAGTTAAAAGAGAAACTGAAGGAAACAGAGACACACCTGGAAATGCTGCAGAAGGC
TCAGGGCTTTGGCAAAGCTTACGCGGCTACGTATCCACGTCAGCTATCTCCTTACTTCTGTC
CTCCCTCACTTGGAGCTTCGTGA

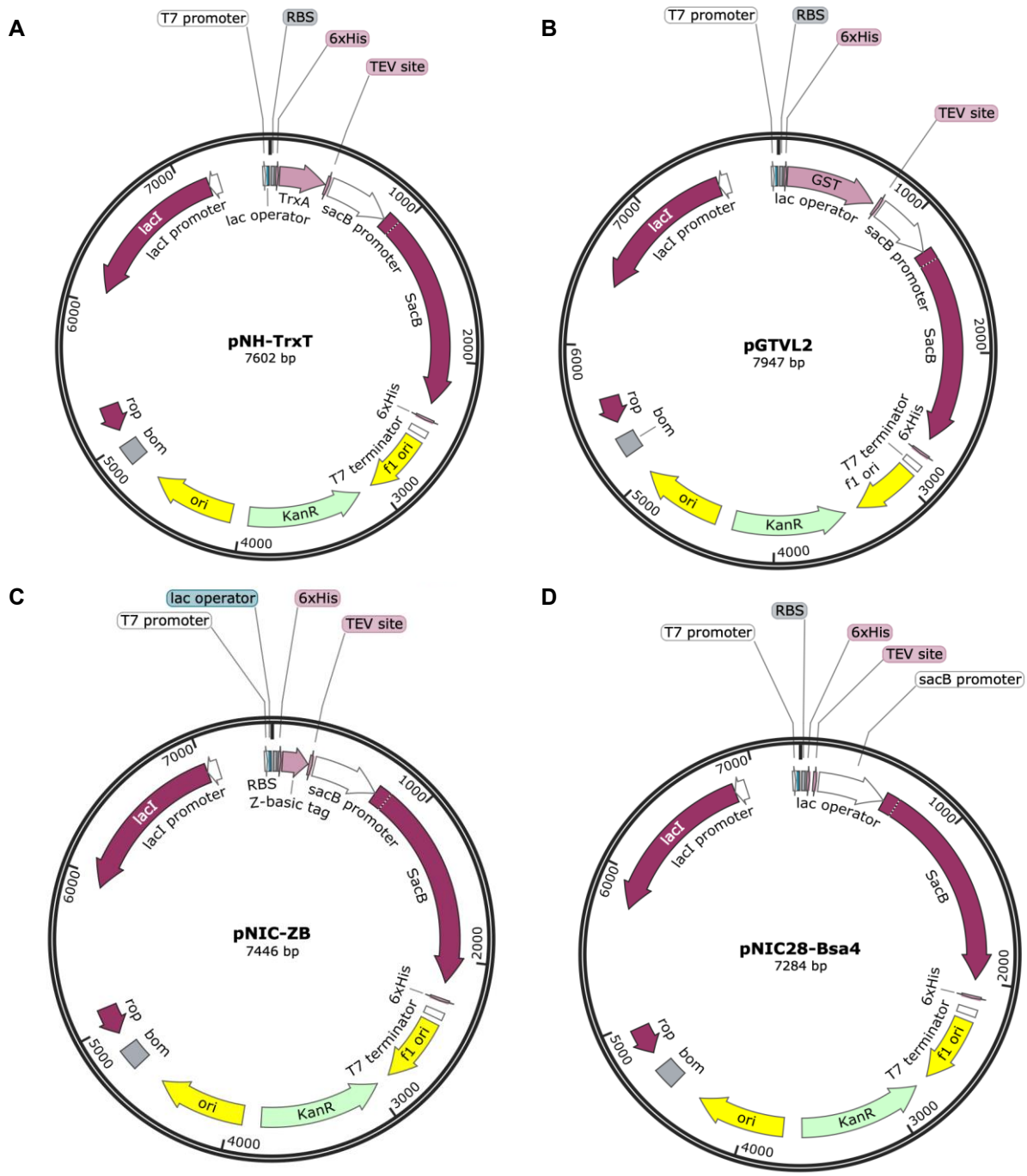
Appendix 8.2: Table showing MORC4 sequences including the sequence containing the mutation and designed mutagenic primers.

Primer	Mutation	Sequence	Primer Length	Forward Primer	Reverse Primer
FLMORC4_N60A_F	N60A	GAT**AAT**G CTGTAGAT CCAGAT	21	GATGCTGCT GTAGATCC AGAT	
FLMORC4_N60A_R	N60A	GCCATCGC GGAGCTG CTA	18		TAGCAGCT CCGCGATG GC
FLMORC4_D88A_F	D88A	ACC**GAT**G ATGGATGT GGGATG	21	ACCGCTGA TGGATGTG GGATG	
FLMORC4_D88A_R	D88A	AAATCTTG TTTGACCT TT	18		AAAGGTCA AACAAGAT TT
FLMORC4_K126R_F	K126R	TTC**AAG**TC AGGCTCC ATGCGG	21	TTCCGGTCA GGCTCCAT GCGG	
FLMORC4_K126R_R	K126R	GGGGTCTT TGGTAATG GT	18		ACCATTACC AAAGACCC C
FLMORC4_W435K_F	W435K	AAA**TGG**A GAAAGCT TCCTGGG	21	AAAAAGAG AAAGCTTC CTGGG	
FLMORC4_W435K_R	W435K	CAGTGTG ATGAGTGT CTT	18		AAGACACT CATCACACT G

Appendix 8.3: **Cell line details** (activated B-cell (ABC), germinal centre B-cell (GCB), MYCN Proto-Oncogene, BHLH Transcription Factor (MYCN), Oestrogen receptor (ER). Maintained with the support of Dr. Christopher Cooper (The University of Huddersfield) indicated with a *.

Type	Cell line	Disease	Media
N/A	COS-1	Monkey Kidney (SV40+)	DMEM
Diffuse large B-cell lymphomas (DLBCL)	OCI-Ly3	ABC DLBCL	RPMI
	OCI-Ly10	ABC DLBCL	
	RIVA	ABC DLBCL	
	SU-DHL-2	ABC DLBCL	
	HBL-1	ABC DLBCL	
	HLY-1	ABC DLBCL	
	SUD-HL6	GCB DLBCL	
	SUD-HL10	GCB DLBCL	
	DB	DLBCL GCB	
	MIEU	DLBCL GCB	
	LIB	DLBCL	
Burkitts Lymphoma	Daudi	Burkitts	RPMI
	Raji	Burkitts	
Myeloma	NCI H929	Myeloma	RPMI
	JJN.3*	Myeloma	
	RPMI8226*	Myeloma	
T-cell acute lymphoblastic leukaemia (T-ALL)	Jurkat	T-ALL	RPMI
	DU528	T-ALL	
	RPMI8402*	T-ALL	
	(CCRF-) CEM*	T-ALL	
	Hut78	Cutaneous T-cell lymphoma	RPMI
Colorectal adenocarcinoma	HCT29	Colorectal adenocarcinoma	DMEM
	HCT116+/+	Colorectal adenocarcinoma	
	SW48	Colorectal adenocarcinoma	
	Hela*	Cervical carcinoma	DMEM
	U2OS*	Osteosarcoma	
	HEK293FT*	Transformed human embryonic kidney	
Neuroblastoma	SK-N-BE(2c)*	NB (MYCN+)	DMEM
	IMR32*	NB (MYCN+)	
	Kelly	NB (MYCN+)	
	NB19	NB (MYCN+)	
	SH-SY5Y*	NB (MYCN-)	
	SH-SY5Y (PM)*	NB (MYCN-)	
	SK-N-SH*	NB (MYCN-)	50/50 DMEM/RPMI
	SK-N-SH (PM)*	NB (MYCN-)	
	SK-N-AS*	NB (MYCN-)	DMEM
	SHEP	NB (MYCN?)	
	PSN1*	Pancreatic adenocarcinoma	DMEM
	Mia-Pa-Ca-2	Pancreatic carcinoma	
Breast adenocarcinoma	MCF7*	Breast adenocarcinoma (ER+ve)	DMEM

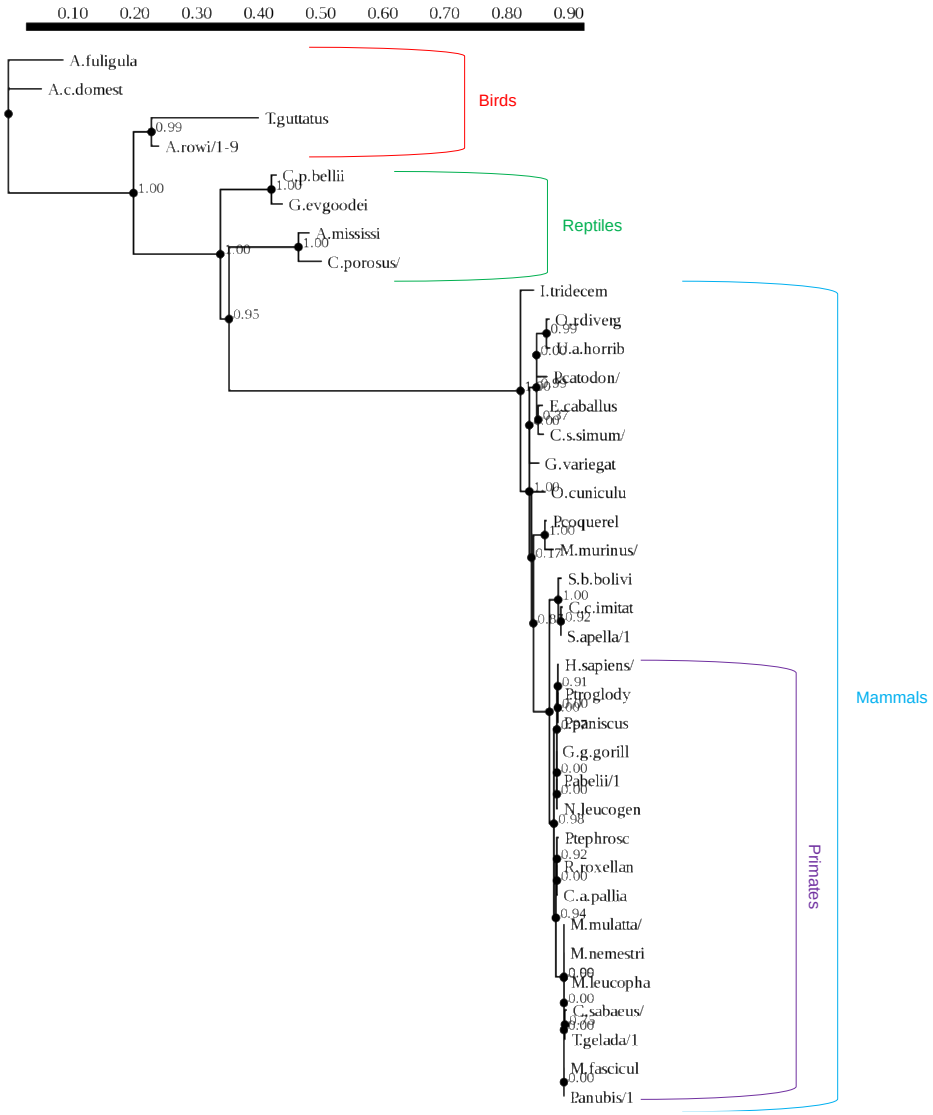
	MDA-MB-231*	Breast adenocarcinoma (ER-ve)
	MDA-MB-468*	Breast adenocarcinoma (ER-ve)



Appendix 8.4: **Vector maps of pET *E. coli* expression vectors used in LIC. (A)** pNH-TrxT **(B)** pGTVL2 **(C)** pNIC-ZB **(D)** pNIC28-Bsa4. All pET expression vectors were made at the Structural Genomics Consortium, Oxford, United Kingdom. Vector maps created by author using SnapGene software.

Appendix 8.5: **Table of PCNA primers.**

Primer name	Primer sequence (5'-3')
PCNA-f000	TACTTCCAATCCATGTTCGAGGCGCGCCTGGTCC
PCNA-r000	TATCCACCTTTACTGTCAAGATCCTTCTTCATCCTCGATC



Appendix 8.5: **Phylogenetic tree analysis of MORC4 species in vertebrates.** MORC4 protein sequences in vertebrate species were identified using NCBI and aligned using MUSCLE in Jalview software. Gblocks was used to remove gaps in sequences and PhyML online server was used to investigate the maximum likelihood and to generate the tree. Birds are highlighted in red; reptiles are highlighted in green; mammals are highlighted in blue and primates are highlighted in purple.

Appendix 8.6: **Zf-CW MORC4 proteins crystallisation screens details.** Showing the screen details, protein concentrations, temperature information and protein purification details.

Domain	Protein[(aa)]	Temp (°C)	Screen	Concentration of protein (mg/ml)	Size drops (µL)	Drop ratio upper well (protein: mother liquor) (µL)	Drop ratio lower well (protein: mother liquor) (µL)	Protein purification	Final protein buffer
Zf-CW	MORC4$^{420-757}$	20	JCSG-*plus*™	14.8	2	2:1	1:1	IMAC-TEV-IMAC-QIEX-SEC	SEC buffer - 20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT
Zf-CW	MORC4$^{420-677}$	20	JCSG-*plus*™	9.0	2	2:1	1:1	IMAC-TEV-IMAC-QIEX-SEC	SEC buffer - 20 mM HEPES pH 7.5, 500 mM NaCl, 5% glycerol, 1 mM DTT

			PEG 400 (%)											
			34	36	38	40	42	44	34	36	38	40	42	44
		Well	1	2	3	4	5	6	7	8	9	10	11	12
Li2S04 (mM)	150	A	pH 7.5						pH 8.0					
	175	B												
	200	C												
	225	D												
	150	E	pH 8.5						pH 9.0					
	175	F												
	200	G												
	225	H												

Appendix 8.7: **Design of MORC4$^{29\text{-}480}$ protein crystallography trial optimisation screen, JCSG_D7_FUP_1, created from the initial JCSG-*plus*™ screen.** Li_2SO_4 was used in a range of 150-225 mM, PEG 400 was used in a range of 34-44% and pH range was from 7.5-9.0.

			PEG 400 (%)											
			16	18	20	22	24	26	28	30	32	34	36	38
		Well	1	2	3	4	5	6	7	8	9	10	11	12
Li2S04 (mM)	100	A	pH 9.0											
	125	B												
	150	C												
	175	D												
	100	E	pH 9.5											
	125	F												
	150	G												
	175	H												

Appendix 8.8: **Design of MORC4$^{29\text{-}480}$ protein crystallography trial optimisation screen, JCSG_D7_FUP_2, created from the JCSG_D7_FUP_1 optimisation screen.** Li_2SO_4 was used in a range of 100-175 mM, PEG 400 was used in a range of 16-38% and pH range was from 9.0 to 9.5.

				Jeffamine M-600 (%)											
				8	10	12	14	16	18	20	22	24	26	28	30
			Well	1	2	3	4	5	6	7	8	9	10	11	12
	0.1M HEPES	pH 7.0	A												
		pH 7.5	B												
	0.1M TRIS	pH 7.5	C												
		pH 8.0	D												
			E												
			F												
			G												
			H												

Appendix 8.9: **Design of MORC4$^{29\text{-}480}$ protein crystallisation trial optimisation screen, JCSG_F04_FUP_1, created from the JCSG-*Plus*™ screen.** Jeffamine® M-600 was used in concentrations varying from 8-30%, 0.1 M HEPES and Tris were used with the pH ranging from 7.0-8.0.

Appendix 8.10: **ATPase-Zf-CW MORC4 protein JCSG-*plus*™ and optimisation crystallisation screens details.** Showing the screen details, protein concentrations, temperature information and protein purification details.

Domain	Protein[(aa)]	Temp (°C)	Screen	Concentration of protein (mg/ml)	Preparation of protein and ligand	Fresh or frozen	Protein purification	Final protein buffer
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG-*plus*™	6.2	Dilute-mix-concentrate	Frozen	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG_D07_FUP_1	5.8	Dilute-mix-concentrate	Frozen	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG_D07_FUP_2	5.7	Dilute-mix-concentrate	Frozen	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG-*plus*™	4.0	Dilute-concentrate-mix	Frozen	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG-*plus*™	4.0	Dilute-concentrate-mix	Fresh	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	JCSG_F04_FUP_1	2.3	Dilute-concentrate-mix	Frozen	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT

			PEG 1500 (%)											
			20	22	24	26	28	30	20	22	24	26	28	30
		Well	1	2	3	4	5	6	7	8	9	10	11	12
pH	5.5	A	0.1 M MIB						0.1 M MMT					
	6.0	B												
	6.5	C												
	7.0	D												
	7.5	E												
	8.0	F												
	8.5	G												
	9.0	H												

Appendix 8.11: **Design of MORC4$^{29\text{-}480}$ protein crystallisation trial optimisation screen, PACT_B03-B06_D03-D06_FUP_1, created from the PACT-*Premier*™ screen.** 0.1 M Sodium malonate dibasic monohydrate, Imidazole, Boric acid (MIB) and DL-Malic acid, 2-(N-morpholino)ethanesulfonic acid monohydrate, Tris (MMT) were used as well as varying concentrations of PEG 1500 in the range of 20-30% at a pH ranging from 5.5-9.0.

Appendix 8.12: **Types of PEG smears and composition**

PEG smear	Composition
PEG Smear Low (50% v/v)	12.5% v/v PEG 400, 12.5% v/v PEG 500 MME, 12.5% v/v PEG 600, 12.5% w/v PEG 1000.
PEG Smear Medium (50% v/v)	12.5% w/v PEG 3350, 12.5% w/v PEG 4000, 12.5% w/v PEG 2000, 12.5% w/v PEG 5000 MME.
PEG Smear High (50% v/v)	16.67% w/v PEG 8000, 16.67% w/v PEG 10000, 16.67% w/v PEG 6000.
PEG Smear Broad (50% v/v)	4.55% v/v PEG 400, 4.55% v/v PEG 500 MME, 4.55% v/v PEG 600, 4.55% w/v PEG 1000, 4.55% w/v PEG 2000, 4.55% w/v PEG 3350, 4.55% w/v PEG 4000, 4.55% w/v PEG 5000 MME, 4.55% w/v PEG 6000, 4.55% w/v PEG 8000, 4.55% w/v PEG 10000.

			PEG (%)											
			18	20	22	24	26	28	30	32	34	36	38	40
		Well	1	2	3	4	5	6	7	8	9	10	11	12
Type of PEG	PEG300	A												
	PEG400	B												
	PEG500 MME	C												
	PEG600	D												
	PEG1000	E												
	PEG smear low	F												
	ADDITIVE SCREEN	G												
	ADDITIVE SCREEN	H												

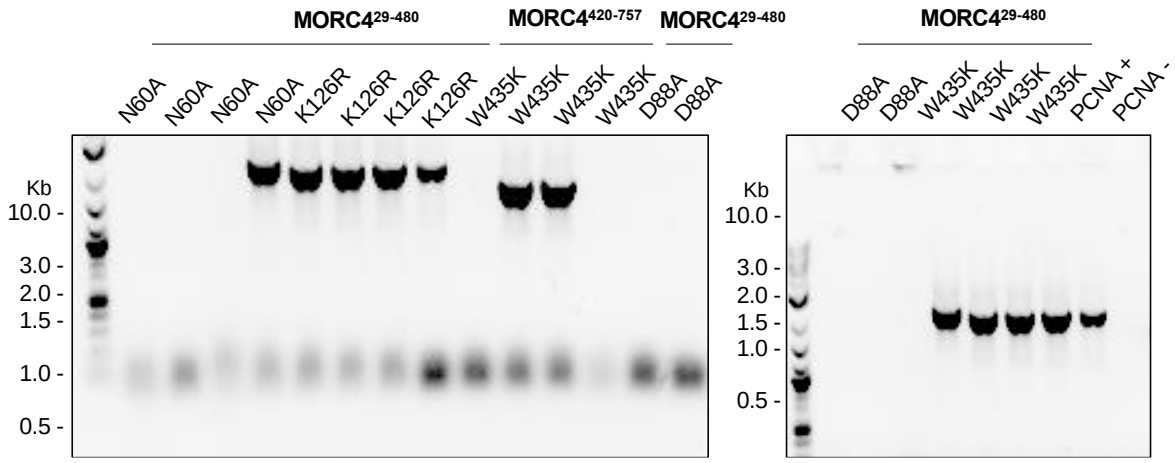
Appendix 8.13: **Design of MORC4$^{29\text{-}480}$ protein crystallisation trial optimisation screen, BCS_F02_FUP_1, created from the BCS screen.** The base in each well was 0.2 M NaCl, 100 mM Tris pH 8.5 and 5% v/v glycerol. Different types of PEGs were also used in this screen as well as a PEG smear of low molecular weight and an additive screen. PEGs were used in the range of 18-40%.

			PEG (%)						PEG (%)					
			10	12	14	16	18	20	10	12	14	16	18	20
		Well	1	2	3	4	5	6	7	8	9	10	11	12
Type of PEG	PEG600	A	5% glycerol						10% glycerol					
	PEG1000	B												
	PEG smear low	C												
	PEG600	D	15% glycerol						25% glycerol					
	PEG1000	E												
	PEG smear low	F												
	ADDITIVE SCREEN	G												
	ADDITIVE SCREEN	H												

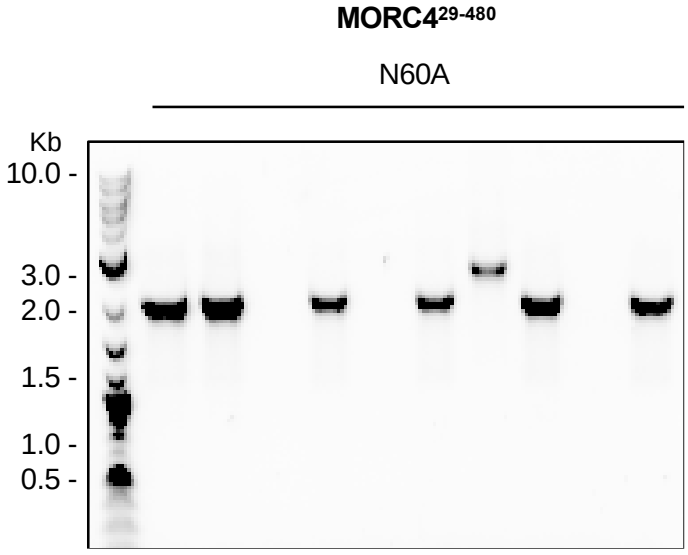
Appendix 8.14: **Design of MORC4$^{29\text{-}480}$ protein crystallisation trial optimisation screen, BCS_F02_FUP_2, created from the BCS_F02_FUP_1 optimisation screen.** The base in each well was 0.2 M NaCl, 100 mM Tris pH 8.5 and 5% v/v glycerol. Low molecular weight PEGs were also used in this screen as well as a PEG smear of low molecular weight and an additive screen. PEGs were used in the range of 10-20% with the addition of 5%, 10%, 15% or 25% PEG.

