Understanding Interactors of the pRB Tumor Suppressor During S/G2/M Phases

A Major Qualifying Project Report

Submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the
Degree of Bachelor of Science
In Biology and Biotechnology

28 April 2022

Submitted by:

Jagruthi Maroju

Advisor:
Amity L. Manning
Worcester Polytechnic Institute, PhD
Biology and Biotechnology

Table of Contents

Table of Figures

Table of Tables

Abstract

The retinoblastoma (RB) protein is a well characterized tumor suppressor, which is found to be mutated in various types of cancer. The best characterized role of pRB relates to its role as a key regulator of the G1 checkpoint. However, recent studies have implicated pRB activity during later phases of the cell cycle, when pRB is present in a hyperphosphorylated state. Recent work has experimentally identified proteins that specifically interact with either hypophosphorylated pRB, as would be found during G1, or hyperphosphorylated pRB, as would be found in G1, S, or G2 stages of the cell cycle. To better understand the functional subcomplexes they may form, I used the web-based Search Tool for the Retrieval of Interacting Genes (STRING) software to map and identify distinct protein subcomplexes within the hypo- and hyperphosphorylated pRB interactomes. Next, to define the nature of the interactions with pRB, I evaluated each protein within these two interactomes for the presence of an LxCxE sequence motif, which has been demonstrated to moderate protein interactions with pRB.

Many of the best characterized pRB-associated proteins that cooperate with pRB in cell cycle control contain an LxCxE motif. Consistent with this, my analysis identified LxCxE binding motifs in 20% of proteins within the hypophosphorylated pRB interactome, with at least one LxCxE domain-containing protein in each identified interacting cluster, suggesting interactions of the remaining 80% of proteins with pRB may be moderated through an LxCxE-containing protein. In contrast, less than 1% of proteins and only a quarter of defined protein clusters within the hyperphosphorylated pRB interactome contained LxCxE motifs. These analyses propose that the majority of proteins that interact with pRB at later stages of the cell cycle, when pRB is hyperphosphorylated, do so in a manner that is independent of pRB's LxCxE binding domain.

Acknowledgements

I want to extend my deepest appreciation and gratitude to Professor Amity Manning for giving me the opportunity to work with her on this MQP research project and for her academic guidance. I want to thank her for the encouragement and motivation she has given me throughout this project.

I am grateful to my academic advisor Mrs. Elizabeth Jacoby for her continued support throughout my tenure at WPI. She was always available and guided me to be on track with my undergraduate course work.

Finally, I am deeply grateful to my parents, Vijaya and Venkat Maroju, and my sister Pragathi Maroju, who have been a constant source of encouragement and inspiration in my studies.

1. Background

1.1 pRB Protein Is An Important Tumor Suppressor.

Retinoblastoma (pRB) protein is a tumor suppressor encoded by the RB1 gene. The protein functions as a key regulator of the G1 checkpoint of the cell cycle (Giacinti & Giordano, 2006), preventing the cell from progressing from G1 into S phase, where the cell synthesizes a complete copy of its DNA (Harbour, 2000). Thus, cells containing mutation or functional inactivation of pRB typically exhibit loss of cell cycle control and uncontrollable cellular entry into S phase (Dick & Rubin, 2013). Since RB1 is a recessive gene, complete functional loss occurs when both alleles of the gene are lost, thus, facilitating the initiation of cancer formation. Mechanisms of mutation can occur through an entire deletion or direct mutation to the protein-coding sequence (Giacinti & Giordano, 2006). Increased susceptibility to total functional loss of RB1 occurs when there is a pre-existing hereditary germline mutation, where one RB1 gene is already mutated within all cells of the body (Goodrich, 2006). However, pRB, is much more commonly misregulated as a result of mutations to upstream regulators that control the activity of pRB.

1.2 pRB Contains Many Cellular Functions.

Although pRB is best known for its regulation of cell cycle progression, the protein forms active complexes with over 300 proteins and has been implicated in dozens of cellular functions.