Appendix 8.15: **ATPase-Zf-CW MORC4 protein BCS screen and optimisation crystallisation screens details.** Showing the screen details, protein concentrations, temperature information and protein purification details.

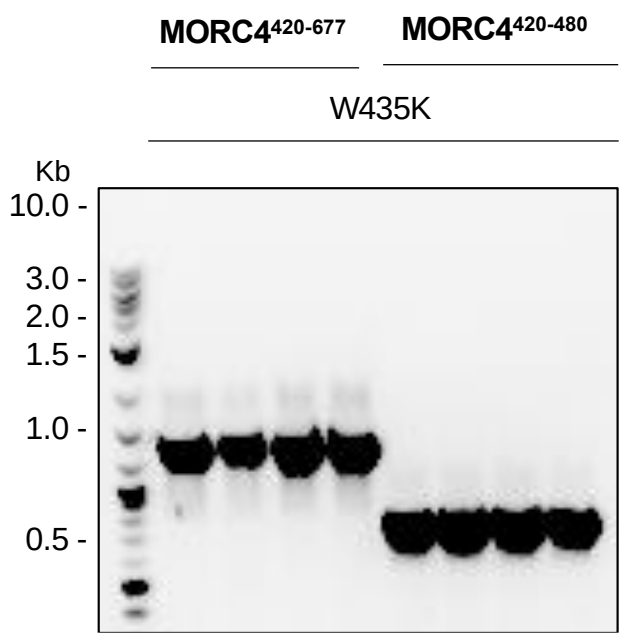
Domain	Protein[(aa)]	Temp (°C)	Screen	Concentration of protein (mg/ml)	Preparation of protein and ligand	Fresh or frozen	Protein purification	Final protein buffer
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	BCS screen	3.5	Dilute-concentrate-mix	Fresh	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	BCS_F02_FUP_1	4.0	Dilute-concentrate-mix	Fresh	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT
ATPase-Zf-CW	MORC4$^{29\text{-}480}$	20	BCS_F02_FUP_2	4.0	Dilute-concentrate-mix	Fresh	IMAC-TEV-IMAC-SEC	20 mM HEPES pH 7.5, 5% glycerol, 1 mM DTT



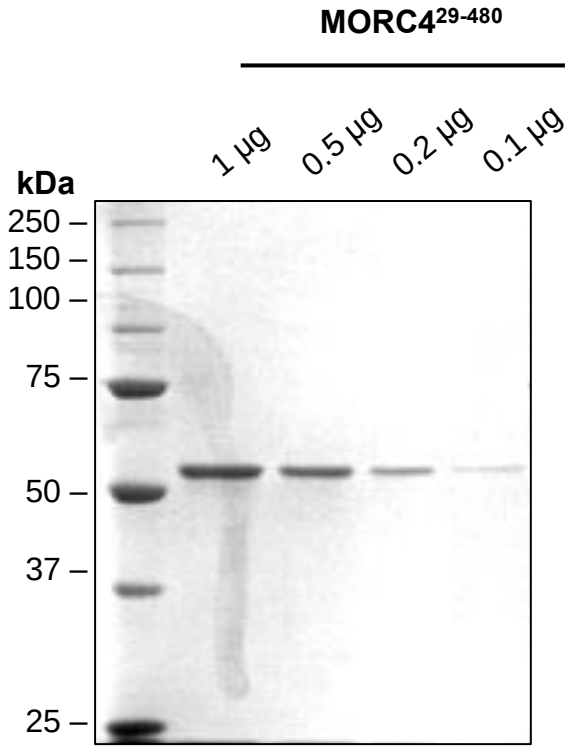
Appendix 8.16: **PCR screen of mutated truncated MORC4 fragments.** Following amplification of truncated fragments of MORC4 DNA with mutagenic primers, treatment of DNA, transformation in *E. coli* and PCR screening, samples were resolved on a 0.8% agarose gel. MORC4$^{29\text{-}480}$ samples were screened for N60A, K126R, D88A and W345K mutations. MORC4$^{420\text{-}757}$ samples were screened for the W345K mutation only. PCNA negative and positive were used as controls.



Appendix 8.17: **PCR screen of MORC4$^{29\text{-}480}$ N60A mutant.** Following amplification of truncated fragments of MORC4 DNA with mutagenic primers, treatment of DNA, transformation in *E. coli* and PCR screening, samples were resolved on a 0.8% agarose gel. MORC4$^{29\text{-}480}$ samples were screened for N60A.



Appendix 8.18: **PCR screen of MORC4$^{420-677}$ and MORC4$^{420-480}$ W435K.** Amplification of truncated fragments of MORC4 DNA with mutagenic primers, treatment of DNA, transformation in *E. coli* and PCR screening, samples were resolved on a 0.8% agarose gel. MORC4$^{420-677}$ and MORC4$^{420-480}$ samples were screened for W435K.



Appendix 8.19: **Determination of suitable concentrations of MORC4$^{29-480}$ required for subsequent chemical crosslinking.** 1 µg, 0.5 µg, 0.2 µg and 0.1 µg of MORC4$^{29-480}$ was loaded onto a 10% tris-glycine SDS (TGS)-PAGE gel.

Appendix 8.20: **Primer mapping of all known splice variants of MORC4.**
MORC4-202 ENST00000355610.9 5' and 3' open reading frame (ORF) regions are highlighted in yellow and blue, respectively.

```
  1 S  R  R  V  G  R  G  A  D  R  A  E  P  P  L  A  S  A  G  L
  1 AGTCGCCGGGTCGGTAGAGGTGCTGACAGGGCGGAGCCGCCGCTTGCCTCAGCCGGCCTC
  1          10        20        30        40        50
  1 TCAGCGGCCCAGCCATCTCCACGACTGTCCCGCCTCGGCGGCGAACGGAGTCGGCCGGAG
 21 T  P  P  A  R  D  R  W  L  E  R  A  P  P  P  P  A  A  S  G
 61 ACCCCTCCCGCCCGGGACAGGTGGCTGGAGCGCGCTCCGCCTCCGCCAGCTGCGAGTGGG
 61          70        80        90        100       110
 61 TGGGGAGGGCGGGCCCTGTCCACCGACCTCGCGCGAGGCGGAGGCGGTCGACGCTCACCC
 41 S  E  G  G  P  V  P  D  G  R  R  Q  L  R  V  E  V  S  E  G
121 AGCGAGGGAGGGCCGGTCCCGGACGGTCGCCGCCAGCTGCGAGTGGAAGTGAGCGAGGGC
121          130       140       150       160       170
121 TCGCTCCCTCCCGGCCAGGGCCTGCCAGCGGCGGTCGACGCTCACCTTCACTCGCTCCCG
 61 *  S  R  T  A  G  P  G  V  R  R  Q  Q  T  G  T  V  A  A  K
181 TAGTCCCGGACCGCCGGGCCAGGGGTCCGGCGGCAGCAGACGGGTACCGTGGCGGCCAAA
181          190       200       210       220       230
181 ATCAGGGCCTGGCGGCCCGGTCCCCAGGCCGCCGTCGTCTGCCCATGGCACCGCCGGTTT
 81 K  M  L  L  Y  R  G  A  P  A  G  P  G  A  P  G  C  G  L  A
241 AAAATGCTCCTGTACCGAGGGGCCCCCGCCGGCCCTGGCGCGCCGGGCTGCGGGCTGGCC
241          250       260       270       280       290
241 TTTTACGAGGACATGGCTCCCCGGGGGCGGCCGGGACCGCGCGGCCCGACGCCCGACCGG
                                >>>MORC4_f1>>> 325 to 342
101 R  P  G  G  G  P  Q  A  F  G  I  R  L  S  T  M  S  P  R  Y
301 CGGCCCGGCGGCGGCCCGCAGGCCTTCGGGATCCGCCTGAGCACGATGAGCCCCCGCTAC
301          310       320       330       340       350
301 GCCGGGCCGCCGCCGGGCGTCCGGAAGCCCTAGGCGGACTCGTGCTACTCGGGGGCGATG
121 L  Q  S  N  S  S  S  H  T  R  P  F  S  A  I  A  E  L  L  D
361 CTCCAGAGCAACTCCAGCAGCCACACGCGACCCTTCAGTGCCATCGCGGAGCTGCTAGAT
361          370       380       390       400       410
361 GAGGTCTCGTTGAGGTCGTCGGTGTGCGCTGGGAAGTCACGGTAGCGCCTCGACGATCTA
141 N  A  V  D  P  D  V  S  A  R  T  V  F  I  D  V  E  E  V  K
421 AATGCTGTAGATCCAGATGTATCTGCCAGGACGGTCTTTATAGATGTTGAGGAGGTCAAG
421          430       440       450       460       470
421 TTACGACATCTAGGTCTACATAGACGGTCCTGCCAGAAATATCTACAACTCCTCCAGTTC
161 N  K  S  C  L  T  F  T  D  D  G  C  G  M  T  P  H  K  L  H
481 AATAAATCTTGTTTGACCTTTACCGATGATGGATGTGGGATGACACCTCATAAACTACAC
481          490       500       510       520       530
481 TTATTTAGAACAAACTGGAAATGGCTACTACCTACACCCTACTGTGGAGTATTTGATGTG
181 R  M  L  S  F  G  F  T  D  K  V  I  K  K  S  Q  C  P  I  G
541 CGAATGCTCAGCTTTGGCTTTACAGATAAAGTAATAAAGAAGAGCCAGTGTCCCATTGGG
541          550       560       570       580       590
541 GCTTACGAGTCGAAACCGAAATGTCTATTTCATTATTTCTTCTCGGTCACAGGGTAACCC
201 V  F  G  N  G  F  K  S  G  S  M  R  L  G  K  D  A  L  V  F
601 GTCTTTGGTAATGGTTTCAAGTCAGGCTCCATGCGGCTAGGAAAGGACGCCCTTGTCTTC
601          610       620       630       640       650
601 CAGAAACCATTACCAAAGTTCAGTCCGAGGTACGCCGATCCTTTCCTGCGGGAACAGAAG
221 T  K  N  G  G  T  L  T  V  G  L  L  S  Q  T  Y  L  E  C  V
661 ACCAAGAATGGGGGTACTCTCACTGTTGGACTTCTATCACAGACCTATCTGGAATGTGTC
661          670       680       690       700       710
661 TGGTTCTTACCCCCATGAGAGTGACAACCTGAAGATAGTGTCTGGATAGACCTTACACAG
241 Q  A  Q  A  V  I  V  P  I  V  P  F  N  Q  Q  N  K  K  M  I
721 CAGGCCCAGGCAGTTATTGTACCAATTGTTCCATTCAACCAGCAAAACAAAAAAATGATT
721          730       740       750       760       770
721 GTCCGGGTCCGTCAATAACATGGTTAACAAGGTAAGTTGGTCGTTTTGTTTTTTTACTAA
261 I  T  E  D  S  L  P  S  L  E  A  I  L  N  Y  S  I  F  N  R
781 ATTACCGAGGATTCATTGCCCAGCCTAGAAGCCATCTTGAACTATTCCATTTTCAACCGT
781          790       800       810       820       830
```

```
 781 TAATGGCTCCTAAGTAACGGGTCGGATCTTCGGTAGAACTTGATAAGGTAAAAGTTGGCA
 281 E  N  D  L  L  A  Q  F  D  A  I  P  G  K  K  G  T  R  V  L
 841 GAAAATGACCTGCTGGCCCAGTTTGATGCCATCCCAGGCAAAAAAGGCACTCGTGTTCTC
 841          850       860       870       880       890
 841 CTTTTACTGGACGACCGGGTCAAACTACGGTAGGGTCCGTTTTTTCCGTGAGCACAAGAG
                                  <<<MORC4_r1<<< 928 to 945
 301 I  W  N  I  R  R  N  K  N  G  K  S  E  L  D  F  D  T  D  Q
 901 ATTTGGAACATCCGCAGAAATAAAAATGGAAAATCTGAGTTGGACTTTGATACAGATCAA
 901          910       920       930       940       950
 901 TAAACCTTGTAGGCGTCTTTATTTTTACCTTTTAGACTCAACCTGAAACTATGTCTAGTT
 321 Y  D  I  L  V  S  D  F  D  T  E  E  K  M  T  G  G  V  T  S
 961 TATGACATCCTGGTATCAGACTTTGACACAGAAGAAAAAATGACTGGCGGTGTTACCTCT
 961          970       980       990       1000      1010
 961 ATACTGTAGGACCATAGTCTGAAACTGTGTCTTCTTTTTTACTGACCGCCACAATGGAGA
 341 E  L  P  E  T  E  Y  S  L  R  A  F  C  G  I  L  Y  M  K  P
1021 GAGCTACCAGAAACAGAATATTCTTTAAGGGCATTTTGTGGTATTCTATACATGAAGCCA
1021          1030      1040      1050      1060      1070
1021 CTCGATGGTCTTTGTCTTATAAGAAATTCCCGTAAAACACCATAAGATATGTACTTCGGT
 361 R  M  K  I  F  L  R  Q  K  K  V  T  T  Q  M  I  A  K  S  L
1081 CGCATGAAAATTTTTCTGCGTCAAAAGAAGGTGACTACCCAGATGATTGCCAAGAGCCTG
1081          1090      1100      1110      1120      1130
1081 GCGTACTTTTAAAAAGACGCAGTTTTCTTCCACTGATGGGTCTACTAACGGTTCTCGGAC
 381 A  N  V  E  Y  D  T  Y  K  P  T  F  T  N  K  Q  V  R  I  T
1141 GCCAATGTAGAATATGATACATATAAACCTACCTTCACAAATAAGCAGGTGAGAATCACC
1141          1150      1160      1170      1180      1190
1141 CGGTTACATCTTATACTATGTATATTTGGATGGAAGTGTTTATTCGTCCACTCTTAGTGG
 401 F  G  F  S  C  K  N  S  N  Q  F  G  I  M  M  Y  H  N  N  R
1201 TTTGGGTTCTCTTGCAAGAATAGTAACCAGTTTGGAATAATGATGTATCATAACAACCGA
1201          1210      1220      1230      1240      1250
1201 AAACCCAAGAGAACGTTCTTATCATTGGTCAAACCTTATTACTACATAGTATTGTTGGCT
 421 L  I  K  S  F  E  K  V  G  C  Q  V  K  P  T  R  G  E  G  V
1261 CTCATAAAATCTTTTGAGAAGGTGGGGTGCCAGGTGAAGCCAACTCGTGGAGAAGGTGTA
1261          1270      1280      1290      1300      1310
1261 GAGTATTTTAGAAAACTCTTCCACCCCACGGTCCACTTCGGTTGAGCACCTCTTCCACAT
 441 G  V  I  G  V  I  E  C  N  F  L  K  P  A  Y  N  K  Q  D  F
1321 GGAGTAATTGGAGTCATTGAGTGCAATTTCCTAAAACCTGCCTACAACAAACAAGACTTT
1321          1330      1340      1350      1360      1370
1321 CCTCATTAACCTCAGTAACTCACGTTAAAGGATTTTGGACGGATGTTGTTTGTTCTGAAA
 461 E  Y  T  K  E  Y  R  L  T  I  N  A  L  A  Q  K  L  N  A  Y
1381 GAGTATACCAAGGAGTACCGGCTAACAATAAATGCCCTTGCCCAGAAGCTCAATGCTTAC
1381          1390      1400      1410      1420      1430
1381 CTCATATGGTTCCTCATGGCCGATTGTTATTTACGGGAACGGGTCTTCGAGTTACGAATG
 481 W  K  E  K  T  S  Q  D  N  F  E  T  S  T  V  A  R  P  I  P
1441 TGGAAGGAAAAAACATCTCAAGATAATTTTGAGACCTCAACTGTAGCCAGGCCAATACCG
1441          1450      1460      1470      1480      1490
1441 ACCTTCCTTTTTTGTAGAGTTCTATTAAAACTCTGGAGTTGACATCGGTCCGGTTATGGC
 501 K  V  P  D  Q  T  W  V  Q  C  D  E  C  L  K  W  R  K  L  P
1501 AAGGTTCCTGACCAGACATGGGTTCAGTGTGATGAGTGTCTTAAATGGAGAAAGCTTCCT
1501          1510      1520      1530      1540      1550
1501 TTCCAAGGACTGGTCTGTACCCAAGTCACACTACTCACAGAATTTACCTCTTTCGAAGGA
 521 G  K  I  D  P  S  M  L  P  A  R  W  F  C  Y  Y  N  S  H  P
1561 GGGAAGATTGATCCATCCATGTTACCTGCAAGATGGTTTTGTTATTATAATTCCCATCCA
1561          1570      1580      1590      1600      1610
1561 CCCTTCTAACTAGGTAGGTACAATGGACGTTCTACCAAAACAATAATATTAAGGGTAGGT
 541 K  Y  R  R  C  S  V  P  E  E  Q  E  L  T  D  E  D  L  C  L
1621 AAGTACAGGAGATGCTCTGTTCCAGAGGAACAAGAACTCACTGATGAAGACCTGTGCTTG
1621          1630      1640      1650      1660      1670
1621 TTCATGTCCTCTACGAGACAAGGTCTCCTTGTTCTTGAGTGACTACTTCTGGACACGAAC
 561 S  K  A  K  K  Q  E  Q  T  V  E  E  K  K  K  M  P  M  E  N
1681 AGCAAAGCTAAGAAACAAGAACAAACTGTTGAGGAGAAGAAGAAGATGCCTATGGAAAAT
1681          1690      1700      1710      1720      1730
```

```
1681 TCGTTTCGATTCTTTGTTCTTGTTTGACAACTCCTCTTCTTCTTCTACGGATACCTTTTA
 581 E  N  H  Q  V  F  S  N  P  P  K  I  L  T  V  Q  E  M  A  G
1741 GAGAACCACCAGGTATTCAGTAATCCACCAAAGATCCTTACTGTTCAAGAAATGGCTGGA
1741          1750      1760      1770      1780      1790
1741 CTCTTGGTGGTCCATAAGTCATTAGGTGGTTTCTAGGAATGACAAGTTCTTTACCGACCT
 601 L  N  N  K  T  I  G  Y  E  G  I  H  S  P  S  V  L  P  S  G
1801 TTGAATAACAAGACAATTGGATATGAGGGAATTCATAGCCCTAGTGTGCTTCCTTCTGGT
1801          1810      1820      1830      1840      1850
1801 AACTTATTGTTCTGTTAACCTATACTCCCTTAAGTATCGGGATCACACGAAGGAAGACCA
 621 G  E  E  S  R  S  P  S  L  Q  L  K  P  L  D  S  S  V  L  Q
1861 GGAGAAGAAAGCAGATCACCATCTCTTCAACTTAAGCCTCTGGATTCCAGTGTTTTACAG
1861          1870      1880      1890      1900      1910
1861 CCTCTTCTTTCGTCTAGTGGTAGAGAAGTTGAATTCGGAGACCTAAGGTCACAAAATGTC
 641 F  S  S  K  Y  K  W  I  L  G  E  E  P  V  E  K  R  R  R  L
1921 TTTTCCAGTAAGTACAAATGGATCCTAGGTGAAGAACCGGTGGAGAAACGAAGAAGGCTC
1921          1930      1940      1950      1960      1970
1921 AAAAGGTCATTCATGTTTACCTAGGATCCACTTCTTGGCCACCTCTTTGCTTCTTCCGAG
 661 Q  N  E  M  T  T  P  S  L  D  Y  S  M  P  A  P  Y  R  R  V
1981 CAGAATGAGATGACAACACCTTCTCTAGATTATTCCATGCCTGCTCCTTACAGGAGGGTA
1981          1990      2000      2010      2020      2030
1981 GTCTTACTCTACTGTTGTGGAAGAGATCTAATAAGGTACGGACGAGGAATGTCCTCCCAT
 681 E  A  P  V  A  Y  P  E  G  E  N  S  H  D  K  S  S  S  E  R
2041 GAAGCACCTGTTGCCTACCCAGAAGGGGAGAACAGCCATGATAAATCGAGTTCTGAGAGA
2041          2050      2060      2070      2080      2090
2041 CTTCGTGGACAACGGATGGGTCTTCCCCTCTTGTCGGTACTATTTAGCTCAAGACTCTCT
 701 S  T  P  P  Y  L  F  P  E  Y  P  E  A  S  K  N  T  G  Q  N
2101 AGTACACCACCATACCTTTTCCCAGAATACCCAGAAGCAAGCAAGAATACAGGTCAGAAT
2101          2110      2120      2130      2140      2150
2101 TCATGTGGTGGTATGGAAAAGGGTCTTATGGGTCTTCGTTCGTTCTTATGTCCAGTCTTA
 721 R  E  V  S  I  L  Y  P  G  A  K  D  Q  R  Q  G  S  L  L  P
2161 AGGGAGGTTTCAATTCTGTATCCAGGGGCCAAAGACCAACGCCAGGGGTCCCTGCTTCCT
2161          2170      2180      2190      2200      2210
2161 TCCCTCCAAAGTTAAGACATAGGTCCCCGGTTTCTGGTTGCGGTCCCCAGGGACGAAGGA
 741 E  E  L  E  D  Q  M  P  R  L  V  A  E  E  S  N  R  G  S  T
2221 GAAGAATTAGAAGATCAGATGCCAAGATTGGTGGCAGAAGAATCTAACAGAGGTAGCACA
2221          2230      2240      2250      2260      2270
2221 CTTCTTAATCTTCTAGTCTACGGTTCTAACCACCGTCTTCTTAGATTGTCTCCATCGTGT
 761 T  I  N  K  E  E  V  N  K  G  P  F  V  A  V  V  G  V  A  K
2281 ACCATAAACAAAGAAGAAGTCAACAAGGGACCTTTTGTAGCTGTTGTGGGTGTTGCCAAA
2281          2290      2300      2310      2320      2330
2281 TGGTATTTGTTTCTTCTTCAGTTGTTCCCTGGAAAACATCGACAACACCCACAACGGTTT
           >>>MORC4_f2>>> 2346 to 2363
 781 G  V  R  D  S  G  A  P  I  Q  L  I  P  F  N  R  E  E  L  A
2341 GGTGTTAGAGATTCAGGAGCTCCCATTCAGCTGATCCCTTTTAACAGAGAGGAGCTTGCT
2341          2350      2360      2370      2380      2390
2341 CCACAATCTCTAAGTCCTCGAGGGTAAGTCGACTAGGGAAAATTGTCTCTCCTCGAACGA
 801 E  R  R  K  A  V  E  S  W  N  P  V  P  Y  S  V  A  S  A  A
2401 GAGAGACGAAAAGCAGTTGAATCCTGGAACCCAGTGCCTTATTCTGTGGCCTCTGCTGCA
2401          2410      2420      2430      2440      2450
2401 CTCTCTGCTTTTCGTCAACTTAGGACCTTGGGTCACGGAATAAGACACCGGAGACGACGT
 821 I  P  A  A  A  I  G  E  K  A  R  G  Y  E  E  S  E  G  H  N
2461 ATCCCTGCTGCAGCCATTGGGGAGAAAGCAAGAGGCTATGAGGAGAGCGAAGGTCATAAT
2461          2470      2480      2490      2500      2510
2461 TAGGGACGACGTCGGTAACCCCTCTTTCGTTCTCCGATACTCCTCTCGCTTCCAGTATTA
 841 T  P  K  L  K  N  Q  R  E  L  E  E  L  K  R  T  T  E  K  L
2521 ACACCAAAGTTGAAGAACCAGAGAGAGCTGGAAGAATTGAAGAGAACCACAGAAAAATTG
2521          2530      2540      2550      2560      2570
2521 TGTGGTTTCAACTTCTTGGTCTCTCTCGACCTTCTTAACTTCTCTTGGTGTCTTTTTAAC
 861 E  R  V  L  A  E  R  N  L  F  Q  Q  K  V  E  E  L  E  Q  E
2581 GAACGTGTTTTGGCTGAAAGGAATTTGTTCCAGCAAAAGGTGGAGGAGCTGGAACAGGAG
2581          2590      2600      2610      2620      2630
```

```
2581 CTTGCACAAAACCGACTTTCCTTAAACAAGGTCGTTTTCCACCTCCTCGACCTTGTCCTC
 881 R  N  H  W  Q  S  E  F  K  K  V  Q  H  E  L  V  I  Y  S  T
2641 AGGAATCACTGGCAGTCTGAATTCAAGAAAGTCCAACATGAATTGGTGATCTACAGTACC
2641          2650      2660      2670      2680      2690
2641 TCCTTAGTGACCGTCAGACTTAAGTTCTTTCAGGTTGTACTTAACCACTAGATGTCATGG
 901 Q  E  A  E  G  L  Y  W  S  K  K  H  M  G  Y  R  Q  A  E  F
2701 CAGGAGGCGGAAGGCTTGTACTGGAGCAAGAAACACATGGGTTATCGCCAAGCTGAATTC
2701          2710      2720      2730      2740      2750
2701 GTCCTCCGCCTTCCGAACATGACCTCGTTCTTTGTGTACCCAATAGCGGTTCGACTTAAG
 921 Q  I  L  K  A  E  L  E  R  T  K  E  E  K  Q  E  L  K  E  K
2761 CAGATTCTGAAAGCTGAGCTGGAAAGAACCAAAGAGGAAAAGCAAGAGTTAAAAGAGAAA
2761          2770      2780      2790      2800      2810
2761 GTCTAAGACTTTCGACTCGACCTTTCTTGGTTTCTCCTTTTCGTTCTCAATTTTCTCTTT
           <<<MORC4_r4<<< 2826 to 2843
 941 L  K  E  T  E  T  H  L  E  M  L  Q  K  A  Q  V  S  Y  R  T
2821 CTGAAGGAAACAGAGACACACCTGGAAATGCTGCAGAAGGCTCAGGTCTCCTACCGGACC
2821          2830      2840      2850      2860      2870
2821 GACTTCCTTTGTCTCTGTGTGGACCTTTACGACGTCTTCCGAGTCCAGAGGATGGCCTGG
 961 P  E  G  D  D  L  E  R  A  L  A  K  L  T  R  L  R  I  H  V
2881 CCAGAGGGAGATGACCTAGAAAGGGCTTTGGCAAAGCTTACGCGGCTACGTATCCACGTC
2881          2890      2900      2910      2920      2930
2881 GGTCTCCCTCTACTGGATCTTTCCCGAAACCGTTTCGAATGCGCCGATGCATAGGTGCAG
                        <<<MORC4_r3<<< 2964 to 2981
 981 S  Y  L  L  T  S  V  L  P  H  L  E  L  R  E  I  G  Y  D  S
2941 AGCTATCTCCTTACTTCTGTCCTCCCTCACTTGGAGCTTCGTGAGATCGGGTATGACTCA
2941          2950      2960      2970      2980      2990
2941 TCGATAGAGGAATGAAGACAGGAGGGAGTGAACCTCGAAGCACTCTAGCCCATACTGAGT
                                              <<<MORC4_r2<<< 3037 to 3054
1001 E  Q  V  D  G  I  L  Y  T  V  L  E  A  N  H  I  L  D  *
```