1.2.1 pRB Acts As A Transcriptional co-Regulator.

pRB's best characterized function is in the modulation of the E2F family of transcriptional regulators. Prior to entering S phase and committing to replicating the genome a

cell must ensure that the conditions are favorable (Harbour & Dean, 2000). pRb prevents a cell's progression to S phase by binding to and inhibiting E2F. The pRB-E2F complex binds to the promoter of cell cycle genes, inhibiting their transcription (Bertoli et al., 2013). Since many of the genes that are regulated by E2F transcription factors contain crucial roles in S phase progression and DNA synthesis, cells where pRB is bound to E2F remain arrested within G1 (Foster et al., 2010).

1.2.2 Regulation of pRB Occurs Through CDK Phosphorylation.

In order to transition through the R point, and into S phase, a cell makes a growth factor-dependent decision to express Cyclin D and E (Bertoli et al., 2013) (Figure 1.1)

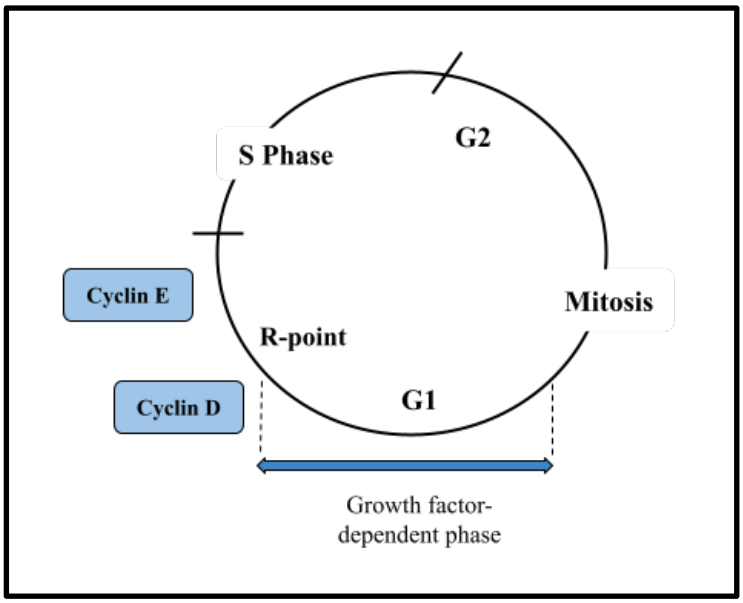

Figure 1.1. *pRB makes a growth factor dependent decision to express Cyclin D and E to allow cell progression from G1 to S phase.*

Cyclin D appears earlier in the G1/S phase transition, while Cyclin E is present in late stages of G1 (Figure 1.1). Cyclins activate cyclin dependent kinases (CDKs) and pRB phosphorylation is regulated through specified CDK phosphorylation. pRB contains 16 distinct CDK phosphorylation sites, which are differently phosphorylated at various stages

of the cell cycle (Knudsen & Wang, 1996). During early stages of G1, pRB remains in a hypophosphorylated state. As the cell moves closer to S phase, pRB undergoes phosphorylation by CDK4-Cyclin D and CDK2-Cyclin E (Gubern et al., 2016). Once pRB is phosphorylated, the protein releases from E2F, allowing transcription of proteins necessary for cell cycle progression to proceed (Komori et al., 2018). In presence of DNA damage, p53 creates a transduction pathway to inhibit progression through G1 by activating p16 and p21, which act as CDK inhibitors, prevent CDKs from phosphorylating pRB (Gubern et al., 2016).

1.2.3 Structural Components of pRB Contain Crucial Binding Domains.

pRB contains a protein binding domain, termed the large pocket, to which many interactors bind.