MORC4-201 ENST00000255495.7. 5’ and 3’ open reading frame (ORF) regions are highlighted in yellow and blue, respectively.

```
   1 M  L  L  Y  R  G  A  P  A  G  P  G  A  P  G  C  G  L  A  R
   1 ATGCTCCTGTACCGAGGGGCCCCCGCCGGCCCTGGCGCGCCGGGCTGCGGGCTGGCCCGG
   1          10        20        30        40        50
   1 TACGAGGACATGGCTCCCCGGGGGCGGCCGGGACCGCGCGGCCCGACGCCCGACCGGGCC
                        >>>MORC4_f1>>> 82 to 99
  21 P  G  G  G  P  Q  A  F  G  I  R  L  S  T  M  S  P  R  Y  L
  61 CCCGGCGGCGGCCCGCAGGCCTTCGGGATCCGCCTGAGCACGATGAGCCCCCGCTACCTC
  61          70        80        90        100       110
  61 GGGCCGCCGCCGGGCGTCCGGAAGCCCTAGGCGGACTCGTGCTACTCGGGGGCGATGGAG
  41 Q  S  N  S  S  S  H  T  R  P  F  S  A  I  A  E  L  L  D  N
 121 CAGAGCAACTCCAGCAGCCACACGCGACCCTTCAGTGCCATCGCGGAGCTGCTAGATAAT
 121          130       140       150       160       170
 121 GTCTCGTTGAGGTCGTCGGTGTGCGCTGGGAAGTCACGGTAGCGCCTCGACGATCTATTA
  61 A  V  D  P  D  V  S  A  R  T  V  F  I  D  V  E  E  V  K  N
 181 GCTGTAGATCCAGATGTATCTGCCAGGACGGTCTTTATAGATGTTGAGGAGGTCAAGAAT
 181          190       200       210       220       230
 181 CGACATCTAGGTCTACATAGACGGTCCTGCCAGAAATATCTACAACTCCTCCAGTTCTTA
  81 K  S  C  L  T  F  T  D  D  G  C  G  M  T  P  H  K  L  H  R
 241 AAATCTTGTTTGACCTTTACCGATGATGGATGTGGGATGACACCTCATAAACTACACCGA
 241          250       260       270       280       290
 241 TTTAGAACAAACTGGAAATGGCTACTACCTACACCCTACTGTGGAGTATTTGATGTGGCT
 101 M  L  S  F  G  F  T  D  K  V  I  K  K  S  Q  C  P  I  G  V
 301 ATGCTCAGCTTTGGCTTTACAGATAAAGTAATAAAGAAGAGCCAGTGTCCCATTGGGGTC
 301          310       320       330       340       350
 301 TACGAGTCGAAACCGAAATGTCTATTTCATTATTTCTTCTCGGTCACAGGGTAACCCCAG
 121 F  G  N  G  F  K  S  G  S  M  R  L  G  K  D  A  L  V  F  T
 361 TTTGGTAATGGTTTCAAGTCAGGCTCCATGCGGCTAGGAAAGGACGCCCTTGTCTTCACC
 361          370       380       390       400       410
```

```
 361 AAACCATTACCAAAGTTCAGTCCGAGGTACGCCGATCCTTTCCTGCGGGAACAGAAGTGG
 141 K  N  G  G  T  L  T  V  G  L  L  S  Q  T  Y  L  E  C  V  Q
 421 AAGAATGGGGGTACTCTCACTGTTGGACTTCTATCACAGACCTATCTGGAATGTGTCCAG
 421          430       440       450       460       470
 421 TTCTTACCCCCATGAGAGTGACAACCTGAAGATAGTGTCTGGATAGACCTTACACAGGTC
 161 A  Q  A  V  I  V  P  I  V  P  F  N  Q  Q  N  K  K  M  I  I
 481 GCCCAGGCAGTTATTGTACCAATTGTTCCATTCAACCAGCAAAACAAAAAAATGATTATT
 481          490       500       510       520       530
 481 CGGGTCCGTCAATAACATGGTTAACAAGGTAAGTTGGTCGTTTTGTTTTTTTACTAATAA
 181 T  E  D  S  L  P  S  L  E  A  I  L  N  Y  S  I  F  N  R  E
 541 ACCGAGGATTCATTGCCCAGCCTAGAAGCCATCTTGAACTATTCCATTTTCAACCGTGAA
 541          550       560       570       580       590
 541 TGGCTCCTAAGTAACGGGTCGGATCTTCGGTAGAACTTGATAAGGTAAAAGTTGGCACTT
 201 N  D  L  L  A  Q  F  D  A  I  P  G  K  K  G  T  R  V  L  I
 601 AATGACCTGCTGGCCCAGTTTGATGCCATCCCAGGCAAAAAAGGCACTCGTGTTCTCATT
 601          610       620       630       640       650
 601 TTACTGGACGACCGGGTCAAACTACGGTAGGGTCCGTTTTTTCCGTGAGCACAAGAGTAA
                                <<<MORC4_r1<<< 685 to 702
 221 W  N  I  R  R  N  K  N  G  K  S  E  L  D  F  D  T  D  Q  Y
 661 TGGAACATCCGCAGAAATAAAAATGGAAAATCTGAGTTGGACTTTGATACAGATCAATAT
 661          670       680       690       700       710
 661 ACCTTGTAGGCGTCTTTATTTTTACCTTTTAGACTCAACCTGAAACTATGTCTAGTTATA
 241 D  I  L  V  S  D  F  D  T  E  E  K  M  T  G  G  V  T  S  E
 721 GACATCCTGGTATCAGACTTTGACACAGAAGAAAAAATGACTGGCGGTGTTACCTCTGAG
 721          730       740       750       760       770
 721 CTGTAGGACCATAGTCTGAAACTGTGTCTTCTTTTTTACTGACCGCCACAATGGAGACTC
 261 L  P  E  T  E  Y  S  L  R  A  F  C  G  I  L  Y  M  K  P  R
 781 CTACCAGAAACAGAATATTCTTTAAGGGCATTTTGTGGTATTCTATACATGAAGCCACGC
 781          790       800       810       820       830
 781 GATGGTCTTTGTCTTATAAGAAATTCCCGTAAAACACCATAAGATATGTACTTCGGTGCG
 281 M  K  I  F  L  R  Q  K  K  V  T  T  Q  M  I  A  K  S  L  A
 841 ATGAAAATTTTTCTGCGTCAAAAGAAGGTGACTACCCAGATGATTGCCAAGAGCCTGGCC
 841          850       860       870       880       890
 841 TACTTTTAAAAAGACGCAGTTTTCTTCCACTGATGGGTCTACTAACGGTTCTCGGACCGG
 301 N  V  E  Y  D  T  Y  K  P  T  F  T  N  K  Q  V  R  I  T  F
 901 AATGTAGAATATGATACATATAAACCTACCTTCACAAATAAGCAGGTGAGAATCACCTTT
 901          910       920       930       940       950
 901 TTACATCTTATACTATGTATATTTGGATGGAAGTGTTTATTCGTCCACTCTTAGTGGAAA
 321 G  F  S  C  K  N  S  N  Q  F  G  I  M  M  Y  H  N  N  R  L
 961 GGGTTCTCTTGCAAGAATAGTAACCAGTTTGGAATAATGATGTATCATAACAACCGACTC
 961          970       980       990       1000      1010
 961 CCCAAGAGAACGTTCTTATCATTGGTCAAACCTTATTACTACATAGTATTGTTGGCTGAG
 341 I  K  S  F  E  K  V  G  C  Q  V  K  P  T  R  G  E  G  V  G
1021 ATAAAATCTTTTGAGAAGGTGGGGTGCCAGGTGAAGCCAACTCGTGGAGAAGGTGTAGGA
1021          1030      1040      1050      1060      1070
1021 TATTTTAGAAAACTCTTCCACCCCACGGTCCACTTCGGTTGAGCACCTCTTCCACATCCT
 361 V  I  G  V  I  E  C  N  F  L  K  P  A  Y  N  K  Q  D  F  E
1081 GTAATTGGAGTCATTGAGTGCAATTTCCTAAAACCTGCCTACAACAAACAAGACTTTGAG
1081          1090      1100      1110      1120      1130
1081 CATTAACCTCAGTAACTCACGTTAAAGGATTTTGGACGGATGTTGTTTGTTCTGAAACTC
 381 Y  T  K  E  Y  R  L  T  I  N  A  L  A  Q  K  L  N  A  Y  W
1141 TATACCAAGGAGTACCGGCTAACAATAAATGCCCTTGCCCAGAAGCTCAATGCTTACTGG
1141          1150      1160      1170      1180      1190
1141 ATATGGTTCCTCATGGCCGATTGTTATTTACGGGAACGGGTCTTCGAGTTACGAATGACC
 401 K  E  K  T  S  Q  D  N  F  E  T  S  T  V  A  R  P  I  P  K
1201 AAGGAAAAAACATCTCAAGATAATTTTGAGACCTCAACTGTAGCCAGGCCAATACCGAAG
1201          1210      1220      1230      1240      1250
1201 TTCCTTTTTTGTAGAGTTCTATTAAAACTCTGGAGTTGACATCGGTCCGGTTATGGCTTC
 421 V  P  D  Q  T  W  V  Q  C  D  E  C  L  K  W  R  K  L  P  G
1261 GTTCCTGACCAGACATGGGTTCAGTGTGATGAGTGTCTTAAATGGAGAAAGCTTCCTGGG
1261          1270      1280      1290      1300      1310
```

```
1261 CAAGGACTGGTCTGTACCCAAGTCACACTACTCACAGAATTTACCTCTTTCGAAGGACCC
 441 K  I  D  P  S  M  L  P  A  R  W  F  C  Y  Y  N  S  H  P  K
1321 AAGATTGATCCATCCATGTTACCTGCAAGATGGTTTTGTTATTATAATTCCCATCCAAAG
1321          1330      1340      1350      1360      1370
1321 TTCTAACTAGGTAGGTACAATGGACGTTCTACCAAAACAATAATATTAAGGGTAGGTTTC
 461 Y  R  R  C  S  V  P  E  E  Q  E  L  T  D  E  D  L  C  L  S
1381 TACAGGAGATGCTCTGTTCCAGAGGAACAAGAACTCACTGATGAAGACCTGTGCTTGAGC
1381          1390      1400      1410      1420      1430
1381 ATGTCCTCTACGAGACAAGGTCTCCTTGTTCTTGAGTGACTACTTCTGGACACGAACTCG
 481 K  A  K  K  Q  E  Q  T  V  E  E  K  K  K  M  P  M  E  N  E
1441 AAAGCTAAGAAACAAGAACAAACTGTTGAGGAGAAGAAGAAGATGCCTATGGAAAATGAG
1441          1450      1460      1470      1480      1490
1441 TTTCGATTCTTTGTTCTTGTTTGACAACTCCTCTTCTTCTTCTACGGATACCTTTTACTC
 501 N  H  Q  V  F  S  N  P  P  K  I  L  T  V  Q  E  M  A  G  L
1501 AACCACCAGGTATTCAGTAATCCACCAAAGATCCTTACTGTTCAAGAAATGGCTGGATTG
1501          1510      1520      1530      1540      1550
1501 TTGGTGGTCCATAAGTCATTAGGTGGTTTCTAGGAATGACAAGTTCTTTACCGACCTAAC
 521 N  N  K  T  I  G  Y  E  G  I  H  S  P  S  V  L  P  S  G  G
1561 AATAACAAGACAATTGGATATGAGGGAATTCATAGCCCTAGTGTGCTTCCTTCTGGTGGA
1561          1570      1580      1590      1600      1610
1561 TTATTGTTCTGTTAACCTATACTCCCTTAAGTATCGGGATCACACGAAGGAAGACCACCT
 541 E  E  S  R  S  P  S  L  Q  L  K  P  L  D  S  S  V  L  Q  F
1621 GAAGAAAGCAGATCACCATCTCTTCAACTTAAGCCTCTGGATTCCAGTGTTTTACAGTTT
1621          1630      1640      1650      1660      1670
1621 CTTCTTTCGTCTAGTGGTAGAGAAGTTGAATTCGGAGACCTAAGGTCACAAAATGTCAAA
 561 S  S  K  Y  K  W  I  L  G  E  E  P  V  E  K  R  R  R  L  Q
1681 TCCAGTAAGTACAAATGGATCCTAGGTGAAGAACCGGTGGAGAAACGAAGAAGGCTCCAG
1681          1690      1700      1710      1720      1730
1681 AGGTCATTCATGTTTACCTAGGATCCACTTCTTGGCCACCTCTTTGCTTCTTCCGAGGTC
 581 N  E  M  T  T  P  S  L  D  Y  S  M  P  A  P  Y  R  R  V  E
1741 AATGAGATGACAACACCTTCTCTAGATTATTCCATGCCTGCTCCTTACAGGAGGGTAGAA
1741          1750      1760      1770      1780      1790
1741 TTACTCTACTGTTGTGGAAGAGATCTAATAAGGTACGGACGAGGAATGTCCTCCCATCTT
 601 A  P  V  A  Y  P  E  G  E  N  S  H  D  K  S  S  S  E  R  S
1801 GCACCTGTTGCCTACCCAGAAGGGGAGAACAGCCATGATAAATCGAGTTCTGAGAGAAGT
1801          1810      1820      1830      1840      1850
1801 CGTGGACAACGGATGGGTCTTCCCCTCTTGTCGGTACTATTTAGCTCAAGACTCTCTTCA
 621 T  P  P  Y  L  F  P  E  Y  P  E  A  S  K  N  T  G  Q  N  R
1861 ACACCACCATACCTTTTCCCAGAATACCCAGAAGCAAGCAAGAATACAGGTCAGAATAGG
1861          1870      1880      1890      1900      1910
1861 TGTGGTGGTATGGAAAAGGGTCTTATGGGTCTTCGTTCGTTCTTATGTCCAGTCTTATCC
 641 E  V  S  I  L  Y  P  G  A  K  D  Q  R  Q  G  S  L  L  P  E
1921 GAGGTTTCAATTCTGTATCCAGGGGCCAAAGACCAACGCCAGGGGTCCCTGCTTCCTGAA
1921          1930      1940      1950      1960      1970
1921 CTCCAAAGTTAAGACATAGGTCCCCGGTTTCTGGTTGCGGTCCCCAGGGACGAAGGACTT
 661 E  L  E  D  Q  M  P  R  L  V  A  E  E  S  N  R  G  S  T  T
1981 GAATTAGAAGATCAGATGCCAAGATTGGTGGCAGAAGAATCTAACAGAGGTAGCACAACC
1981          1990      2000      2010      2020      2030
1981 CTTAATCTTCTAGTCTACGGTTCTAACCACCGTCTTCTTAGATTGTCTCCATCGTGTTGG
 681 I  N  K  E  E  V  N  K  G  P  F  V  A  V  V  G  V  A  K  G
2041 ATAAACAAAGAAGAAGTCAACAAGGGACCTTTTGTAGCTGTTGTGGGTGTTGCCAAAGGT
2041          2050      2060      2070      2080      2090
2041 TATTTGTTTCTTCTTCAGTTGTTCCCTGGAAAACATCGACAACACCCACAACGGTTTCCA
       >>>MORC4_f2>>> 2103 to 2120
 701 V  R  D  S  G  A  P  I  Q  L  I  P  F  N  R  E  E  L  A  E
2101 GTTAGAGATTCAGGAGCTCCCATTCAGCTGATCCCTTTTAACAGAGAGGAGCTTGCTGAG
2101          2110      2120      2130      2140      2150
2101 CAATCTCTAAGTCCTCGAGGGTAAGTCGACTAGGGAAAATTGTCTCTCCTCGAACGACTC
 721 R  R  K  A  V  E  S  W  N  P  V  P  Y  S  V  A  S  A  A  I
2161 AGACGAAAAGCAGTTGAATCCTGGAACCCAGTGCCTTATTCTGTGGCCTCTGCTGCAATC
2161          2170      2180      2190      2200      2210
```

```
2161 TCTGCTTTTCGTCAACTTAGGACCTTGGGTCACGGAATAAGACACCGGAGACGACGTTAG
 741 P  A  A  A  I  G  E  K  A  R  G  Y  E  E  S  E  G  H  N  T
2221 CCTGCTGCAGCCATTGGGGAGAAAGCAAGAGGCTATGAGGAGAGCGAAGGTCATAATACA
2221          2230      2240      2250      2260      2270
2221 GGACGACGTCGGTAACCCCTCTTTCGTTCTCCGATACTCCTCTCGCTTCCAGTATTATGT
 761 P  K  L  K  N  Q  R  E  L  E  E  L  K  R  T  T  E  K  L  E
2281 CCAAAGTTGAAGAACCAGAGAGAGCTGGAAGAATTGAAGAGAACCACAGAAAAATTGGAA
2281          2290      2300      2310      2320      2330
2281 GGTTTCAACTTCTTGGTCTCTCTCGACCTTCTTAACTTCTCTTGGTGTCTTTTTAACCTT
 781 R  V  L  A  E  R  N  L  F  Q  Q  K  V  E  E  L  E  Q  E  R
2341 CGTGTTTTGGCTGAAAGGAATTTGTTCCAGCAAAAGGTGGAGGAGCTGGAACAGGAGAGG
2341          2350      2360      2370      2380      2390
2341 GCACAAAACCGACTTTCCTTAAACAAGGTCGTTTTCCACCTCCTCGACCTTGTCCTCTCC
 801 N  H  W  Q  S  E  F  K  K  V  Q  H  E  L  V  I  Y  S  T  Q
2401 AATCACTGGCAGTCTGAATTCAAGAAAGTCCAACATGAATTGGTGATCTACAGTACCCAG
2401          2410      2420      2430      2440      2450
2401 TTAGTGACCGTCAGACTTAAGTTCTTTCAGGTTGTACTTAACCACTAGATGTCATGGGTC
 821 E  A  E  G  L  Y  W  S  K  K  H  M  G  Y  R  Q  A  E  F  Q
2461 GAGGCGGAAGGCTTGTACTGGAGCAAGAAACACATGGGTTATCGCCAAGCTGAATTCCAG
2461          2470      2480      2490      2500      2510
2461 CTCCGCCTTCCGAACATGACCTCGTTCTTTGTGTACCCAATAGCGGTTCGACTTAAGGTC
 841 I  L  K  A  E  L  E  R  T  K  E  E  K  Q  E  L  K  E  K  L
2521 ATTCTGAAAGCTGAGCTGGAAAGAACCAAAGAGGAAAAGCAAGAGTTAAAAGAGAAACTG
2521          2530      2540      2550      2560      2570
2521 TAAGACTTTCGACTCGACCTTTCTTGGTTTCTCCTTTTCGTTCTCAATTTTCTCTTTGAC
       <<<MORC4_r4<<< 2583 to 2600
 861 K  E  T  E  T  H  L  E  M  L  Q  K  A  Q  G  F  G  K  A  Y
2581 AAGGAAACAGAGACACACCTGGAAATGCTGCAGAAGGCTCAGGGCTTTGGCAAAGCTTAC
2581          2590      2600      2610      2620      2630
2581 TTCCTTTGTCTCTGTGTGGACCTTTACGACGTCTTCCGAGTCCCGAAACCGTTTCGAATG
                                                  <<<MORC4_r3<<< 2683 to
2700
 881 A  A  T  Y  P  R  Q  L  S  P  Y  F  C  P  P  S  L  G  A  S
2641 GCGGCTACGTATCCACGTCAGCTATCTCCTTACTTCTGTCCTCCCTCACTTGGAGCTTCG
2641          2650      2660      2670      2680      2690
2641 CGCCGATGCATAGGTGCAGTCGATAGAGGAATGAAGACAGGAGGGAGTGAACCTCGAAGC
 901 *
2701 TGA
```

MORC4-204 ENST00000604604.1

```
   1 X  S  A  R  T  V  F  I  D  V  E  E  V  K  N  K  S  C  L  T
   1 NNATCTGCCAGGACGGTCTTTATAGATGTTGAGGAGGTCAAGAATAAATCTTGTTTGACC
   1          10        20        30        40        50
   1 NNTAGACGGTCCTGCCAGAAATATCTACAACTCCTCCAGTTCTTATTTAGAACAAACTGG
  21 F  T  D  D  G  C  G  M  T  P  H  K  L  H  R  M  L  S  *
  61 TTTACCGATGATGGATGTGGGATGACACCTCATAAACTACACCGAATGCTCAGTTAG
  61          70        80        90        100       110
  61 AAATGGCTACTACCTACACCCTACTGTGGAGTATTTGATGTGGCTTACGAGTCAATC
```

MORC4 **Q8TE76-3**

```
   1 M  T  G  G  V  T  S  E  L  P  E  T  E  Y  S  L  R  A  F  C
   1 ATGACTGGCGGTGTTACCTCTGAGCTACCAGAAACAGAATATTCTTTAAGGGCATTTTGT
   1          10        20        30        40        50
   1 TACTGACCGCCACAATGGAGACTCGATGGTCTTTGTCTTATAAGAAATTCCCGTAAAACA
  21 G  I  L  Y  M  K  P  R  M  K  I  F  L  R  Q  K  K  V  T  T
  61 GGTATTCTATACATGAAGCCACGCATGAAAATTTTTCTGCGTCAAAAGAAGGTGACTACC
  61          70        80        90        100       110
  61 CCATAAGATATGTACTTCGGTGCGTACTTTTAAAAAGACGCAGTTTTCTTCCACTGATGG
  41 Q  M  I  A  K  S  L  A  N  V  E  Y  D  T  Y  K  P  T  F  T
```

```
121 CAGATGATTGCCAAGAGCCTGGCCAATGTAGAATATGATACATATAAACCTACCTTCACA
121          130       140       150       160       170
121 GTCTACTAACGGTTCTCGGACCGGTTACATCTTATACTATGTATATTTGGATGGAAGTGT
 61 N  K  Q  V  R  I  T  F  G  F  S  C  K  N  S  N  Q  F  G  I
181 AATAAGCAGGTGAGAATCACCTTTGGGTTCTCTTGCAAGAATAGTAACCAGTTTGGAATA
181          190       200       210       220       230
181 TTATTCGTCCACTCTTAGTGGAAACCCAAGAGAACGTTCTTATCATTGGTCAAACCTTAT
 81 M  M  Y  H  N  N  R  L  I  K  S  F  E  K  V  G  C  Q  V  K
241 ATGATGTATCATAACAACCGACTCATAAAATCTTTTGAGAAGGTGGGGTGCCAGGTGAAG
241          250       260       270       280       290
241 TACTACATAGTATTGTTGGCTGAGTATTTTAGAAAACTCTTCCACCCCACGGTCCACTTC
101 P  T  R  G  E  G  V  G  V  I  G  V  I  E  C  N  F  L  K  P
301 CCAACTCGTGGAGAAGGTGTAGGAGTAATTGGAGTCATTGAGTGCAATTTCCTAAAACCT
301          310       320       330       340       350
301 GGTTGAGCACCTCTTCCACATCCTCATTAACCTCAGTAACTCACGTTAAAGGATTTTGGA
121 A  Y  N  K  Q  D  F  E  Y  T  K  E  Y  R  L  T  I  N  A  L
361 GCCTACAACAAACAAGACTTTGAGTATACCAAGGAGTACCGGCTAACAATAAATGCCCTT
361          370       380       390       400       410
361 CGGATGTTGTTTGTTCTGAAACTCATATGGTTCCTCATGGCCGATTGTTATTTACGGGAA
141 A  Q  K  L  N  A  Y  W  K  E  K  T  S  Q  D  N  F  E  T  S
421 GCCCAGAAGCTCAATGCTTACTGGAAGGAAAAAACATCTCAAGATAATTTTGAGACCTCA
421          430       440       450       460       470
421 CGGGTCTTCGAGTTACGAATGACCTTCCTTTTTTGTAGAGTTCTATTAAAACTCTGGAGT
161 T  V  A  R  P  I  P  K  V  P  D  Q  T  W  V  Q  C  D  E  C
481 ACTGTAGCCAGGCCAATACCGAAGGTTCCTGACCAGACATGGGTTCAGTGTGATGAGTGT
481          490       500       510       520       530
481 TGACATCGGTCCGGTTATGGCTTCCAAGGACTGGTCTGTACCCAAGTCACACTACTCACA
181 L  K  W  R  K  L  P  G  K  I  D  P  S  M  L  P  A  R  W  F
541 CTTAAATGGAGAAAGCTTCCTGGGAAGATTGATCCATCCATGTTACCTGCAAGATGGTTT
541          550       560       570       580       590
541 GAATTTACCTCTTTCGAAGGACCCTTCTAACTAGGTAGGTACAATGGACGTTCTACCAAA
201 C  Y  Y  N  S  H  P  K  Y  R  R  C  S  V  P  E  E  Q  E  L
601 TGTTATTATAATTCCCATCCAAAGTACAGGAGATGCTCTGTTCCAGAGGAACAAGAACTC
601          610       620       630       640       650
601 ACAATAATATTAAGGGTAGGTTTCATGTCCTCTACGAGACAAGGTCTCCTTGTTCTTGAG
221 T  D  E  D  L  C  L  S  K  A  K  K  Q  E  Q  T  V  E  E  K
661 ACTGATGAAGACCTGTGCTTGAGCAAAGCTAAGAAACAAGAACAAACTGTTGAGGAGAAG
661          670       680       690       700       710
661 TGACTACTTCTGGACACGAACTCGTTTCGATTCTTTGTTCTTGTTTGACAACTCCTCTTC
241 K  K  M  P  M  E  N  E  N  H  Q  V  F  S  N  P  P  K  I  L
721 AAGAAGATGCCTATGGAAAATGAGAACCACCAGGTATTCAGTAATCCACCAAAGATCCTT
721          730       740       750       760       770
721 TTCTTCTACGGATACCTTTTACTCTTGGTGGTCCATAAGTCATTAGGTGGTTTCTAGGAA
261 T  V  Q  E  M  A  G  L  N  N  K  T  I  G  Y  E  G  I  H  S
781 ACTGTTCAAGAAATGGCTGGATTGAATAACAAGACAATTGGATATGAGGGAATTCATAGC
781          790       800       810       820       830
781 TGACAAGTTCTTTACCGACCTAACTTATTGTTCTGTTAACCTATACTCCCTTAAGTATCG
281 P  S  V  L  P  S  G  G  E  E  S  R  S  P  S  L  Q  L  K  P
841 CCTAGTGTGCTTCCTTCTGGTGGAGAAGAAAGCAGATCACCATCTCTTCAACTTAAGCCT
841          850       860       870       880       890
841 GGATCACACGAAGGAAGACCACCTCTTCTTTCGTCTAGTGGTAGAGAAGTTGAATTCGGA
301 L  D  S  S  V  L  Q  F  S  S  K  Y  K  W  I  L  G  E  E  P
901 CTGGATTCCAGTGTTTTACAGTTTTCCAGTAAGTACAAATGGATCCTAGGTGAAGAACCG
901          910       920       930       940       950
901 GACCTAAGGTCACAAAATGTCAAAAGGTCATTCATGTTTACCTAGGATCCACTTCTTGGC
321 V  E  K  R  R  R  L  Q  N  E  M  T  T  P  S  L  D  Y  S  M
961 GTGGAGAAACGAAGAAGGCTCCAGAATGAGATGACAACACCTTCTCTAGATTATTCCATG
961          970       980       990       1000      1010
961 CACCTCTTTGCTTCTTCCGAGGTCTTACTCTACTGTTGTGGAAGAGATCTAATAAGGTAC
341 P  A  P  Y  R  R  V  E  A  P  V  A  Y  P  E  G  E  N  S  H
1021 CCTGCTCCTTACAGGAGGGTAGAAGCACCTGTTGCCTACCCAGAAGGGGAGAACAGCCAT
```

```
1021          1030      1040      1050      1060      1070
1021 GGACGAGGAATGTCCTCCCATCTTCGTGGACAACGGATGGGTCTTCCCCTCTTGTCGGTA
 361 D  K  S  S  S  E  R  S  T  P  P  Y  L  F  P  E  Y  P  E  A
1081 GATAAATCGAGTTCTGAGAGAAGTACACCACCATACCTTTTCCCAGAATACCCAGAAGCA
1081          1090      1100      1110      1120      1130
1081 CTATTTAGCTCAAGACTCTCTTCATGTGGTGGTATGGAAAAGGGTCTTATGGGTCTTCGT
 381 S  K  N  T  G  Q  N  R  E  V  S  I  L  Y  P  G  A  K  D  Q
1141 AGCAAGAATACAGGTCAGAATAGGGAGGTTTCAATTCTGTATCCAGGGGCCAAAGACCAA
1141          1150      1160      1170      1180      1190
1141 TCGTTCTTATGTCCAGTCTTATCCCTCCAAAGTTAAGACATAGGTCCCCGGTTTCTGGTT
 401 R  Q  G  S  L  L  P  E  E  L  E  D  Q  M  P  R  L  V  A  E
1201 CGCCAGGGGTCCCTGCTTCCTGAAGAATTAGAAGATCAGATGCCAAGATTGGTGGCAGAA
1201          1210      1220      1230      1240      1250
1201 GCGGTCCCCAGGGACGAAGGACTTCTTAATCTTCTAGTCTACGGTTCTAACCACCGTCTT
 421 E  S  N  R  G  S  T  T  I  N  K  E  E  V  N  K  G  P  F  V
1261 GAATCTAACAGAGGTAGCACAACCATAAACAAAGAAGAAGTCAACAAGGGACCTTTTGTA
1261          1270      1280      1290      1300      1310
1261 CTTAGATTGTCTCCATCGTGTTGGTATTTGTTTCTTCTTCAGTTGTTCCCTGGAAAACAT
                                >>>MORC4_f2>>> 1347 to 1364
 441 A  V  V  G  V  A  K  G  V  R  D  S  G  A  P  I  Q  L  I  P
1321 GCTGTTGTGGGTGTTGCCAAAGGTGTTAGAGATTCAGGAGCTCCCATTCAGCTGATCCCT
1321          1330      1340      1350      1360      1370
1321 CGACAACACCCACAACGGTTTCCACAATCTCTAAGTCCTCGAGGGTAAGTCGACTAGGGA
 461 F  N  R  E  E  L  A  E  R  R  K  A  V  E  S  W  N  P  V  P
1381 TTTAACAGAGAGGAGCTTGCTGAGAGACGAAAAGCAGTTGAATCCTGGAACCCAGTGCCT
1381          1390      1400      1410      1420      1430
1381 AAATTGTCTCTCCTCGAACGACTCTCTGCTTTTCGTCAACTTAGGACCTTGGGTCACGGA
 481 Y  S  V  A  S  A  A  I  P  A  A  A  I  G  E  K  A  R  G  Y
1441 TATTCTGTGGCCTCTGCTGCAATCCCTGCTGCAGCCATTGGGGAGAAAGCAAGAGGCTAT
1441          1450      1460      1470      1480      1490
1441 ATAAGACACCGGAGACGACGTTAGGGACGACGTCGGTAACCCCTCTTTCGTTCTCCGATA
 501 E  E  S  E  G  H  N  T  P  K  L  K  N  Q  R  E  L  E  E  L
1501 GAGGAGAGCGAAGGTCATAATACACCAAAGTTGAAGAACCAGAGAGAGCTGGAAGAATTG
1501          1510      1520      1530      1540      1550
1501 CTCCTCTCGCTTCCAGTATTATGTGGTTTCAACTTCTTGGTCTCTCTCGACCTTCTTAAC
 521 K  R  T  T  E  K  L  E  R  V  L  A  E  R  N  L  F  Q  Q  K
1561 AAGAGAACCACAGAAAAATTGGAACGTGTTTTGGCTGAAAGGAATTTGTTCCAGCAAAAG
1561          1570      1580      1590      1600      1610
1561 TTCTCTTGGTGTCTTTTTAACCTTGCACAAAACCGACTTTCCTTAAACAAGGTCGTTTTC
 541 V  E  E  L  E  Q  E  R  N  H  W  Q  S  E  F  K  K  V  Q  H
1621 GTGGAGGAGCTGGAACAGGAGAGGAATCACTGGCAGTCTGAATTCAAGAAAGTCCAACAT
1621          1630      1640      1650      1660      1670
1621 CACCTCCTCGACCTTGTCCTCTCCTTAGTGACCGTCAGACTTAAGTTCTTTCAGGTTGTA
 561 E  L  V  I  Y  S  T  Q  E  A  E  G  L  Y  W  S  K  K  H  M
1681 GAATTGGTGATCTACAGTACCCAGGAGGCGGAAGGCTTGTACTGGAGCAAGAAACACATG
1681          1690      1700      1710      1720      1730
1681 CTTAACCACTAGATGTCATGGGTCCTCCGCCTTCCGAACATGACCTCGTTCTTTGTGTAC
 581 G  Y  R  Q  A  E  F  Q  I  L  K  A  E  L  E  R  T  K  E  E
1741 GGTTATCGCCAAGCTGAATTCCAGATTCTGAAAGCTGAGCTGGAAAGAACCAAAGAGGAA
1741          1750      1760      1770      1780      1790
1741 CCAATAGCGGTTCGACTTAAGGTCTAAGACTTTCGACTCGACCTTTCTTGGTTTCTCCTT
                                <<<MORC4_r4<<< 1827 to 1844
 601 K  Q  E  L  K  E  K  L  K  E  T  E  T  H  L  E  M  L  Q  K
1801 AAGCAAGAGTTAAAAGAGAAACTGAAGGAAACAGAGACACACCTGGAAATGCTGCAGAAG
1801          1810      1820      1830      1840      1850
1801 TTCGTTCTCAATTTTCTCTTTGACTTCCTTTGTCTCTGTGTGGACCTTTACGACGTCTTC
 621 A  Q  G  F  G  K  A  Y  A  A  T  Y  P  R  Q  L  S  P  Y  F
1861 GCTCAGGGCTTTGGCAAAGCTTACGCGGCTACGTATCCACGTCAGCTATCTCCTTACTTC
1861          1870      1880      1890      1900      1910
1861 CGAGTCCCGAAACCGTTTCGAATGCGCCGATGCATAGGTGCAGTCGATAGAGGAATGAAG
       <<<MORC4_r3<<< 1927 to 1944
```

```
 641 C  P  P  S  L  G  A  S  *
1921 TGTCCTCCCTCACTTGGAGCTTCGTGA
```