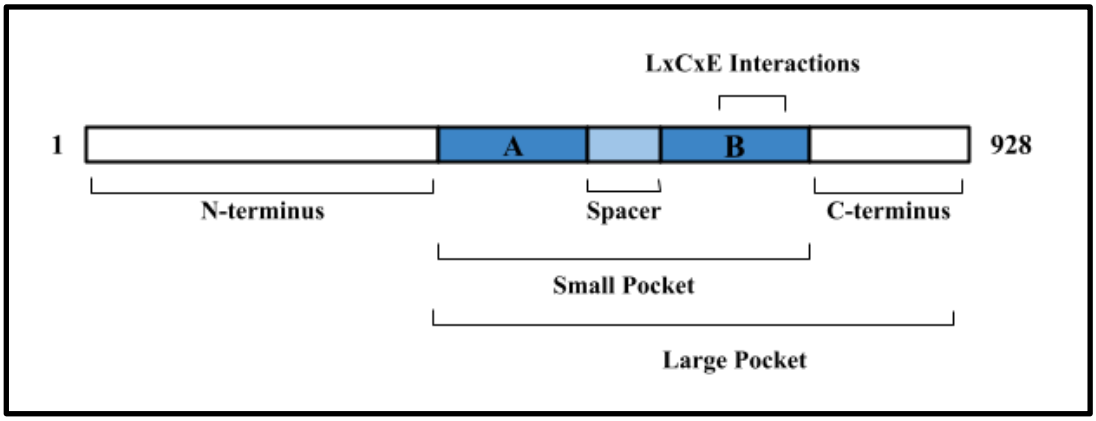

Figure 1.2*. Genomic map of RB1 containing small and large pocket domains. LxCxE interaction domain in right arm of B pocket region.*

The large pocket domain of pRB includes the small pocket plus additional sequence at the C-terminus of the protein (Henley, 2012). Within the pocket domain of pRB is a sequence referred to as the LxCxE binding domain (Dick & Rubin, 2013) (Figure 1.2). The role of the LxCxE binding domain is to recognize and bidn to proteins that contain a six residue amino acid

sequence containing a Lysine (L), Cysteine (C), and Glutamic Acid (E), each separated by any other amino acid (Chan, La Thangue & Smith, 2001). Some proteins which contain the LxCxE motif include HPV E7, adenovirus E1a, HDAC1 and 2, and E2F (Dahiya et al., 2000). pRB's capacity to bind and restrict E2F transcription factor activity, a key aspect of pRB's role as a tumor suppressor, are dependent on the LxCxE binding domain (Dick, 2007). Phosphorylation of pRB by CDKs disrupts binding of LxCxE motif-containing proteins. Additionally, the binding of viral oncoproteins, such as adenovirus E1a and HPV E7, preclude pRB from associating with LxCxE domain-containing proteins. Thus, the pocket domain is deemed the minimal growth suppressing domain of pRB (Henley, 2012).

1.3 pRB Contains Other Non-LxCxE Binding Protein Interactors.

pRB can additionally interact with some proteins independent of the LxCxE domain. Cyclin D forms a complex with CDK4, which directly interacts with pRB to hyperphosphorylate the protein and further, induce entry into S phase (Kato et al., 1993). Cyclin D family proteins bind to pRB via the functional T/E1A/E7-pocket and carboxy-terminal-terminal sequences. Since entry into S phase is highly dependent on phosphorylation of pRB, cells where Cyclin D is mutated or lost will disrupt efficient cell cycle entry (Kato et al., 1993).

1.4 Depletion of pRB in Cancer Cells Containing Hyperphosphorylated pRB.

Within a cancer context, pRB has been shown to be mutated in addition to highly expressed in its hyperphosphorylated state. Cancerous cells containing hyperphosphorylated pRB, also express a high level of pRB (Chatterjee, 2004). When pRB is overly phosphorylated, the G1 regulatory capabilities are lost, but pRB still has functionality within G2/S/M phases. Thus, cancer cells which express hyperphosphorylated pRB have slower cancer progression than

cells containing pRB depletions as pRB has some tumor suppressive ability (Tamrakar, 2000). Since much of pRB's activity in G2/M/S phase are unknown, discovering these binding partners are essential to understanding the protein's entire tumor suppressive capabilities.

2. Project Goal

The goal of the project was to utilize recently identified pRB interactomes to predict functional pRB complexes during G1 and S/G2/M phases of the cell cycle. Individual proteins within the complexes were further analyzed to understand mechanism of interaction between binding partners, with focus on interaction made with hyperphosphorylated pRB during S/G2/M phases.