MORC4 NM_001085354.3

```
   1 M  L  L  Y  R  G  A  P  A  G  P  G  A  P  G  C  G  L  A  R
   1 atgctcctgtaccgaggggccccccgccggccctggcgcgccgggctgcgggctggcccgg
   1          10        20        30        40        50
   1 tacgaggacatggctccccgggggcggccgggaccgcgcggcccgacgcccgaccgggcc
                          >>>MORC4_f1>>> 82 to 99
  21 P  G  G  G  P  Q  A  F  G  I  R  L  S  T  M  S  P  R  Y  L
  61 cccggcggcggcccgcaggccttcgggatccgcctgagcacgatgagcccccgctacctc
  61          70        80        90        100       110
  61 gggccgccgccgggcgtccggaagccctaggcggactcgtgctactcgggggcgatggag
  41 Q  S  N  S  S  S  H  T  R  P  F  S  A  I  A  E  L  L  D  N
 121 cagagcaactccagcagccacacgcgacccttcagtgccatcgcggagctgctagataat
 121          130       140       150       160       170
 121 gtctcgttgaggtcgtcggtgtgcgctgggaagtcacggtagcgcctcgacgatctatta
  61 A  V  D  P  D  V  S  A  R  T  V  F  I  D  V  E  E  V  K  N
 181 gctgtagatccagatgtatctgccaggacggtctttatagatgttgaggaggtcaagaat
 181          190       200       210       220       230
 181 cgacatctaggtctacatagacggtcctgccagaaatatctacaactcctccagttctta
  81 K  S  C  L  T  F  T  D  D  G  C  G  M  T  P  H  K  L  H  R
 241 aaatcttgtttgacctttaccgatgatggatgtgggatgacacctcataaactacaccga
 241          250       260       270       280       290
 241 tttagaacaaactggaaatggctactacctacaccctactgtggagtatttgatgtggct
 101 M  L  S  F  G  F  T  D  K  V  I  K  K  S  Q  C  P  I  G  V
 301 atgctcagctttggctttacagataaagtaataaagaagagccagtgtcccattggggtc
 301          310       320       330       340       350
 301 tacgagtcgaaaccgaaatgtctatttcattatttcttctcggtcacagggtaaccccag
 121 F  G  N  G  F  K  S  G  S  M  R  L  G  K  D  A  L  V  F  T
 361 tttggtaatggtttcaagtcaggctccatgcggctaggaaaggacgcccttgtcttcacc
 361          370       380       390       400       410
 361 aaaccattaccaaagttcagtccgaggtacgccgatcctttcctgcgggaacagaagtgg
 141 K  N  G  G  T  L  T  V  G  L  L  S  Q  T  Y  L  E  C  V  Q
 421 aagaatgggggtactctcactgttggacttctatcacagacctatctggaatgtgtccag
 421          430       440       450       460       470
 421 ttcttacccccatgagagtgacaacctgaagatagtgtctggatagaccttacacaggtc
 161 A  Q  A  V  I  V  P  I  V  P  F  N  Q  Q  N  K  K  M  I  I
 481 gcccaggcagttattgtaccaattgttccattcaaccagcaaaacaaaaaaatgattatt
 481          490       500       510       520       530
 481 cgggtccgtcaataacatggttaacaaggtaagttggtcgttttgttttttactaataa
 181 T  E  D  S  L  P  S  L  E  A  I  L  N  Y  S  I  F  N  R  E
 541 accgaggattcattgcccagcctagaagccatcttgaactattccattttcaaccgtgaa
 541          550       560       570       580       590
 541 tggctcctaagtaacgggtcggatcttcggtagaacttgataaggtaaaagttggcactt
 201 N  D  L  L  A  Q  F  D  A  I  P  G  K  K  G  T  R  V  L  I
 601 aatgacctgctggcccagtttgatgccatcccaggcaaaaaaggcactcgtgttctcatt
 601          610       620       630       640       650
 601 ttactggacgaccgggtcaaactacggtagggtccgtttttttccgtgagcacaagagtaa
                          <<<MORC4_r1<<< 685 to 702
 221 W  N  I  R  R  N  K  N  G  K  S  E  L  D  F  D  T  D  Q  Y
 661 tggaacatccgcagaaataaaaatggaaaatctgagttggactttgatacagatcaatat
 661          670       680       690       700       710
 661 accttgtaggcgtctttatttttacctttagactcaacctgaaactatgtctagttata
 241 D  I  L  V  S  D  F  D  T  E  E  K  M  T  G  G  V  T  S  E
 721 gacatcctggtatcagactttgacacagaagaaaaaatgactggcggtgttacctctgag
 721          730       740       750       760       770
 721 ctgtaggaccatagtctgaaactgtgtcttctttttactgaccgccacaatggagactc
 261 L  P  E  T  E  Y  S  L  R  A  F  C  G  I  L  Y  M  K  P  R
 781 ctaccagaaacagaatattctttaagggcattttgtggtattctatacatgaagccacgc
```

```
 781          790       800       810       820       830
 781 gatggtctttgtcttataagaaattcccgtaaaacaccataagatatgtacttcggtgcg
 281 M   K   I   F   L   R   Q   K   K   V   T   T   Q   M   I   A   K   S   L   A
 841 atgaaaatttttctgcgtcaaaagaaggtgactacccagatgattgccaagagcctggcc
 841          850       860       870       880       890
 841 tacttttaaaaagacgcagttttcttccactgatgggtctactaacggttctcggaccgg
 301 N   V   E   Y   D   T   Y   K   P   T   F   T   N   K   Q   V   R   I   T   F
 901 aatgtagaatatgatacatataaacctaccttcacaaataagcaggtgagaatcaccttt
 901          910       920       930       940       950
 901 ttacatcttatactatgtatatttggatggaagtgtttattcgtccactcttagtggaaa
 321 G   F   S   C   K   N   S   N   Q   F   G   I   M   M   Y   H   N   N   R   L
 961 gggttctcttgcaagaatagtaaccagtttggaataatgatgtatcataacaaccgactc
 961          970       980       990       1000      1010
 961 cccaagagaacgttcttatcattggtcaaaccttattactacatagtattgttggctgag
 341 I   K   S   F   E   K   V   G   C   Q   V   K   P   T   R   G   E   G   V   G
1021 ataaaatcttttgagaaggtggggtgccaggtgaagccaactcgtggagaaggtgtagga
1021          1030      1040      1050      1060      1070
1021 tattttagaaaactcttccaccccacggtccacttcggttgagcacctcttccacatcct
 361 V   I   G   V   I   E   C   N   F   L   K   P   A   Y   N   K   Q   D   F   E
1081 gtaattggagtcattgagtgcaatttcctaaaacctgcctacaacaaacaagactttgag
1081          1090      1100      1110      1120      1130
1081 cattaacctcagtaactcacgttaaaggattttggacggatgttgtttgttctgaaactc
 381 Y   T   K   E   Y   R   L   T   I   N   A   L   A   Q   K   L   N   A   Y   W
1141 tataccaaggagtaccggctaacaataaatgcccttgcccagaagctcaatgcttactgg
1141          1150      1160      1170      1180      1190
1141 atatggttcctcatggccgattgttatttacgggaacgggtcttcgagttacgaatgacc
 401 K   E   K   T   S   Q   D   N   F   E   T   S   T   V   A   R   P   I   P   K
1201 aaggaaaaaacatctcaagataattttgagacctcaactgtagccaggccaataccgaag
1201          1210      1220      1230      1240      1250
1201 ttccttttttgtagagttctattaaaactctggagttgacatcggtccggttatggcttc
 421 V   P   D   Q   T   W   V   Q   C   D   E   C   L   K   W   R   K   L   P   G
1261 gttcctgaccagacatgggttcagtgtgatgagtgtcttaaatggagaaagcttcctggg
1261          1270      1280      1290      1300      1310
1261 caaggactggtctgtacccaagtcacactactcacagaatttacctctttcgaaggaccc
 441 K   I   D   P   S   M   L   P   A   R   W   F   C   Y   Y   N   S   H   P   K
1321 aagattgatccatccatgttacctgcaagatggttttgttattataattcccatccaaag
1321          1330      1340      1350      1360      1370
1321 ttctaactaggtaggtacaatggacgttctaccaaaacaataatattaagggtaggtttc
 461 Y   R   R   C   S   V   P   E   E   Q   E   L   T   D   E   D   L   C   L   S
1381 tacaggagatgctctgttccagaggaacaagaactcactgatgaagacctgtgcttgagc
1381          1390      1400      1410      1420      1430
1381 atgtcctctacgagacaaggtctccttgttcttgagtgactacttctggacacgaactcg
 481 K   A   K   K   Q   E   Q   T   V   E   E   K   K   K   M   P   M   E   N   E
1441 aaagctaagaaacaagaacaaactgttgaggagaagaagaagatgcctatggaaaatgag
1441          1450      1460      1470      1480      1490
1441 tttcgattctttgttcttgtttgacaactcctcttcttcttctacggatacctttactc
 501 N   H   Q   V   F   S   N   P   P   K   I   L   T   V   Q   E   M   A   G   L
1501 aaccaccaggtattcagtaatccaccaaagatccttactgttcaagaaatggctggattg
1501          1510      1520      1530      1540      1550
1501 ttggtggtccataagtcattaggtggtttctaggaatgacaagttctttaccgacctaac
 521 N   N   K   T   I   G   Y   E   G   I   H   S   P   S   V   L   P   S   G   G
1561 aataacaagacaattggatatgagggaattcatagccctagtgtgcttccttctggtgga
1561          1570      1580      1590      1600      1610
1561 ttattgttctgttaacctatactcccttaagtatcgggatcacacgaaggaagaccacct
 541 E   E   S   R   S   P   S   L   Q   L   K   P   L   D   S   S   V   L   Q   F
1621 gaagaaagcagatcaccatctcttcaacttaagcctctggattccagtgttttacagttt
1621          1630      1640      1650      1660      1670
1621 cttctttcgtctagtggtagagaagttgaattcggagacctaaggtcacaaaatgtcaaa
 561 S   S   K   Y   K   W   I   L   G   E   E   P   V   E   K   R   R   R   L   Q
1681 tccagtaagtacaaatggatcctaggtgaagaaccggtggagaaacgaagaaggctccag
1681          1690      1700      1710      1720      1730
```

```
1681 aggtcattcatgtttacctaggatccacttcttggccacctctttgcttcttccgaggtc
 581 N  E  M  T  T  P  S  L  D  Y  S  M  P  A  P  Y  R  R  V  E
1741 aatgagatgacaacaccttctctagattattccatgcctgctccttacaggagggtagaa
1741          1750      1760      1770      1780      1790
1741 ttactctactgttgtggaagagatctaataaggtacggacgaggaatgtcctcccatctt
 601 A  P  V  A  Y  P  E  G  E  N  S  H  D  K  S  S  S  E  R  S
1801 gcacctgttgcctacccagaaggggagaacagccatgataaatcgagttctgagagaagt
1801          1810      1820      1830      1840      1850
1801 cgtggacaacggatgggtcttcccctcttgtcggtactatttagctcaagactctcttca
 621 T  P  P  Y  L  F  P  E  Y  P  E  A  S  K  N  T  G  Q  N  R
1861 acaccaccatacctttcccagaatacccagaagcaagcaagaatacaggtcagaatagg
1861          1870      1880      1890      1900      1910
1861 tgtggtggtatggaaaagggtcttatgggtcttcgttcgttcttatgtccagtcttatcc
 641 E  V  S  I  L  Y  P  G  A  K  D  Q  R  Q  G  S  L  L  P  E
1921 gaggtttcaattctgtatccaggggccaaagaccaacgccaggggtccctgcttcctgaa
1921          1930      1940      1950      1960      1970
1921 ctccaaagttaagacataggtccccggtttctggttgcggtccccagggacgaaggactt
 661 E  L  E  D  Q  M  P  R  L  V  A  E  E  S  N  R  G  S  T  T
1981 gaattagaagatcagatgccaagattggtggcagaagaatctaacagaggtagcacaacc
1981          1990      2000      2010      2020      2030
1981 cttaatcttctagtctacggttctaaccaccgtcttcttagattgtctccatcgtgttgg
 681 I  N  K  E  E  V  N  K  G  P  F  V  A  V  V  G  V  A  K  G
2041 ataaacaaagaagaagtcaacaagggacctttgtagctgttgtgggtgttgccaaaggt
2041          2050      2060      2070      2080      2090
2041 tatttgtttcttcttcagttgttccctggaaaacatcgacaacacccacaacggtttcca
       >>>MORC4_f2>>> 2103 to 2120
 701 V  R  D  S  G  A  P  I  Q  L  I  P  F  N  R  E  E  L  A  E
2101 gttagagattcaggagctcccattcagctgatccctttaacagagaggagcttgctgag
2101          2110      2120      2130      2140      2150
2101 caatctctaagtcctcgagggtaagtcgactagggaaaattgtctctcctcgaacgactc
 721 R  R  K  A  V  E  S  W  N  P  V  P  Y  S  V  A  S  A  A  I
2161 agacgaaaagcagttgaatcctggaacccagtgccttattctgtggcctctgctgcaatc
2161          2170      2180      2190      2200      2210
2161 tctgcttttcgtcaacttaggaccttgggtcacggaataagacaccggagacgacgttag
 741 P  A  A  A  I  G  E  K  A  R  G  Y  E  E  S  E  G  H  N  T
2221 cctgctgcagccattggggagaaagcaagaggctatgaggagagcgaaggtcataataca
2221          2230      2240      2250      2260      2270
2221 ggacgacgtcggtaacccctctttcgttctccgatactcctctcgcttccagtattatgt
 761 P  K  L  K  N  Q  R  E  L  E  E  L  K  R  T  T  E  K  L  E
2281 ccaaagttgaagaaccagagagagctggaagaattgaagagaaccacagaaaaattggaa
2281          2290      2300      2310      2320      2330
2281 ggtttcaacttcttggtctctctcgaccttcttaacttctcttggtgtctttttaaccttt
 781 R  V  L  A  E  R  N  L  F  Q  Q  K  V  E  E  L  E  Q  E  R
2341 cgtgttttggctgaaaggaatttgttccagcaaaaggtggaggagctggaacaggagagg
2341          2350      2360      2370      2380      2390
2341 gcacaaaaccgactttccttaaacaaggtcgttttccacctcctcgaccttgtcctctcc
 801 N  H  W  Q  S  E  F  K  K  V  Q  H  E  L  V  I  Y  S  T  Q
2401 aatcactggcagtctgaattcaagaaagtccaacatgaattggtgatctacagtacccag
2401          2410      2420      2430      2440      2450
2401 ttagtgaccgtcagacttaagttctttcaggttgtacttaaccactagatgtcatgggtc
 821 E  A  E  G  L  Y  W  S  K  K  H  M  G  Y  R  Q  A  E  F  Q
2461 gaggcggaaggcttgtactggagcaagaaacacatgggttatcgccaagctgaattccag
2461          2470      2480      2490      2500      2510
2461 ctccgccttccgaacatgacctcgttctttgtgtacccaatagcggttcgacttaaggtc
 841 I  L  K  A  E  L  E  R  T  K  E  E  K  Q  E  L  K  E  K  L
2521 attctgaaagctgagctggaaagaaccaaagaggaaaagcaagagttaaaagagaaactg
2521          2530      2540      2550      2560      2570
2521 taagactttcgactcgacctttcttggtttctccttttcgttctcaattttctctttgac
       <<<MORC4_r4<<< 2583 to 2600
 861 K  E  T  E  T  H  L  E  M  L  Q  K  A  Q  G  F  G  K  A  Y
2581 aaggaaacagagacacacctggaaatgctgcagaaggctcagggctttggcaaagcttac
```

```
2581          2590      2600      2610      2620      2630
2581 ttcctttgtctctgtgtggacctttacgacgtcttccgagtcccgaaaccgtttcgaatg
                                                <<<MORC4_r3<<< 2683 to
2700
 881 A  A  T  Y  P  R  Q  L  S  P  Y  F  C  P  P  S  L  G  A  S
2641 gcggctacgtatccacgtcagctatctccttacttctgtcctccctcacttggagcttcg
2641          2650      2660      2670      2680      2690
2641 cgccgatgcataggtgcagtcgatagaggaatgaagacaggagggagtgaacctcgaagc
 901 *
2701 tga
```

MORC4 XM_011531027.2

```
   1 M  I  I  T  E  D  S  L  P  S  L  E  A  I  L  N  Y  S  I  F
   1 atgattattaccgaggattcattgcccagcctagaagccatcttgaactattccattttc
   1          10        20        30        40        50
   1 tactaataatggctcctaagtaacgggtcggatcttcggtagaacttgataaggtaaaag
  21 N  R  E  N  D  L  L  A  Q  F  D  A  I  P  G  K  K  G  T  R
  61 aaccgtgaaaatgacctgctggcccagtttgatgccatcccaggcaaaaaaggcactcgt
  61          70        80        90        100       110
  61 ttggcacttttactggacgaccgggtcaaactacggtagggtccgtttttttccgtgagca
                                      <<<MORC4_r1<<< 154 to 171
  41 V  L  I  W  N  I  R  R  N  K  N  G  K  S  E  L  D  F  D  T
 121 gttctcatttggaacatccgcagaaataaaaatggaaaatctgagttggactttgataca
 121          130       140       150       160       170
 121 caagagtaaaccttgtaggcgtctttatttttaccttttagactcaacctgaaactatgt
  61 D  Q  Y  D  I  L  V  S  D  F  D  T  E  E  K  M  T  G  G  V
 181 gatcaatatgacatcctggtatcagactttgacacagaagaaaaaatgactggcggtgtt
 181          190       200       210       220       230
 181 ctagttatactgtaggaccatagtctgaaactgtgtcttctttttttactgaccgccacaa
  81 T  S  E  L  P  E  T  E  Y  S  L  R  A  F  C  G  I  L  Y  M
 241 acctctgagctaccagaaacagaatattctttaagggcattttgtggtattctatacatg
 241          250       260       270       280       290
 241 tggagactcgatggtctttgtcttataagaaattcccgtaaaacaccataagatatgtac
 101 K  P  R  M  K  I  F  L  R  Q  K  K  V  T  T  Q  M  I  A  K
 301 aagccacgcatgaaaatttttctgcgtcaaaagaaggtgactacccagatgattgccaag
 301          310       320       330       340       350
 301 ttcggtgcgtacttttaaaaagacgcagttttcttccactgatgggtctactaacggttc
 121 S  L  A  N  V  E  Y  D  T  Y  K  P  T  F  T  N  K  Q  V  R
 361 agcctggccaatgtagaatatgatacatataaacctaccttcacaaataagcaggtgaga
 361          370       380       390       400       410
 361 tcggaccggttacatcttatactatgtatatttggatggaagtgtttattcgtccactct
 141 I  T  F  G  F  S  C  K  N  S  N  Q  F  G  I  M  M  Y  H  N
 421 atcacctttgggttctcttgcaagaatagtaaccagtttggaataatgatgtatcataac
 421          430       440       450       460       470
 421 tagtggaaacccaagagaacgttcttatcattggtcaaaccttattactacatagtattg
 161 N  R  L  I  K  S  F  E  K  V  G  C  Q  V  K  P  T  R  G  E
 481 aaccgactcataaaatcttttgagaaggtggggtgccaggtgaagccaactcgtggagaa
 481          490       500       510       520       530
 481 ttggctgagtattttagaaaactcttccaccccacggtccacttcggttgagcacctctt
 181 G  V  G  V  I  G  V  I  E  C  N  F  L  K  P  A  Y  N  K  Q
 541 ggtgtaggagtaattggagtcattgagtgcaatttcctaaaacctgcctacaacaaacaa
 541          550       560       570       580       590
 541 ccacatcctcattaacctcagtaactcacgttaaaggattttggacggatgttgtttgtt
 201 D  F  E  Y  T  K  E  Y  R  L  T  I  N  A  L  A  Q  K  L  N
 601 gactttgagtataccaaggagtaccggctaacaataaatgcccttgcccagaagctcaat
 601          610       620       630       640       650
 601 ctgaaactcatatggttcctcatggccgattgttatttacgggaacgggtcttcgagtta
 221 A  Y  W  K  E  K  T  S  Q  D  N  F  E  T  S  T  V  A  R  P
```

```
 661 gcttactggaaggaaaaaacatctcaagataattttgagacctcaactgtagccaggcca
 661          670       680       690       700       710
 661 cgaatgaccttccttttttgtagagttctattaaaactctggagttgacatcggtccggt
 241 I   P   K   V   P   D   Q   T   W   V   Q   C   D   E   C   L   K   W   R   K
 721 ataccgaaggttcctgaccagacatgggttcagtgtgatgagtgtcttaaatggagaaag
 721          730       740       750       760       770
 721 tatggcttccaaggactggtctgtacccaagtcacactactcacagaatttacctctttc
 261 L   P   G   K   I   D   P   S   M   L   P   A   R   W   F   C   Y   Y   N   S
 781 cttcctgggaagattgatccatccatgttacctgcaagatggttttgttattataattcc
 781          790       800       810       820       830
 781 gaaggacccttctaactaggtaggtacaatggacgttctaccaaaacaataatattaagg
 281 H   P   K   Y   R   R   C   S   V   P   E   E   Q   E   L   T   D   E   D   L
 841 catccaaagtacaggagatgctctgttccagaggaacaagaactcactgatgaagacctg
 841          850       860       870       880       890
 841 gtaggtttcatgtcctctacgagacaaggtctccttgttcttgagtgactacttctggac
 301 C   L   S   K   A   K   K   Q   E   Q   T   V   E   E   K   K   K   M   P   M
 901 tgcttgagcaaagctaagaaacaagaacaaactgttgaggagaagaagaagatgcctatg
 901          910       920       930       940       950
 901 acgaactcgtttcgattctttgttcttgtttgacaactcctcttcttcttctacggatac
 321 E   N   E   N   H   Q   V   F   S   N   P   P   K   I   L   T   V   Q   E   M
 961 gaaaatgagaaccaccaggtattcagtaatccaccaaagatccttactgttcaagaaatg
 961          970       980       990       1000      1010
 961 cttttactcttggtggtccataagtcattaggtggtttctaggaatgacaagttctttac
 341 A   G   L   N   N   K   T   I   G   Y   E   G   I   H   S   P   S   V   L   P
1021 gctggattgaataacaagacaattggatatgagggaattcatagccctagtgtgcttcct
1021          1030      1040      1050      1060      1070
1021 cgacctaacttattgttctgttaacctatactcccttaagtatcgggatcacacgaagga
 361 S   G   G   E   E   S   R   S   P   S   L   Q   L   K   P   L   D   S   S   V
1081 tctggtggagaagaaagcagatcaccatctcttcaacttaagcctctggattccagtgtt
1081          1090      1100      1110      1120      1130
1081 agaccacctcttctttcgtctagtggtagagaagttgaattcggagacctaaggtcacaa
 381 L   Q   F   S   S   K   Y   K   W   I   L   G   E   E   P   V   E   K   R   R
1141 ttacagttttccagtaagtacaaatggatcctaggtgaagaaccggtggagaaacgaaga
1141          1150      1160      1170      1180      1190
1141 aatgtcaaaaggtcattcatgtttacctaggatccacttcttggccacctctttgcttct
 401 R   L   Q   N   E   M   T   T   P   S   L   D   Y   S   M   P   A   P   Y   R
1201 aggctccagaatgagatgacaacaccttctctagattattccatgcctgctccttacagg
1201          1210      1220      1230      1240      1250
1201 tccgaggtcttactctactgttgtggaagagatctaataaggtacggacgaggaatgtcc
 421 R   V   E   A   P   V   A   Y   P   E   G   E   N   S   H   D   K   S   S   S
1261 agggtagaagcacctgttgcctacccagaaggggagaacagccatgataaatcgagttct
1261          1270      1280      1290      1300      1310
1261 tcccatcttcgtggacaacggatgggtcttcccctcttgtcggtactatttagctcaaga
 441 E   R   S   T   P   P   Y   L   F   P   E   Y   P   E   A   S   K   N   T   G
1321 gagagaagtacaccaccatacctttcccagaatacccagaagcaagcaagaatacaggt
1321          1330      1340      1350      1360      1370
1321 ctctcttcatgtggtggtatggaaaagggtcttatgggtcttcgttcgttcttatgtcca
 461 Q   N   R   E   V   S   I   L   Y   P   G   A   K   D   Q   R   Q   G   S   L
1381 cagaatagggaggtttcaattctgtatccaggggccaaagaccaacgccaggggtccctg
1381          1390      1400      1410      1420      1430
1381 gtcttatccctccaaagttaagacataggtccccggtttctggttgcggtccccagggac
 481 L   P   E   E   L   E   D   Q   M   P   R   L   V   A   E   E   S   N   R   G
1441 cttcctgaagaattagaagatcagatgccaagattggtggcagaagaatctaacagaggt
1441          1450      1460      1470      1480      1490
1441 gaaggacttcttaatcttctagtctacggttctaaccaccgtcttcttagattgtctcca
 501 S   T   T   I   N   K   E   E   V   N   K   G   P   F   V   A   V   V   G   V
1501 agcacaaccataaacaaagaagaagtcaacaagggacctttgtagctgttgtgggtgtt
1501          1510      1520      1530      1540      1550
1501 tcgtgttggtatttgtttcttcttcagttgttccctggaaaacatcgacaacacccacaa
                  >>>MORC4_f2>>> 1572 to 1589
 521 A   K   G   V   R   D   S   G   A   P   I   Q   L   I   P   F   N   R   E   E
```

```
1561 gccaaaggtgttagagattcaggagctcccattcagctgatccctttaacagagaggag
1561          1570      1580      1590      1600      1610
1561 cggtttccacaatctctaagtcctcgagggtaagtcgactagggaaaattgtctctcctc
 541 L  A  E  R  R  K  A  V  E  S  W  N  P  V  P  Y  S  V  A  S
1621 cttgctgagagacgaaaagcagttgaatcctggaacccagtgccttattctgtggcctct
1621          1630      1640      1650      1660      1670
1621 gaacgactctctgcttttcgtcaacttaggaccttgggtcacggaataagacaccggaga
 561 A  A  I  P  A  A  A  I  G  E  K  A  R  G  Y  E  E  S  E  G
1681 gctgcaatccctgctgcagccattggggagaaagcaagaggctatgaggagagcgaaggt
1681          1690      1700      1710      1720      1730
1681 cgacgttagggacgacgtcggtaacccctctttcgttctccgatactcctctcgcttcca
 581 H  N  T  P  K  L  K  N  Q  R  E  L  E  E  L  K  R  T  T  E
1741 cataatacaccaaagttgaagaaccagagagagctggaagaattgaagagaaccacagaa
1741          1750      1760      1770      1780      1790
1741 gtattatgtggtttcaacttcttggtctctctcgaccttcttaacttctcttggtgtctt
 601 K  L  E  R  V  L  A  E  R  N  L  F  Q  Q  K  V  E  E  L  E
1801 aaattggaacgtgttttggctgaaaggaatttgttccagcaaaaggtggaggagctggaa
1801          1810      1820      1830      1840      1850
1801 tttaaccttgcacaaaaccgactttccttaaacaaggtcgttttccacctcctcgacctt
 621 Q  E  R  N  H  W  Q  S  E  F  K  K  V  Q  H  E  L  V  I  Y
1861 caggagaggaatcactggcagtctgaattcaagaaagtccaacatgaattggtgatctac
1861          1870      1880      1890      1900      1910
1861 gtcctctccttagtgaccgtcagacttaagttctttcaggttgtacttaaccactagatg
 641 S  T  Q  E  A  E  G  L  Y  W  S  K  K  H  M  G  Y  R  Q  A
1921 agtacccaggaggcggaaggcttgtactggagcaagaaacacatgggttatcgccaagct
1921          1930      1940      1950      1960      1970
1921 tcatgggtcctccgccttccgaacatgacctcgttctttgtgtacccaatagcggttcga
 661 E  F  Q  I  L  K  A  E  L  E  R  T  K  E  E  K  Q  E  L  K
1981 gaattccagattctgaaagctgagctggaaagaaccaaagaggaaaagcaagagttaaaa
1981          1990      2000      2010      2020      2030
1981 cttaaggtctaagactttcgactcgacctttcttggtttctccttttcgttctcaatttt
               <<<MORC4_r4<<< 2052 to 2069
 681 E  K  L  K  E  T  E  T  H  L  E  M  L  Q  K  A  Q  V  S  Y
2041 gagaaactgaaggaaacagagacacacctggaaatgctgcagaaggctcaggtctcctac
2041          2050      2060      2070      2080      2090
2041 ctctttgacttccttttgtctctgtgtggacctttacgacgtcttccgagtccagaggatg
 701 R  T  P  E  G  D  D  L  E  R  A  L  A  K  L  T  R  L  R  I
2101 cggaccccagagggagatgacctagaaagggctttggcaaagcttacgcggctacgtatc
2101          2110      2120      2130      2140      2150
2101 gcctggggtctccctctactggatctttcccgaaaccgtttcgaatgcgccgatgcatag
                                  <<<MORC4_r3<<< 2190 to 2207
 721 H  V  S  Y  L  L  T  S  V  L  P  H  L  E  L  R  E  I  G  Y
2161 cacgtcagctatctccttacttctgtcctccctcacttggagcttcgtgagatcgggtat
2161          2170      2180      2190      2200      2210
2161 gtgcagtcgatagaggaatgaagacaggagggagtgaacctcgaagcactctagcccata
                                                  <<<MORC4_r2<<< 2263 to
2280
 741 D  S  E  Q  V  D  G  I  L  Y  T  V  L  E  A  N  H  I  L  D
2221 gactcagaacaagtggatgggatcctgtacacggtgctggaggcaaatcacatactggat
2221          2230      2240      2250      2260      2270
2221 ctgagtcttgttcacctaccctaggacatgtgccacgacctccgtttagtgtatgaccta
 761 *
2281 tga
```

MORC4 XM_017029844.1

```
   1 M  T  P  H  K  L  H  R  M  L  S  F  G  F  T  D  K  V  I  K
   1 atgacacctcataaactacaccgaatgctcagctttggctttacagataaagtaataaag
   1          10        20        30        40        50
   1 tactgtggagtatttgatgtggcttacgagtcgaaaccgaaatgtctatttcattatttc
```

```
 21 K  S  Q  C  P  I  G  V  F  G  N  G  F  K  S  G  S  M  R  L
 61 aagagccagtgtcccattggggtctttggtaatggtttcaagtcaggctccatgcggcta
 61          70        80        90        100       110
 61 ttctcggtcacagggtaaccccagaaaccattaccaaagttcagtccgaggtacgccgat
 41 G  K  D  A  L  V  F  T  K  N  G  G  T  L  T  V  G  L  L  S
121 ggaaaggacgcccttgtcttcaccaagaatgggggtactctcactgttggacttctatca
121          130       140       150       160       170
121 cctttcctgcgggaacagaagtggttcttacccccatgagagtgacaacctgaagatagt
 61 Q  T  Y  L  E  C  V  Q  A  Q  A  V  I  V  P  I  V  P  F  N
181 cagacctatctggaatgtgtccaggcccaggcagttattgtaccaattgttccattcaac
181          190       200       210       220       230
181 gtctggatagaccttacacaggtccgggtccgtcaataacatggttaacaaggtaagttg
 81 Q  Q  N  K  K  M  I  I  T  E  D  S  L  P  S  L  E  A  I  L
241 cagcaaaacaaaaaaatgattattaccgaggattcattgcccagcctagaagccatcttg
241          250       260       270       280       290
241 gtcgttttgttttttactaataatggctcctaagtaacgggtcggatcttcggtagaac
101 N  Y  S  I  F  N  R  E  N  D  L  L  A  Q  F  D  A  I  P  G
301 aactattccattttcaaccgtgaaaatgacctgctggcccagtttgatgccatcccaggc
301          310       320       330       340       350
301 ttgataaggtaaaagttggcacttttactggacgaccgggtcaaactacggtagggtccg
                                                  <<<MORC4_r1<<< 409
to 426
121 K  K  G  T  R  V  L  I  W  N  I  R  R  N  K  N  G  K  S  E
361 aaaaaaggcactcgtgttctcatttggaacatccgcagaaataaaaatggaaaatctgag
361          370       380       390       400       410
361 ttttttccgtgagcacaagagtaaaccttgtaggcgtctttatttttacctttagactc
141 L  D  F  D  T  D  Q  Y  D  I  L  V  S  D  F  D  T  E  E  K
421 ttggactttgatacagatcaatatgacatcctggtatcagactttgacacagaagaaaaa
421          430       440       450       460       470
421 aacctgaaactatgtctagttatactgtaggaccatagtctgaaactgtgtcttctttt
161 M  T  G  G  V  T  S  E  L  P  E  T  E  Y  S  L  R  A  F  C
481 atgactggcggtgttacctctgagctaccagaaacagaatattctttaagggcattttgt
481          490       500       510       520       530
481 tactgaccgccacaatggagactcgatggtctttgtcttataagaaattcccgtaaaaca
181 G  I  L  Y  M  K  P  R  M  K  I  F  L  R  Q  K  K  V  T  T
541 ggtattctatacatgaagccacgcatgaaaatttttctgcgtcaaaagaaggtgactacc
541          550       560       570       580       590
541 ccataagatatgtacttcggtgcgtacttttaaaaagacgcagttttcttccactgatgg
201 Q  M  I  A  K  S  L  A  N  V  E  Y  D  T  Y  K  P  T  F  T
601 cagatgattgccaagagcctggccaatgtagaatatgatacatataaacctaccttcaca
601          610       620       630       640       650
601 gtctactaacggttctcggaccggttacatcttatactatgtatatttggatggaagtgt
221 N  K  Q  V  R  I  T  F  G  F  S  C  K  N  S  N  Q  F  G  I
661 aataagcaggtgagaatcacctttgggttctcttgcaagaatagtaaccagtttggaata
661          670       680       690       700       710
661 ttattcgtccactcttagtggaaacccaagagaacgttcttatcattggtcaaaccttat
241 M  M  Y  H  N  N  R  L  I  K  S  F  E  K  V  G  C  Q  V  K
721 atgatgtatcataacaaccgactcataaaatcttttgagaaggtggggtgccaggtgaag
721          730       740       750       760       770
721 tactacatagtattgttggctgagtattttagaaaactcttccaccccacggtccacttc
261 P  T  R  G  E  G  V  G  V  I  G  V  I  E  C  N  F  L  K  P
781 ccaactcgtggagaaggtgtaggagtaattggagtcattgagtgcaatttcctaaaacct
781          790       800       810       820       830
781 ggttgagcacctcttccacatcctcattaacctcagtaactcacgttaaaggattttgga
281 A  Y  N  K  Q  D  F  E  Y  T  K  E  Y  R  L  T  I  N  A  L
841 gcctacaacaaacaagactttgagtataccaaggagtaccggctaacaataaatgcccttt
841          850       860       870       880       890
841 cggatgttgtttgttctgaaactcatatggttcctcatggccgattgttatttacgggaa
301 A  Q  K  L  N  A  Y  W  K  E  K  T  S  Q  D  N  F  E  T  S
901 gcccagaagctcaatgcttactggaaggaaaaaacatctcaagataattttgagacctca
901          910       920       930       940       950
```

```
 901 cgggtcttcgagttacgaatgaccttccttttttgtagagttctattaaaactctggagt
 321 T  V  A  R  P  I  P  K  V  P  D  Q  T  W  V  Q  C  D  E  C
 961 actgtagccaggccaataccgaaggttcctgaccagacatgggttcagtgtgatgagtgt
 961          970       980       990       1000      1010
 961 tgacatcggtccggttatggcttccaaggactggtctgtacccaagtcacactactcaca
 341 L  K  W  R  K  L  P  G  K  I  D  P  S  M  L  P  A  R  W  F
1021 cttaaatggagaaagcttcctgggaagattgatccatccatgttacctgcaagatggttt
1021          1030      1040      1050      1060      1070
1021 gaatttacctctttcgaaggacccttctaactaggtaggtacaatggacgttctaccaaa
 361 C  Y  Y  N  S  H  P  K  Y  R  R  C  S  V  P  E  E  Q  E  L
1081 tgttattataattcccatccaaagtacaggagatgctctgttccagaggaacaagaactc
1081          1090      1100      1110      1120      1130
1081 acaataatattaagggtaggtttcatgtcctctacgagacaaggtctccttgttcttgag
 381 T  D  E  D  L  C  L  S  K  A  K  K  Q  E  Q  T  V  E  E  K
1141 actgatgaagacctgtgcttgagcaaagctaagaaacaagaacaaactgttgaggagaag
1141          1150      1160      1170      1180      1190
1141 tgactacttctggacacgaactcgtttcgattctttgttcttgtttgacaactcctcttc
 401 K  K  M  P  M  E  N  E  N  H  Q  V  F  S  N  P  P  K  I  L
1201 aagaagatgcctatggaaaatgagaaccaccaggtattcagtaatccaccaaagatcctt
1201          1210      1220      1230      1240      1250
1201 ttcttctacggatacctttttactcttggtggtccataagtcattaggtggtttctaggaa
 421 T  V  Q  E  M  A  G  L  N  N  K  T  I  G  Y  E  G  I  H  S
1261 actgttcaagaaatggctggattgaataacaagacaattggatatgagggaattcatagc
1261          1270      1280      1290      1300      1310
1261 tgacaagttctttaccgacctaacttattgttctgttaacctatactcccttaagtatcg
 441 P  S  V  L  P  S  G  G  E  E  S  R  S  P  S  L  Q  L  K  P
1321 cctagtgtgcttccttctggtggagaagaaagcagatcaccatctcttcaacttaagcct
1321          1330      1340      1350      1360      1370
1321 ggatcacacgaaggaagaccacctcttctttcgtctagtggtagagaagttgaattcgga
 461 L  D  S  S  V  L  Q  F  S  S  K  Y  K  W  I  L  G  E  E  P
1381 ctggattccagtgttttacagttttccagtaagtacaaatggatcctaggtgaagaaccg
1381          1390      1400      1410      1420      1430
1381 gacctaaggtcacaaaatgtcaaaaggtcattcatgtttacctaggatccacttcttggc
 481 V  E  K  R  R  R  L  Q  N  E  M  T  T  P  S  L  D  Y  S  M
1441 gtggagaaacgaagaaggctccagaatgagatgacaacaccttctctagattattccatg
1441          1450      1460      1470      1480      1490
1441 cacctctttgcttcttccgaggtcttactctactgttgtggaagagatctaataaggtac
 501 P  A  P  Y  R  R  V  E  A  P  V  A  Y  P  E  G  E  N  S  H
1501 cctgctccttacaggagggtagaagcacctgttgcctacccagaaggggagaacagccat
1501          1510      1520      1530      1540      1550
1501 ggacgaggaatgtcctcccatcttcgtggacaacggatgggtcttcccctcttgtcggta
 521 D  K  S  S  S  E  R  S  T  P  P  Y  L  F  P  E  Y  P  E  A
1561 gataaatcgagttctgagagaagtacaccaccatacctttcccagaatacccagaagca
1561          1570      1580      1590      1600      1610
1561 ctatttagctcaagactctcttcatgtggtggtatggaaaagggtcttatgggtcttcgt
 541 S  K  N  T  G  Q  N  R  E  V  S  I  L  Y  P  G  A  K  D  Q
1621 agcaagaatacaggtcagaatagggaggtttcaattctgtatccaggggccaaagaccaa
1621          1630      1640      1650      1660      1670
1621 tcgttcttatgtccagtcttatccctccaaagttaagacataggtccccggtttctggtt
 561 R  Q  G  S  L  L  P  E  E  L  E  D  Q  M  P  R  L  V  A  E
1681 cgccaggggtccctgcttcctgaagaattagaagatcagatgccaagattggtggcagaa
1681          1690      1700      1710      1720      1730
1681 gcggtccccagggacgaaggacttcttaatcttctagtctacggttctaaccaccgtctt
 581 E  S  N  R  G  S  T  T  I  N  K  E  E  V  N  K  G  P  F  V
1741 gaatctaacagaggtagcacaaccataaacaaagaagaagtcaacaagggacctttgta
1741          1750      1760      1770      1780      1790
1741 cttagattgtctccatcgtgttggtatttgtttcttcttcagttgttccctggaaaacat
                                >>>MORC4_f2>>> 1827 to 1844
 601 A  V  V  G  V  A  K  G  V  R  D  S  G  A  P  I  Q  L  I  P
1801 gctgttgtgggtgttgccaaaggtgttagagattcaggagctcccattcagctgatccct
1801          1810      1820      1830      1840      1850
```

```
1801 cgacaacacccacaacggtttccacaatctctaagtcctcgagggtaagtcgactaggga
 621 F  N  R  E  E  L  A  E  R  R  K  A  V  E  S  W  N  P  V  P
1861 tttaacagagaggagcttgctgagagacgaaaagcagttgaatcctggaacccagtgcct
1861          1870      1880      1890      1900      1910
1861 aaattgtctctcctcgaacgactctctgcttttcgtcaacttaggaccttgggtcacgga
 641 Y  S  V  A  S  A  A  I  P  A  A  A  I  G  E  K  A  R  G  Y
1921 tattctgtggcctctgctgcaatccctgctgcagccattggggagaaagcaagaggctat
1921          1930      1940      1950      1960      1970
1921 ataagacaccggagacgacgttagggacgacgtcggtaacccctctttcgttctccgata
 661 E  E  S  E  G  H  N  T  P  K  L  K  N  Q  R  E  L  E  E  L
1981 gaggagagcgaaggtcataatacaccaaagttgaagaaccagagagagctggaagaattg
1981          1990      2000      2010      2020      2030
1981 ctcctctcgcttccagtattatgtggtttcaacttcttggtctctctcgaccttcttaac
 681 K  R  T  T  E  K  L  E  R  V  L  A  E  R  N  L  F  Q  Q  K
2041 aagagaaccacagaaaaattggaacgtgttttggctgaaaggaatttgttccagcaaaag
2041          2050      2060      2070      2080      2090
2041 ttctcttggtgtctttttaaccttgcacaaaaccgactttccttaaacaaggtcgttttc
 701 V  E  E  L  E  Q  E  R  N  H  W  Q  S  E  F  K  K  V  Q  H
2101 gtggaggagctggaacaggagaggaatcactggcagtctgaattcaagaaagtccaacat
2101          2110      2120      2130      2140      2150
2101 cacctcctcgaccttgtcctctccttagtgaccgtcagacttaagttctttcaggttgta
 721 E  L  V  I  Y  S  T  Q  E  A  E  G  L  Y  W  S  K  K  H  M
2161 gaattggtgatctacagtacccaggaggcggaaggcttgtactggagcaagaaacacatg
2161          2170      2180      2190      2200      2210
2161 cttaaccactagatgtcatgggtcctccgccttccgaacatgacctcgttctttgtgtac
 741 G  Y  R  Q  A  E  F  Q  I  L  K  A  E  L  E  R  T  K  E  E
2221 ggttatcgccaagctgaattccagattctgaaagctgagctggaaagaaccaaagaggaa
2221          2230      2240      2250      2260      2270
2221 ccaatagcggttcgacttaaggtctaagactttcgactcgacctttcttggtttctcctt
                                <<<MORC4_r4<<< 2307 to 2324
 761 K  Q  E  L  K  E  K  L  K  E  T  E  T  H  L  E  M  L  Q  K
2281 aagcaagagttaaaagagaaactgaaggaaacagagacacacctggaaatgctgcagaag
2281          2290      2300      2310      2320      2330
2281 ttcgttctcaattttctctttgacttcctttgtctctgtgtggacctttacgacgtcttc
 781 A  Q  G  F  G  K  A  Y  A  A  T  Y  P  R  Q  L  S  P  Y  F
2341 gctcagggctttggcaaagcttacgcggctacgtatccacgtcagctatctccttacttc
2341          2350      2360      2370      2380      2390
2341 cgagtcccgaaaccgtttcgaatgcgccgatgcataggtgcagtcgatagaggaatgaag
          <<<MORC4_r3<<< 2407 to 2424
 801 C  P  P  S  L  G  A  S  *
2401 tgtcctccctcacttggagcttcgtga
```

MORC4 XM_006724691.2

```
   1 M  L  L  Y  R  G  A  P  A  G  P  G  A  P  G  C  G  L  A  R
   1 atgctcctgtaccgaggggccccccgccggccctggcgcgccgggctgcgggctggcccgg
   1          10        20        30        40        50
   1 tacgaggacatggctccccgggggcggccgggaccgcgcggcccgacgcccgaccgggcc
                        >>>MORC4_f1>>> 82 to 99
  21 P  G  G  G  P  Q  A  F  G  I  R  L  S  T  M  S  P  R  Y  L
  61 cccggcggcggcccgcaggccttcgggatccgcctgagcacgatgagccccccgctacctc
  61          70        80        90        100       110
  61 gggccgccgccgggcgtccggaagccctaggcggactcgtgctactcgggggcgatggag
  41 Q  S  N  S  S  S  H  T  R  P  F  S  A  I  A  E  L  L  D  N
 121 cagagcaactccagcagccacacgcgacccttcagtgccatcgcggagctgctagataat
 121          130       140       150       160       170
 121 gtctcgttgaggtcgtcggtgtgcgctgggaagtcacggtagcgcctcgacgatctatta
  61 A  V  D  P  D  V  S  A  R  T  V  F  I  D  V  E  E  V  K  N
```

```
181 gctgtagatccagatgtatctgccaggacggtctttatagatgttgaggaggtcaagaat
181          190       200       210       220       230
181 cgacatctaggtctacatagacggtcctgccagaaatatctacaactcctccagttctta
 81 K   S   C   L   T   F   T   D   D   G   C   G   M   T   P   H   K   L   H   R
241 aaatcttgtttgacctttaccgatgatggatgtgggatgacacctcataaactacaccga
241          250       260       270       280       290
241 tttagaacaaactggaaatggctactacctacaccctactgtggagtatttgatgtggct
                     <<<MORC4_r1<<< 319 to 336
101 M   L   R   N   K   N   G   K   S   E   L   D   F   D   T   D   Q   Y   D   I
301 atgctcagaaataaaaatggaaaatctgagttggactttgatacagatcaatatgacatc
301          310       320       330       340       350
301 tacgagtctttatttttacctttttagactcaacctgaaactatgtctagttatactgtag
121 L   V   S   D   F   D   T   E   E   K   M   T   G   G   V   T   S   E   L   P
361 ctggtatcagactttgacacagaagaaaaaatgactggcggtgttacctctgagctacca
361          370       380       390       400       410
361 gaccatagtctgaaactgtgtcttctttttttactgaccgccacaatggagactcgatggt
141 E   T   E   Y   S   L   R   A   F   C   G   I   L   Y   M   K   P   R   M   K
421 gaaacagaatattctttaagggcattttgtggtattctatacatgaagccacgcatgaaa
421          430       440       450       460       470
421 ctttgtcttataagaaattcccgtaaaacaccataagatatgtacttcggtgcgtacttt
161 I   F   L   R   Q   K   K   V   T   T   Q   M   I   A   K   S   L   A   N   V
481 atttttctgcgtcaaaagaaggtgactacccagatgattgccaagagcctggccaatgta
481          490       500       510       520       530
481 taaaaagacgcagttttcttccactgatgggtctactaacggttctcggaccggttacat
181 E   Y   D   T   Y   K   P   T   F   T   N   K   Q   V   R   I   T   F   G   F
541 gaatatgatacatataaacctaccttcacaaataagcaggtgagaatcacctttgggttc
541          550       560       570       580       590
541 cttatactatgtatatttggatggaagtgtttattcgtccactcttagtggaaacccaag
201 S   C   K   N   S   N   Q   F   G   I   M   M   Y   H   N   N   R   L   I   K
601 tcttgcaagaatagtaaccagtttggaataatgatgtatcataacaaccgactcataaaa
601          610       620       630       640       650
601 agaacgttcttatcattggtcaaaccttattactacatagtattgttggctgagtatttt
221 S   F   E   K   V   G   C   Q   V   K   P   T   R   G   E   G   V   G   V   I
661 tcttttgagaaggtggggtgccaggtgaagccaactcgtggagaaggtgtaggagtaatt
661          670       680       690       700       710
661 agaaaactcttccaccccacggtccacttcggttgagcacctcttccacatcctcattaa
241 G   V   I   E   C   N   F   L   K   P   A   Y   N   K   Q   D   F   E   Y   T
721 ggagtcattgagtgcaatttcctaaaacctgcctacaacaaacaagactttgagtatacc
721          730       740       750       760       770
721 cctcagtaactcacgttaaaggattttggacggatgttgtttgttctgaaactcatatgg
261 K   E   Y   R   L   T   I   N   A   L   A   Q   K   L   N   A   Y   W   K   E
781 aaggagtaccggctaacaataaatgcccttgcccagaagctcaatgcttactggaaggaa
781          790       800       810       820       830
781 ttcctcatggccgattgttatttacgggaacgggtcttcgagttacgaatgaccttcctt
281 K   T   S   Q   D   N   F   E   T   S   T   V   A   R   P   I   P   K   V   P
841 aaaacatctcaagataattttgagacctcaactgtagccaggccaataccgaaggttcct
841          850       860       870       880       890
841 ttttgtagagttctattaaaactctggagttgacatcggtccggttatggcttccaagga
301 D   Q   T   W   V   Q   C   D   E   C   L   K   W   R   K   L   P   G   K   I
901 gaccagacatgggttcagtgtgatgagtgtcttaaatggagaaagcttcctgggaagatt
901          910       920       930       940       950
901 ctggtctgtacccaagtcacactactcacagaatttacctctttcgaaggacccttctaa
321 D   P   S   M   L   P   A   R   W   F   C   Y   Y   N   S   H   P   K   Y   R
961 gatccatccatgttacctgcaagatggttttgttattataattcccatccaaagtacagg
961          970       980       990       1000      1010
961 ctaggtaggtacaatggacgttctaccaaaacaataatattaagggtaggtttcatgtcc
341 R   C   S   V   P   E   E   Q   E   L   T   D   E   D   L   C   L   S   K   A
1021 agatgctctgttccagaggaacaagaactcactgatgaagacctgtgcttgagcaaagct
1021         1030      1040      1050      1060      1070
1021 tctacgagacaaggtctccttgttcttgagtgactacttctggacacgaactcgtttcga
361 K   K   Q   E   Q   T   V   E   E   K   K   K   M   P   M   E   N   E   N   H
```

```
1081 aagaaacaagaacaaactgttgaggagaagaagaagatgcctatggaaaatgagaaccac
1081          1090      1100      1110      1120      1130
1081 ttctttgttcttgtttgacaactcctcttcttcttctacggatacctttttactcttggtg
 381 Q  V  F  S  N  P  P  K  I  L  T  V  Q  E  M  A  G  L  N  N
1141 caggtattcagtaatccaccaaagatccttactgttcaagaaatggctggattgaataac
1141          1150      1160      1170      1180      1190
1141 gtccataagtcattaggtggtttctaggaatgacaagttctttaccgacctaacttattg
 401 K  T  I  G  Y  E  G  I  H  S  P  S  V  L  P  S  G  G  E  E
1201 aagacaattggatatgagggaattcatagccctagtgtgcttccttctggtggagaagaa
1201          1210      1220      1230      1240      1250
1201 ttctgttaacctatactccccttaagtatcgggatcacacgaaggaagaccacctcttctt
 421 S  R  S  P  S  L  Q  L  K  P  L  D  S  S  V  L  Q  F  S  S
1261 agcagatcaccatctcttcaacttaagcctctggattccagtgttttacagttttccagt
1261          1270      1280      1290      1300      1310
1261 tcgtctagtggtagagaagttgaattcggagacctaaggtcacaaaatgtcaaaaggtca
 441 K  Y  K  W  I  L  G  E  E  P  V  E  K  R  R  R  L  Q  N  E
1321 aagtacaaatggatcctaggtgaagaaccggtggagaaacgaagaaggctccagaatgag
1321          1330      1340      1350      1360      1370
1321 ttcatgtttacctaggatccacttcttggccacctctttgcttcttccgaggtcttactc
 461 M  T  T  P  S  L  D  Y  S  M  P  A  P  Y  R  R  V  E  A  P
1381 atgacaacaccttctctagattattccatgcctgctccttacaggagggtagaagcacct
1381          1390      1400      1410      1420      1430
1381 tactgttgtggaagagatctaataaggtacggacgaggaatgtcctcccatcttcgtgga
 481 V  A  Y  P  E  G  E  N  S  H  D  K  S  S  S  E  R  S  T  P
1441 gttgcctacccagaaggggagaacagccatgataaatcgagttctgagagaagtacacca
1441          1450      1460      1470      1480      1490
1441 caacggatgggtcttcccctcttgtcggtactatttagctcaagactctcttcatgtggt
 501 P  Y  L  F  P  E  Y  P  E  A  S  K  N  T  G  Q  N  R  E  V
1501 ccataccttttcccagaatacccagaagcaagcaagaatacaggtcagaatagggaggtt
1501          1510      1520      1530      1540      1550
1501 ggtatggaaaagggtcttatgggtcttcgttcgttcttatgtccagtcttatccctccaa
 521 S  I  L  Y  P  G  A  K  D  Q  R  Q  G  S  L  L  P  E  E  L
1561 tcaattctgtatccaggggccaaagaccaacgccaggggtccctgcttcctgaagaatta
1561          1570      1580      1590      1600      1610
1561 agttaagacataggtccccggtttctggttgcggtccccagggacgaaggacttcttaat
 541 E  D  Q  M  P  R  L  V  A  E  E  S  N  R  G  S  T  T  I  N
1621 gaagatcagatgccaagattggtggcagaagaatctaacagaggtagcacaaccataaac
1621          1630      1640      1650      1660      1670
1621 cttctagtctacggttctaaccaccgtcttcttagattgtctccatcgtgttggtatttg
```