3. Materials and Methods

3.1 STRING Protein Interaction Database.

In order to initially map pRB protein interactions, STRING was used to assess strong interactions between pRB and known cell cycle binding partners. The Search Tool for the Retrieval of Interacting Genes (STRING) is a protein interactome mapping database, which maps strength of interactions based on a database of physical and functional interactions (Szklarczyk et al., 2019). My analysis was based on a list of pRB-interacting proteins that were experimentally defined by Sanidas et al. In this study, cell lines in which the endogenous RB1 gene had been homozygously deleted were used. The authors then rescued loss of endogenous pRB protein through inducible expression of a construct carrying a wild-type RB1 allele, or one carrying a mutated RB1 gene. (Sanidas et al., 2019). The mutated RB1, called RB**Δ**cdk, contained alanine mutations at serine residues within the large pocket domain that are known to be phosphorylated by cyclin dependent kinases (CDK). The authors analyzed cells that expressed the wild type or CDK mutant pRB protein by mass spectrometry to identify protein interactors of RB1 during G1 and S/G2/M phases (Sanidas et al., 2019). The authors confirmed they were identifying pRB interactors through conducting immunoprecipitation of active RB. As shown in Figure 3.1 below, the two sets of protein interactors, G1 and S/G2/M phase, were further analyzed to understand the two pRB interactomes.