>>>MORC4_f2>>> 1737 to 1754

```
 561 K  E  E  V  N  K  G  P  F  V  A  V  V  G  V  A  K  G  V  R
1681 aaagaagaagtcaacaagggacctttgtagctgttgtgggtgttgccaaaggtgttaga
1681          1690      1700      1710      1720      1730
1681 tttcttcttcagttgttccctggaaaacatcgacaacacccacaacggtttccacaatct
 581 D  S  G  A  P  I  Q  L  I  P  F  N  R  E  E  L  A  E  R  R
1741 gattcaggagctcccattcagctgatccctttaacagagaggagcttgctgagagacga
1741          1750      1760      1770      1780      1790
1741 ctaagtcctcgagggtaagtcgactagggaaaattgtctctcctcgaacgactctctgct
 601 K  A  V  E  S  W  N  P  V  P  Y  S  V  A  S  A  A  I  P  A
1801 aaagcagttgaatcctggaacccagtgccttattctgtggcctctgctgcaatccctgct
1801          1810      1820      1830      1840      1850
1801 tttcgtcaacttaggaccttgggtcacggaataagacaccggagacgacgttagggacga
 621 A  A  I  G  E  K  A  R  G  Y  E  E  S  E  G  H  N  T  P  K
1861 gcagccattggggagaaagcaagaggctatgaggagagcgaaggtcataatacaccaaag
1861          1870      1880      1890      1900      1910
1861 cgtcggtaacccctctttcgttctccgatactcctctcgcttccagtattatgtggtttc
 641 L  K  N  Q  R  E  L  E  E  L  K  R  T  T  E  K  L  E  R  V
1921 ttgaagaaccagagagagctggaagaattgaagagaaccacagaaaaattggaacgtgtt
1921          1930      1940      1950      1960      1970
1921 aacttcttggtctctctcgaccttcttaacttctcttggtgtctttttaaccttgcacaa
```

```
 661 L  A  E  R  N  L  F  Q  Q  K  V  E  E  L  E  Q  E  R  N  H
1981 ttggctgaaaggaatttgttccagcaaaaggtggaggagctggaacaggagaggaatcac
1981          1990      2000      2010      2020      2030
1981 aaccgactttccttaaacaaggtcgttttccacctcctcgaccttgtcctctccttagtg
 681 W  Q  S  E  F  K  K  V  Q  H  E  L  V  I  Y  S  T  Q  E  A
2041 tggcagtctgaattcaagaaagtccaacatgaattggtgatctacagtacccaggaggcg
2041          2050      2060      2070      2080      2090
2041 accgtcagacttaagttctttcaggttgtacttaaccactagatgtcatgggtcctccgc
 701 E  G  L  Y  W  S  K  K  H  M  G  Y  R  Q  A  E  F  Q  I  L
2101 gaaggcttgtactggagcaagaaacacatgggttatcgccaagctgaattccagattctg
2101          2110      2120      2130      2140      2150
2101 cttccgaacatgacctcgttctttgtgtacccaatagcggttcgacttaaggtctaagac

<<<MORC4_r4<<< 2217 to 2234
 721 K  A  E  L  E  R  T  K  E  E  K  Q  E  L  K  E  K  L  K  E
2161 aaagctgagctggaaagaaccaaagaggaaaagcaagagttaaaagagaaactgaaggaa
2161          2170      2180      2190      2200      2210
2161 tttcgactcgacctttcttggtttctccttttcgttctcaattttctctttgacttcctt
 741 T  E  T  H  L  E  M  L  Q  K  A  Q  V  S  Y  R  T  P  E  G
2221 acagagacacacctggaaatgctgcagaaggctcaggtctcctaccggaccccagaggga
2221          2230      2240      2250      2260      2270
2221 tgtctctgtgtggacctttacgacgtcttccgagtccagaggatggcctggggtctccct
 761 D  D  L  E  R  A  L  A  K  L  T  R  L  R  I  H  V  S  Y  L
2281 gatgacctagaaagggctttggcaaagcttacgcggctacgtatccacgtcagctatctc
2281          2290      2300      2310      2320      2330
2281 ctactggatctttcccgaaaccgtttcgaatgcgccgatgcataggtgcagtcgatagag
                   <<<MORC4_r3<<< 2355 to 2372
 781 L  T  S  V  L  P  H  L  E  L  R  E  I  G  Y  D  S  E  Q  V
2341 cttacttctgtcctccctcacttggagcttcgtgagatcgggtatgactcagaacaagtg
2341          2350      2360      2370      2380      2390
2341 gaatgaagacaggagggagtgaacctcgaagcactctagcccatactgagtcttgttcac
                                 <<<MORC4_r2<<< 2428 to 2445
 801 D  G  I  L  Y  T  V  L  E  A  N  H  I  L  D  *
2401 gatgggatcctgtacacggtgctggaggcaaatcacatactggattga
```

MORC4 XM_005262190.3

```
   1 M  T  P  H  K  L  H  R  M  L  S  F  G  F  T  D  K  V  I  K
   1 atgacacctcataaactacaccgaatgctcagctttggctttacagataaagtaataaag
   1          10        20        30        40        50
   1 tactgtggagtatttgatgtggcttacgagtcgaaaccgaaatgtctatttcattatttc
  21 K  S  Q  C  P  I  G  V  F  G  N  G  F  K  S  G  S  M  R  L
  61 aagagccagtgtcccattggggtctttggtaatggtttcaagtcaggctccatgcggcta
  61          70        80        90        100       110
  61 ttctcggtcacagggtaaccccagaaaccattaccaaagttcagtccgaggtacgccgat
  41 G  K  D  A  L  V  F  T  K  N  G  G  T  L  T  V  G  L  L  S
 121 ggaaaggacgcccttgtcttcaccaagaatgggggtactctcactgttggacttctatca
 121          130       140       150       160       170
 121 cctttcctgcgggaacagaagtggttcttacccccatgagagtgacaacctgaagatagt
  61 Q  T  Y  L  E  C  V  Q  A  Q  A  V  I  V  P  I  V  P  F  N
 181 cagacctatctggaatgtgtccaggcccaggcagttattgtaccaattgttccattcaac
 181          190       200       210       220       230
 181 gtctggatagaccttacacaggtccgggtccgtcaataacatggttaacaaggtaagttg
  81 Q  Q  N  K  K  M  I  I  T  E  D  S  L  P  S  L  E  A  I  L
 241 cagcaaaacaaaaaaatgattattaccgaggattcattgcccagcctagaagccatcttg
 241          250       260       270       280       290
 241 gtcgttttgttttttactaataatggctcctaagtaacgggtcggatcttcggtagaac
 101 N  Y  S  I  F  N  R  E  N  D  L  L  A  Q  F  D  A  I  P  G
 301 aactattccattttcaaccgtgaaaatgacctgctggcccagtttgatgccatcccaggc
```

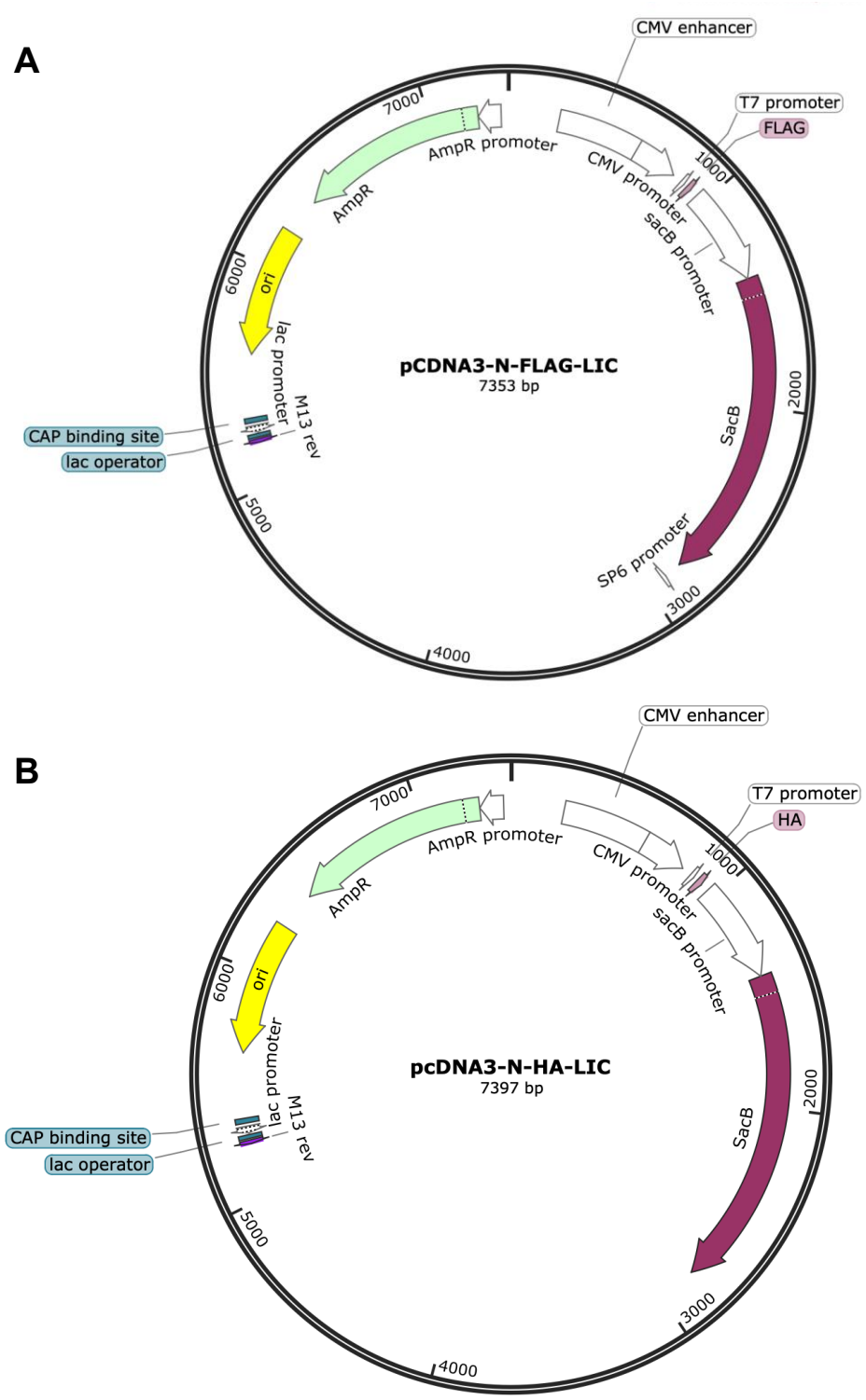
```
 301          310       320       330       340       350
 301 ttgataaggtaaaagttggcacttttactggacgaccgggtcaaactacggtagggtccg
                                                      <<<MORC4_r1<<< 409
to 426
 121 K  K  G  T  R  V  L  I  W  N  I  R  R  N  K  N  G  K  S  E
 361 aaaaaaggcactcgtgttctcatttggaacatccgcagaaataaaaatggaaaatctgag
 361          370       380       390       400       410
 361 ttttttccgtgagcacaagagtaaaccttgtaggcgtctttatttttaccttttagactc
 141 L  D  F  D  T  D  Q  Y  D  I  L  V  S  D  F  D  T  E  E  K
 421 ttggactttgatacagatcaatatgacatcctggtatcagactttgacacagaagaaaaa
 421          430       440       450       460       470
 421 aacctgaaactatgtctagttatactgtaggaccatagtctgaaactgtgtcttcttttt
 161 M  T  G  G  V  T  S  E  L  P  E  T  E  Y  S  L  R  A  F  C
 481 atgactggcggtgttacctctgagctaccagaaacagaatattctttaagggcattttgt
 481          490       500       510       520       530
 481 tactgaccgccacaatggagactcgatggtctttgtcttataagaaattcccgtaaaaca
 181 G  I  L  Y  M  K  P  R  M  K  I  F  L  R  Q  K  K  V  T  T
 541 ggtattctatacatgaagccacgcatgaaaatttttctgcgtcaaaagaaggtgactacc
 541          550       560       570       580       590
 541 ccataagatatgtacttcggtgcgtacttttaaaaagacgcagttttcttccactgatgg
 201 Q  M  I  A  K  S  L  A  N  V  E  Y  D  T  Y  K  P  T  F  T
 601 cagatgattgccaagagcctggccaatgtagaatatgatacatataaacctaccttcaca
 601          610       620       630       640       650
 601 gtctactaacggttctcggaccggttacatcttatactatgtatatttggatggaagtgt
 221 N  K  Q  V  R  I  T  F  G  F  S  C  K  N  S  N  Q  F  G  I
 661 aataagcaggtgagaatcacctttgggttctcttgcaagaatagtaaccagtttggaata
 661          670       680       690       700       710
 661 ttattcgtccactcttagtggaaacccaagagaacgttcttatcattggtcaaaccttat
 241 M  M  Y  H  N  N  R  L  I  K  S  F  E  K  V  G  C  Q  V  K
 721 atgatgtatcataacaaccgactcataaaatcttttgagaaggtggggtgccaggtgaag
 721          730       740       750       760       770
 721 tactacatagtattgttggctgagtattttagaaaactcttccaccccacggtccacttc
 261 P  T  R  G  E  G  V  G  V  I  G  V  I  E  C  N  F  L  K  P
 781 ccaactcgtggagaaggtgtaggagtaattggagtcattgagtgcaatttcctaaaacct
 781          790       800       810       820       830
 781 ggttgagcacctcttccacatcctcattaacctcagtaactcacgttaaaggattttgga
 281 A  Y  N  K  Q  D  F  E  Y  T  K  E  Y  R  L  T  I  N  A  L
 841 gcctacaacaaacaagactttgagtataccaaggagtaccggctaacaataaatgccctt
 841          850       860       870       880       890
 841 cggatgttgtttgttctgaaactcatatggttcctcatggccgattgttatttacgggaa
 301 A  Q  K  L  N  A  Y  W  K  E  K  T  S  Q  D  N  F  E  T  S
 901 gcccagaagctcaatgcttactggaaggaaaaaacatctcaagataatttgagacctca
 901          910       920       930       940       950
 901 cgggtcttcgagttacgaatgaccttccttttttgtagagttctattaaaactctggagt
 321 T  V  A  R  P  I  P  K  V  P  D  Q  T  W  V  Q  C  D  E  C
 961 actgtagccaggccaataccgaaggttcctgaccagacatgggttcagtgtgatgagtgt
 961          970       980       990       1000      1010
 961 tgacatcggtccggttatggcttccaaggactggtctgtacccaagtcacactactcaca
 341 L  K  W  R  K  L  P  G  K  I  D  P  S  M  L  P  A  R  W  F
1021 cttaaatggagaaagcttcctgggaagattgatccatccatgttacctgcaagatggttt
1021          1030      1040      1050      1060      1070
1021 gaatttacctctttcgaaggacccttctaactaggtaggtacaatggacgttctaccaaa
 361 C  Y  Y  N  S  H  P  K  Y  R  R  C  S  V  P  E  E  Q  E  L
1081 tgttattataattcccatccaaagtacaggagatgctctgttccagaggaacaagaactc
1081          1090      1100      1110      1120      1130
1081 acaataatattaagggtaggtttcatgtcctctacgagacaaggtctccttgttcttgag
 381 T  D  E  D  L  C  L  S  K  A  K  K  Q  E  Q  T  V  E  E  K
1141 actgatgaagacctgtgcttgagcaaagctaagaaacaagaacaaactgttgaggagaag
1141          1150      1160      1170      1180      1190
1141 tgactacttctggacacgaactcgtttcgattctttgttcttgtttgacaactcctcttc
 401 K  K  M  P  M  E  N  E  N  H  Q  V  F  S  N  P  P  K  I  L
```

```
1201 aagaagatgcctatggaaaatgagaaccaccaggtattcagtaatccaccaaagatccttt
1201          1210      1220      1230      1240      1250
1201 ttcttctacggatacctttttactcttggtggtccataagtcattaggtggtttctaggaa
 421 T  V  Q  E  M  A  G  L  N  N  K  T  I  G  Y  E  G  I  H  S
1261 actgttcaagaaatggctggattgaataacaagacaattggatatgagggaattcatagc
1261          1270      1280      1290      1300      1310
1261 tgacaagttctttaccgacctaacttattgttctgttaacctatactcccttaagtatcg
 441 P  S  V  L  P  S  G  G  E  E  S  R  S  P  S  L  Q  L  K  P
1321 cctagtgtgcttccttctggtggagaagaaagcagatcaccatctcttcaacttaagcct
1321          1330      1340      1350      1360      1370
1321 ggatcacacgaaggaagaccacctcttctttcgtctagtggtagagaagttgaattcgga
 461 L  D  S  S  V  L  Q  F  S  S  K  Y  K  W  I  L  G  E  E  P
1381 ctggattccagtgttttacagttttccagtaagtacaaatggatcctaggtgaagaaccg
1381          1390      1400      1410      1420      1430
1381 gacctaaggtcacaaaatgtcaaaaggtcattcatgtttacctaggatccacttcttggc
 481 V  E  K  R  R  R  L  Q  N  E  M  T  T  P  S  L  D  Y  S  M
1441 gtggagaaacgaagaaggctccagaatgagatgacaacaccttctctagattattccatg
1441          1450      1460      1470      1480      1490
1441 cacctctttgcttcttccgaggtcttactctactgttgtggaagagatctaataaggtac
 501 P  A  P  Y  R  R  V  E  A  P  V  A  Y  P  E  G  E  N  S  H
1501 cctgctccttacaggagggtagaagcacctgttgcctacccagaaggggagaacagccat
1501          1510      1520      1530      1540      1550
1501 ggacgaggaatgtcctcccatcttcgtggacaacggatgggtcttcccctcttgtcggta
 521 D  K  S  S  S  E  R  S  T  P  P  Y  L  F  P  E  Y  P  E  A
1561 gataaatcgagttctgagagaagtacaccaccatacctttcccagaatacccagaagca
1561          1570      1580      1590      1600      1610
1561 ctatttagctcaagactctcttcatgtggtggtatggaaaagggtcttatgggtcttcgt
 541 S  K  N  T  G  Q  N  R  E  V  S  I  L  Y  P  G  A  K  D  Q
1621 agcaagaatacaggtcagaatagggaggtttcaattctgtatccaggggccaaagaccaa
1621          1630      1640      1650      1660      1670
1621 tcgttcttatgtccagtcttatccctccaaagttaagacataggtccccggtttctggtt
 561 R  Q  G  S  L  L  P  E  E  L  E  D  Q  M  P  R  L  V  A  E
1681 cgccaggggtccctgcttcctgaagaattagaagatcagatgccaagattggtggcagaa
1681          1690      1700      1710      1720      1730
1681 gcggtccccagggacgaaggacttcttaatcttctagtctacggttctaaccaccgtctt
 581 E  S  N  R  G  S  T  T  I  N  K  E  E  V  N  K  G  P  F  V
1741 gaatctaacagaggtagcacaaccataaacaaagaagaagtcaacaagggacctttgta
1741          1750      1760      1770      1780      1790
1741 cttagattgtctccatcgtgttggtatttgtttcttcttcagttgttccctggaaaacat
                                >>>MORC4_f2>>> 1827 to 1844
 601 A  V  V  G  V  A  K  G  V  R  D  S  G  A  P  I  Q  L  I  P
1801 gctgttgtgggtgttgccaaaggtgttagagattcaggagctcccattcagctgatccct
1801          1810      1820      1830      1840      1850
1801 cgacaacacccacaacggtttccacaatctctaagtcctcgagggtaagtcgactaggga
 621 F  N  R  E  E  L  A  E  R  R  K  A  V  E  S  W  N  P  V  P
1861 tttaacagagaggagcttgctgagagacgaaaagcagttgaatcctggaacccagtgcct
1861          1870      1880      1890      1900      1910
1861 aaattgtctctcctcgaacgactctctgcttttcgtcaacttaggaccttgggtcacgga
 641 Y  S  V  A  S  A  A  I  P  A  A  A  I  G  E  K  A  R  G  Y
1921 tattctgtggcctctgctgcaatccctgctgcagccattggggagaaagcaagaggctat
1921          1930      1940      1950      1960      1970
1921 ataagacaccggagacgacgttagggacgacgtcggtaacccctctttcgttctccgata
 661 E  E  S  E  G  H  N  T  P  K  L  K  N  Q  R  E  L  E  E  L
1981 gaggagagcgaaggtcataatacaccaaagttgaagaaccagagagagctggaagaattg
1981          1990      2000      2010      2020      2030
1981 ctcctctcgcttccagtattatgtggtttcaacttcttggtctctctcgaccttcttaac
 681 K  R  T  T  E  K  L  E  R  V  L  A  E  R  N  L  F  Q  Q  K
2041 aagagaaccacagaaaaattggaacgtgttttggctgaaaggaatttgttccagcaaaag
2041          2050      2060      2070      2080      2090
2041 ttctcttggtgtctttttaaccttgcacaaaaccgactttccttaaacaaggtcgttttc
 701 V  E  E  L  E  Q  E  R  N  H  W  Q  S  E  F  K  K  V  Q  H
```