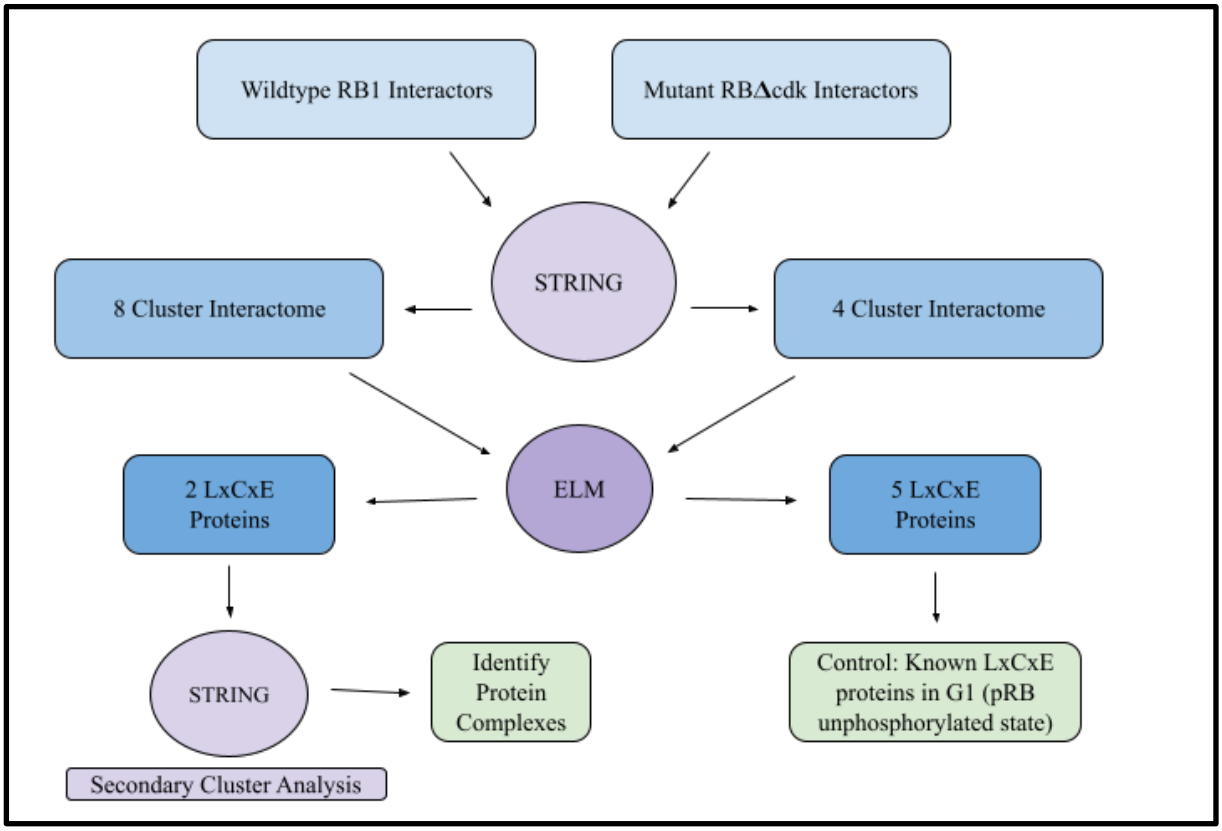

Figure 3.1. *Flowchart of overall methods containing STRING and ELM analyses.*

To further explore specific RB interactions and define potential functional complexes, I formed protein interactomes using the two lists of proteins containing RB1 interactors of either hypophosphorylated or hyperphosphorylated pRB to understand interactions during G1 and S/G2/M phases, respectively. STRING utilizes published databases of protein interaction information to propose physical and functional relationships between two proteins (Szklarczyk et al., 2019). To visualize and compare interactomes of 27 proteins experimentally determined to interact with the RB**Δ**cdk and 243 proteins experimentally determined to interact with wildtype RB1, but not with RB**Δ**cdk mutant, each list of proteins was uploaded into STRING. As depicted in Figure 3.2 below, parameters were set in STRING, such that confidence cutoff for shown

interaction links was 0.400, or the medium setting for confidence. The scores were calculated based on the interaction sources listed below.

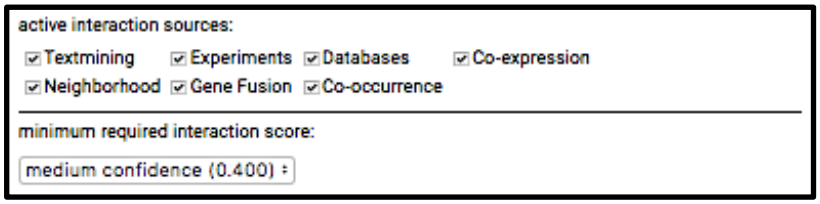

Figure 3.2. *The parameters were set as shown above in STRING for the initial wildtype RB1 and RB**Δ**cdk mutant interactomes. The minimum required interaction score was 0.400, which was the medium confidence setting within STRING.*

In order to understand the role pRB may play in binding these proteins in S/G2/M phases, I next used STRING to define clusters of proteins that interact physically or functionally with each other, independent of pRB. I conducted the same STRING analysis for the RB**Δ**cdk mutant interactors (ie those that interact with pRB in G1) to act as a control for the wildtype-only pRB interactome data (ie those that interact with pRB in S/G2/M). In my analysis, clusters were defined as protein groups containing interactions with a high confidence score of at least 0.900. To do this, I exploited a clustering mechanism provided within STRING where the user may alter the number of clusters present within the interactome; the software forms these clusters based on confidence score, which is generated by experimentally defined interactions amongst two proteins (Szklarczyk et al., 2019). STRING visually signifies strength of interaction through the node thickness, in that a thicker node indicates a more statistically significant interaction (Szklarczyk et al., 2019).

To find the most robust interactions, the number of clusters was tested from a range of two to twelve clusters until proteins within a cluster had a high confidence score of at least

0.900. After testing the range of possible clusters for the S/G2/M phase interactors, eight clusters were chosen for further analysis as this proved to be the optimal number to show nodes with desired confidence interval. Four clusters were chosen for further study for the G1 protein interactors as this number of clustering provided interactions with a confidence of 0.900 or greater.

To further explore the initial STRING data, the eight clusters identified in the complete pRB protein interactome were subsequently analyzed for smaller secondary clusters. The set of proteins within a single cluster were first input into STRING in order to identify protein complexes within the secondary clusters. Performing this tiered clustering, rather than simply defining a large number of clusters in the first level of analysis allowed me to analyze functional units of protein interactions, or potential complexes.