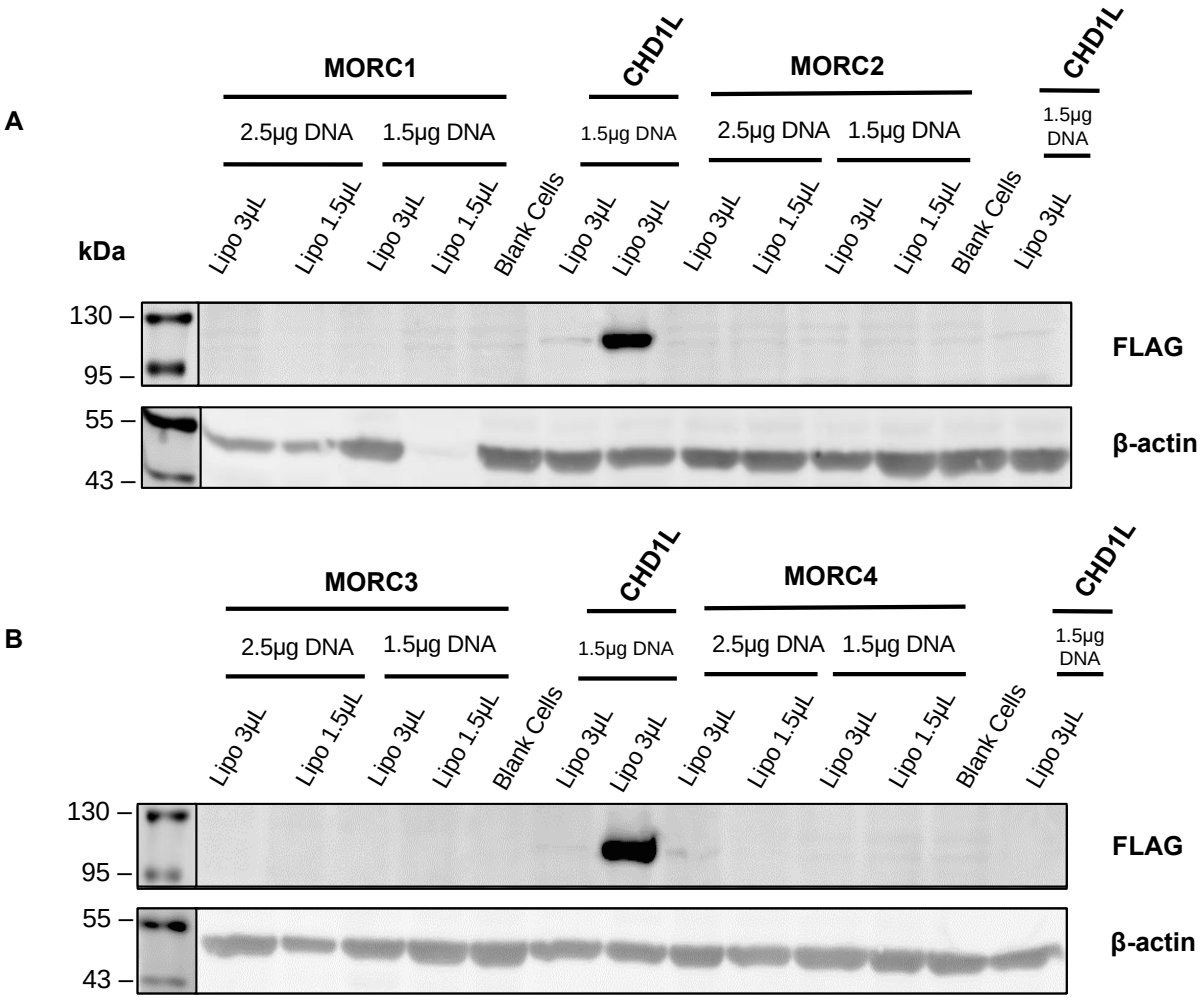
```
2101 gtggaggagctggaacaggagaggaatcactggcagtctgaattcaagaaagtccaacat
2101          2110      2120      2130      2140      2150
2101 cacctcctcgaccttgtcctctccttagtgaccgtcagacttaagttctttcaggttgta
 721 E   L   V   I   Y   S   T   Q   E   A   E   G   L   Y   W   S   K   K   H   M
2161 gaattggtgatctacagtacccaggaggcggaaggcttgtactggagcaagaaacacatg
2161          2170      2180      2190      2200      2210
2161 cttaaccactagatgtcatgggtcctccgccttccgaacatgacctcgttctttgtgtac
 741 G   Y   R   Q   A   E   F   Q   I   L   K   A   E   L   E   R   T   K   E   E
2221 ggttatcgccaagctgaattccagattctgaaagctgagctggaaagaaccaaagaggaa
2221          2230      2240      2250      2260      2270
2221 ccaatagcggttcgacttaaggtctaagactttcgactcgacctttcttggtttctcctt
                                <<<MORC4_r4<<< 2307 to 2324
 761 K   Q   E   L   K   E   K   L   K   E   T   E   T   H   L   E   M   L   Q   K
2281 aagcaagagttaaaagagaaactgaaggaaacagagacacacctggaaatgctgcagaag
2281          2290      2300      2310      2320      2330
2281 ttcgttctcaattttctctttgacttcctttgtctctgtgtggacctttacgacgtcttc
 781 A   Q   V   S   Y   R   T   P   E   G   D   D   L   E   R   A   L   A   K   L
2341 gctcaggtctcctaccggaccccagagggagatgacctagaaagggctttggcaaagctt
2341          2350      2360      2370      2380      2390
2341 cgagtccagaggatggcctggggtctccctctactggatctttcccgaaaccgtttcgaa
                                                      <<<MORC4_r3<<< 2445 to
2462
 801 T   R   L   R   I   H   V   S   Y   L   L   T   S   V   L   P   H   L   E   L
2401 acgcggctacgtatccacgtcagctatctccttacttctgtcctccctcacttggagctt
2401          2410      2420      2430      2440      2450
2401 tgcgccgatgcataggtgcagtcgatagaggaatgaagacaggagggagtgaacctcgaa

<<<MORC4_r2<<< 2518 to 2535
 821 R   E   I   G   Y   D   S   E   Q   V   D   G   I   L   Y   T   V   L   E   A
2461 cgtgagatcgggtatgactcagaacaagtggatgggatcctgtacacggtgctggaggca
2461          2470      2480      2490      2500      2510
2461 gcactctagcccatactgagtcttgttcacctaccctaggacatgtgccacgacctccgt
 841 N   H   I   L   D   *
2521 aatcacatactggattga
```

Appendix 8.21: **Concentrations and quality assessment of RNA preparation of cell lines utilised throughout this study.**

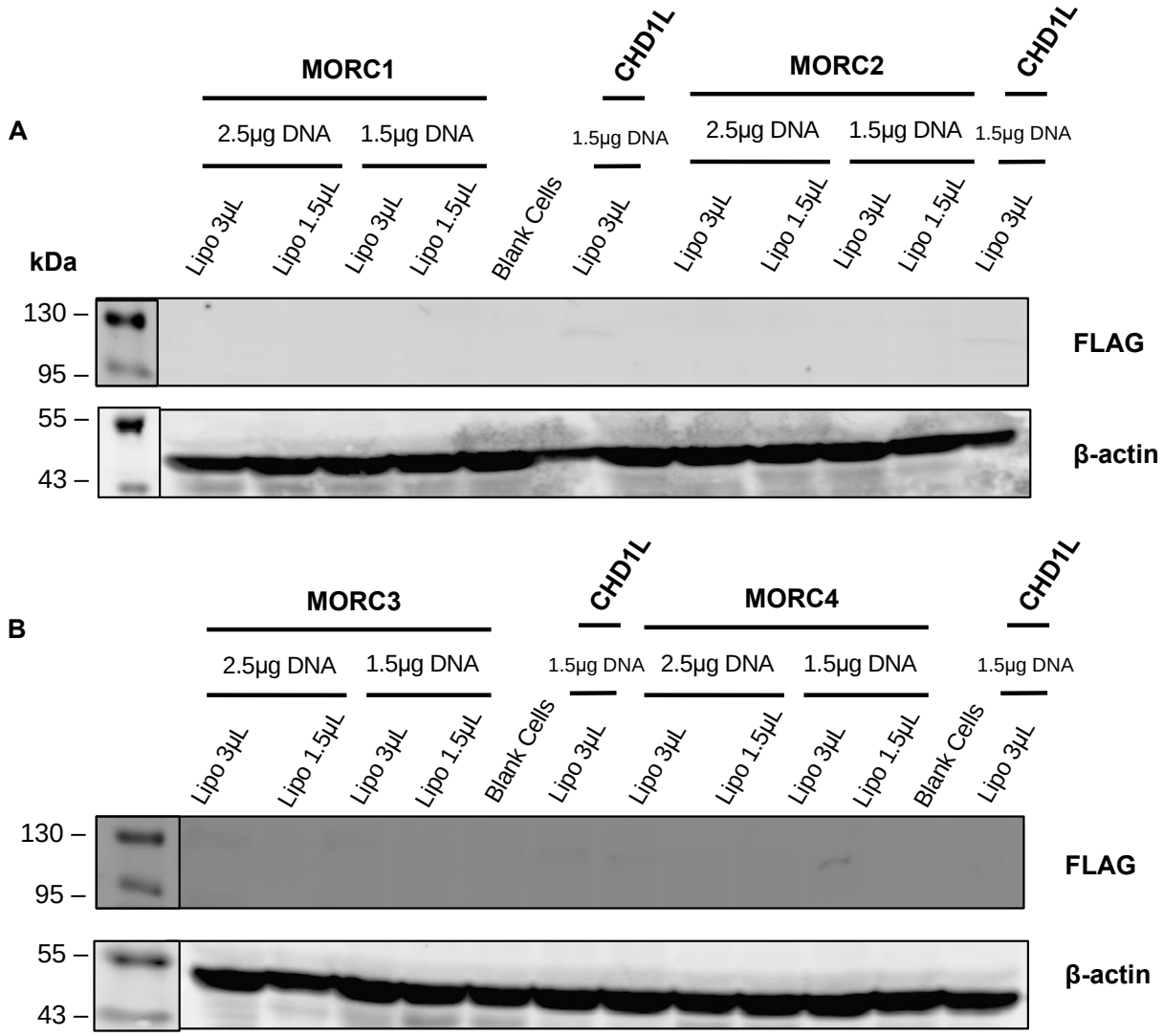
Cell line	Disease	RNA (ng/µL)		260/280 (nm)	
		1	2	1	2
OCI-Ly3	ABC DLBCL	687.6	696.8	2.02	2.02
RIVA	ABC DLBCL	665.8	676.1	2.03	2.04
SU-DHL-2	ABC DLBCL	1170.8	1177.9	2.02	2.02
HBL-1	ABC DLBCL	563.9	564.4	2.03	2.03
HLY-1	ABC DLBCL	1189.2	1211.6	2.02	2.03
SUD-HL6	GCB DLBCL	830.5	843.3	2.03	2.03
SUD-HL10	GCB DLBCL	795.9	799.2	2.04	2.04
DB	GCB DLBCL	610.5	609.1	2.01	2.01
MIEU	GCB DLBCL	1110.0	1103.3	2.03	2.03
LIB	DLBCL	639.6	640.2	2.03	2.03
Daudi	Burkitts	546.7	514.2	2.00	1.93
Raji	Burkitts	989.9	957.6	2.04	2.04
NCI H929	Myeloma	804.7	967.3	2.06	2.07
JJN.3	Myeloma	1114.7	1092.4	2.04	2.03
RPMI8226	Myeloma	879.8	842.7	2.02	2.03
Jurkat	T-ALL	600.6	599.7	2.02	2.00
DU528	T-ALL	560.3	559.3	2.00	2.02
RPMI8402	T-ALL	557.5	532.9	2.01	2.03
Hut78	CTCL-SS	916.2	916.3	2.04	2.04
HCT29	Colorectal adenocarcinoma	918.5	877.4	2.05	2.06
HCT116+/+	Colorectal adenocarcinoma	1423.2	1505.6	2.04	2.04
SW48	Colorectal adenocarcinoma	734.4	732.1	2.05	2.04
HeLa	Cervical carcinoma	891.8	933.0	2.08	2.03
U2OS	Osteosarcoma	415.9	424.1	2.02	2.01
HEK293FT	Transformed human embryonic kidney	1162.2	1169.8	2.03	2.02
SK-N-BE(2c)	NB (MYCN+)	954.5	953.4	2.05	2.05
IMR32	NB (MYCN+)	980.0	969.2	2.04	2.03
Kelly	NB (MYCN+)	1063.0	1072.2	2.04	2.03
NB19	NB (MYCN+)	684.1	694.1	2.04	2.04
SH-SY5Y	NB (MYCN-)	587.8	582.4	2.02	2.06
SK-N-SH	NB (MYCN-)	933.7	938.3	2.02	2.02
SK-N-AS	NB (MYCN-)	921.6	936.2	2.07	2.06
SHEP	NB	853.6	858.0	2.05	2.04
PSN1	Pancreatic adenocarcinoma	775.1	777.3	2.05	2.04
Mia-Pa-Ca-2	Pancreatic carcinoma	702.0	704.6	2.06	2.04
MCF7	Breast adenocarcinoma (ER+)	389.5	393.5	2.02	2.02
MDA-MB-231	Breast adenocarcinoma (ER-)	474.8	484.1	2.04	2.04
MDA-MB-468	Breast adenocarcinoma (ER-)	694.6	704.8	2.04	2.05



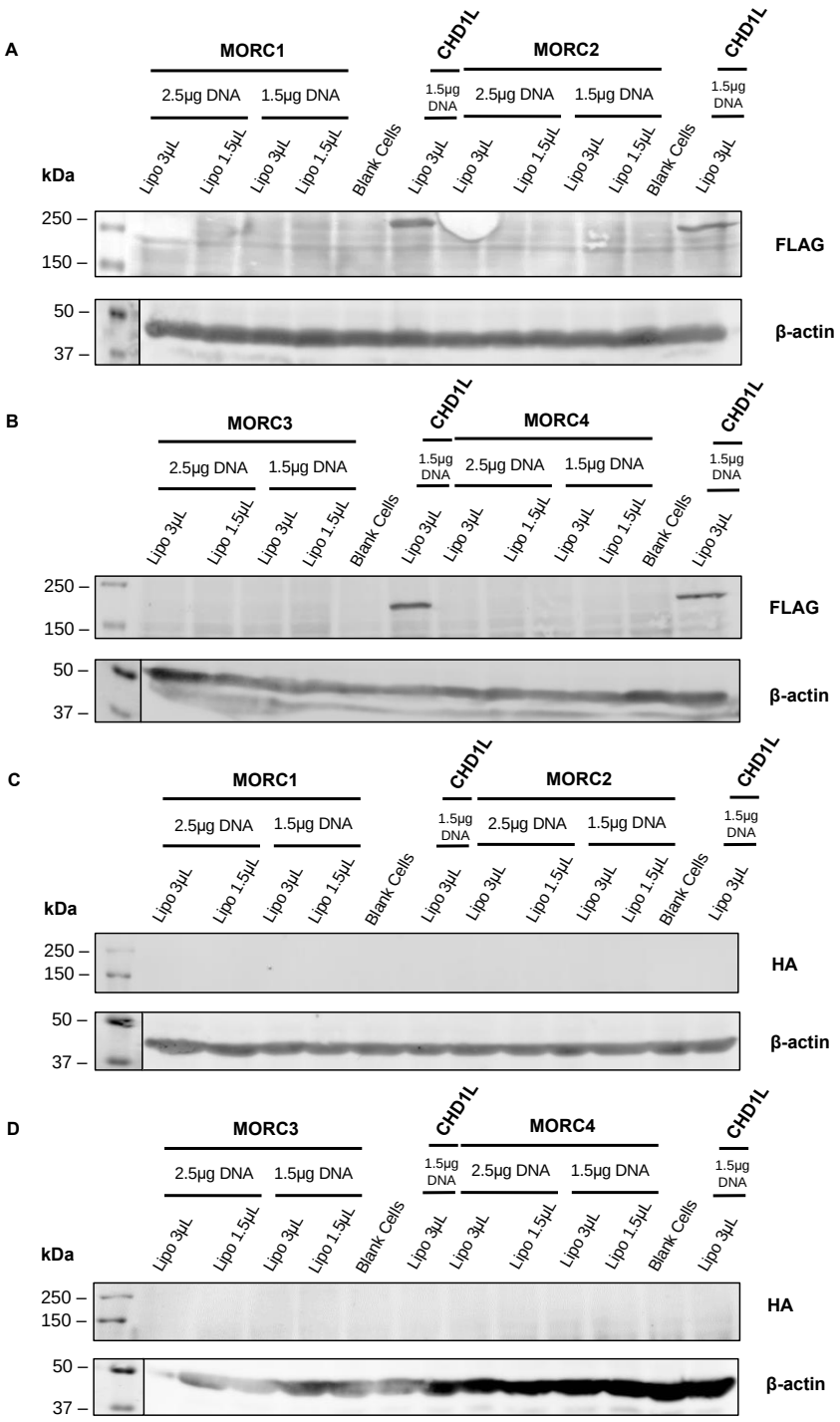
Appendix 8.22: **Vector maps of mammalian expression vectors used in LIC. A** pcDNA3-N-FLAG-LIC **B** pCDNA3-N-HA-LIC. Both mammalian expression vectors were made at the Structural Genomics Consortium, Oxford, United Kingdom. Vector maps Created by author using SnapGene software.



Appendix 8.23: **Fluorescence detection western blotting of transfected full length MORCs into U2OS cell line.** HEK293FT cells were transfected with full length FLAG-tagged MORC DNA, harvested, sonicated and cell lysates were resolved on a 10% SDS-PAGE gel for fluorescence detection western blotting to determine protein expression. Membranes were probed with FLAG and β-actin (1:20,000 dilutions) primary antibodies in 5% milk and TBS-T. **A** FLAG-tagged MORC1 and MORC2. **B** FLAG-tagged MORC3 and MORC4. Membranes were detected using a Li-COR Odyssey imagining system. β-actin loading control and FLAG were imaged using the green channel (800nm) and the marker was imaged using the red channel (700nm).



Appendix 8.24: **Fluorescence detection western blotting of transfected full length MORCs into HeLa cell line.** HEK293FT cells were transfected with full length FLAG-tagged MORC DNA, harvested, sonicated and cell lysates were resolved on a 10% SDS-PAGE gel for fluorescence detection western blotting to determine protein expression. Membranes were probed with FLAG and β-actin (1:20,000 dilutions) primary antibodies in 5% milk and TBS-T. **A** FLAG-tagged MORC1 and MORC2. **B** FLAG-tagged MORC3 and MORC4. Membranes were detected using a Li-COR Odyssey imagining system. β-actin loading control and FLAG were imaged using the green channel (800nm) and the marker was imaged using the red channel (700nm).



Appendix 8.25: **Fluorescence detection western blotting of transfected full length MORCs into COS-1 cell line.** HEK293FT cells were transfected with full length MORC DNA, harvested, sonicated and cell lysates were resolved on a 10% SDS-PAGE gel for fluorescence detection western blotting to determine protein expression. Membranes were probed with FLAG and β-actin (1:20,000 dilutions) primary antibodies in 5% milk and TBS-T. **A** FLAG-tagged MORC1 and MORC2. **B** FLAG-tagged MORC3 and MORC4. **C** HA-tagged MORC1 and MORC2. **D** HA-tagged MORC3 and MORC4. Membranes were detected using a Li-COR Odyssey imagining system. β-actin loading control was imaged using the green channel (800nmand the marker, FLAG and HA were imaged using the red channel (700nm).