3.2 ELM Protein Domain Binding Database.

To further analyze STRING secondary clusters, the Eukaryotic Linear Motifs (ELM) database was used to identify proteins containing an LxCxE motif sequence. The ELM database is a large database of previously defined short linear motifs (SLiMs), which are prevalent in functional regions of proteins (Kumar et al., 2020). Thus, the database can be exploited to search for a specified motif of interest. In order to understand pRB binding interactions, the ELM database was used to identify presence of the LxCxE sequences binding cleft within the proteins included in the G1 and S/G2/M phase interactomes.

To determine if LxCxE domain containing proteins may function as intermediate ‘hubs’ through which proteins that lack an LxCxE domain may interact with pRB, LxCxE domain-containing proteins were input into STRING individually and analyzed for their capacity to form

complexes with other proteins in their respective G1 or S/G2/M interactome lists. Clusters that emerged from this analysis were taken to indicate subcomplexes which could be linked to pRB indirectly via an LxCxE domain-containing protein.

4. Results and Discussion.

4.1 Identification of Subcomplexes Within the pRB Interactome.

The two protein lists to be analyzed for pRB binding activity, were generated from data produced by Sanidas et al, where scientists induced removal of endogenous RB1 and further rescued the lost protein to express a wild-type RB1 allele or a mutated RB1 allele; the mutated RB1, termed RB**Δ**cdk, contained alanine mutations at known CDK phosphorylation sites, which further minimize phosphorylation of pRB (Sanidas et al., 2019). Co-immunoprecipitations of pRb and any interacting proteins were then analyzed using mass spectrometry to identify protein interactors of hyperphosphorylated and hypophosphorylated pRB, which I interpreted to reflect binding partners during S/G2/M phases and G1, respectively. To enable this analysis, the hypophosphorylated RB1 (ie G1) interactome and hyperphosphorylated RB1 (ie S/G2/M) interactome were first redefined by replacing protein names for the respective gene codes and the input into STRING. As shown in Figure 4.1 below, the hypophosphorylated RB1 (G1) interactome contained 27 proteins in addition to RB1.

Within STRING, clustering of proteins is calculated through a confidence score and visually represents strength of interaction through the thickness of the edge connecting two protein nodes. I experimentally defined a confidence cutoff of 0.900 that resulted in the generation of 4 distinct clusters of proteins that interact with hypophosphorylated pRB. Within the hypophosphorylated pRB interactome (Figure 4.1), the clustering appears to be most robust, in terms of edge strength, within the red and green clusters. pRB's role in G1 regulation is very well understood, with many binding partners extensively studied. HDAC and E2F family proteins are amongst the known binding partners of pRB during G1; E2F forms the active transcriptional complex with pRB, which regulates the R point of the cell cycle. Consistent with

these functional studies, my identified G1 clusters contained E2F family proteins, E2F2 and 4, and HDAC proteins, HDAC 1 and 2.

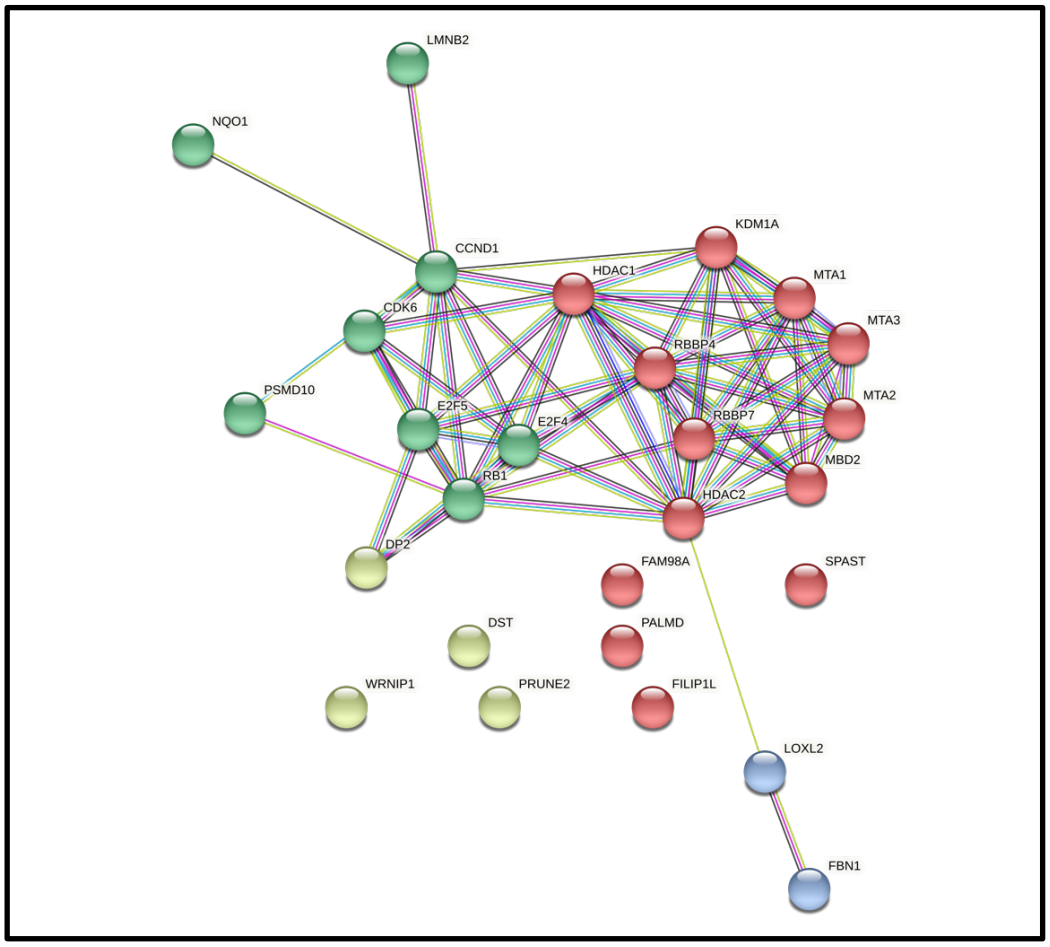

Figure 4.1. *Hypophosphorylated pRB interactome containing protein interactors during G1 phase of the cell cycle. Same colored nodes belong to the same cluster and clustering is dependent on the statistical significance of functional relationships. Thus, proteins may not contain an edge, and identify with a cluster.*

In contrast to well-studied roles for hypophosphorylated pRB during G1, there is considerably less known about the functional complexes formed by hyperphosphorylated pRB during later stages of the cell cycle. The proposed interactome, shown in Figure 4.2, contains 243 proteins and includes many previously unknown binding partners. Using the same 0.900

confidence cutoff, I defined 8 clusters of RB interacting proteins from the S/G2/M list for further analysis.

Amongst the defined clusters above for this interactome, tight binding was observed amongst protein interactors within the orange and olive green clusters, which is indicative of close physical and functional relationships amongst proteins. The initial 8 clusters in the hyperphosphorylated pRB interactome were further analyzed in STRING for the capacity to form secondary or sub-clusters of interacting proteins.

4.2 STRING Analysis of S/G2/M Phase Interactors For Secondary Clusters.

In order to highlight how each cluster is interacting with pRB during S/G2/M phases, STRING was used to separately plot pRB and the proteins within each cluster. The secondary interactomes for clusters 1 and 2, color coded red and yellow in Figure 4.2, are shown below (Figure 4.3). The interactomes visually showed secondary clusters by same color coded nodes.

Cluster 1 contained 25 proteins, which could be clustered into 3 secondary clusters. As depicted in Figure 4.3, cluster 1 shows RB1 in the lime green cluster where it contains many strong binding partners; RB1 contains thick edges to FOS, ACTB, and CTNNB1, which are implicated in maintaining the structural cytoskeleton of cells. Additionally, cluster 2 contained 39 proteins, which were subclustered to show 4 distinct secondary clusters. Cluster 2 localizes RB1 in the red cluster where the protein appears to show strong binding to HIST1H1B and CBX5; both proteins contain functionality in interacting with chromatin.

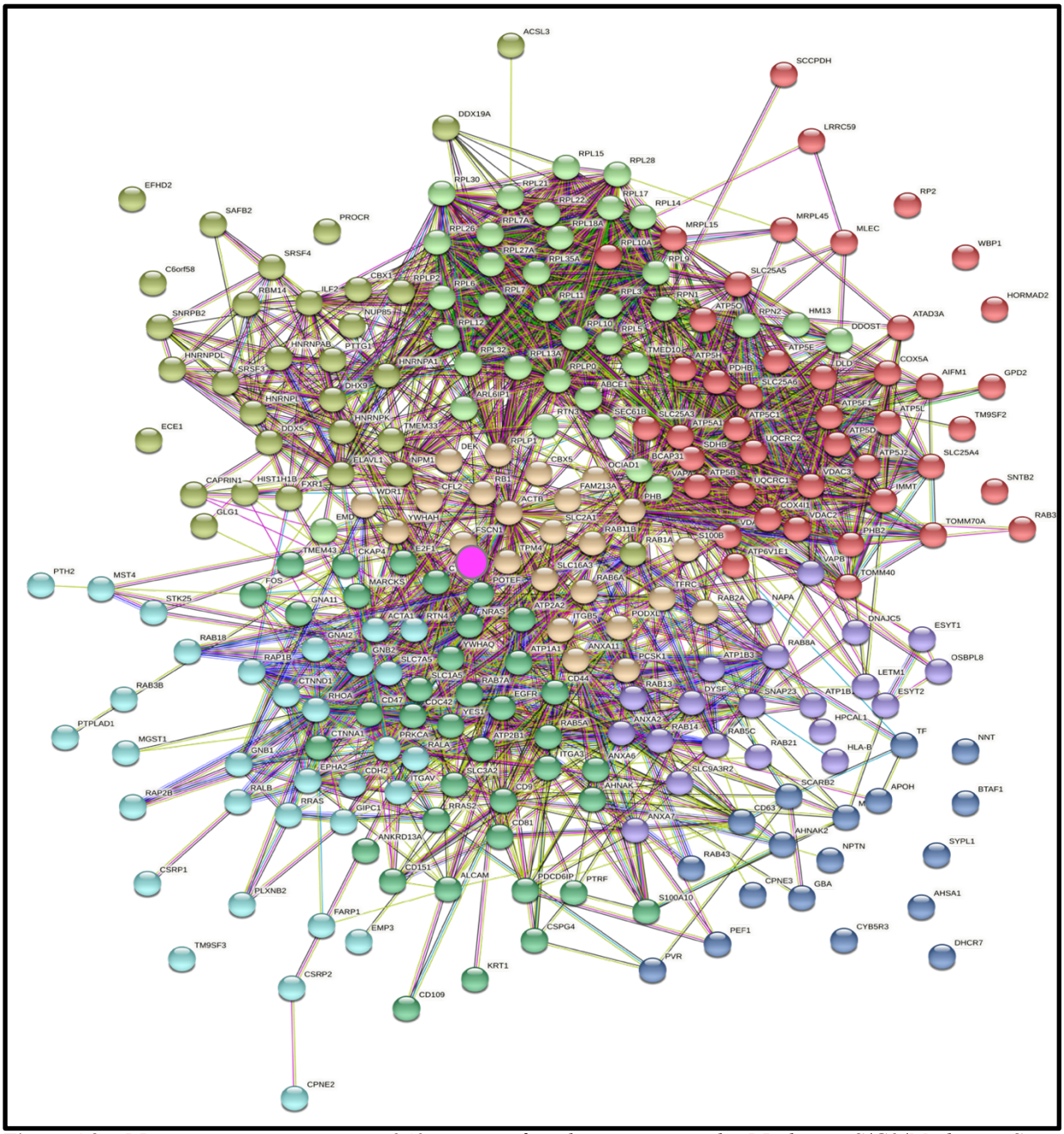

Figure 4.2. *pRB interactome containing 243 proteins found to interact with pRB during S/G2/M phases. Same colored nodes belong to the same cluster. Clustering is dependent on the strength of interactions between proteins, which is indicated by the statistical significance of functional relationships. Thus, some proteins do not contain an edge, but identify in a cluster. RB1 is indicated by the magenta node.*

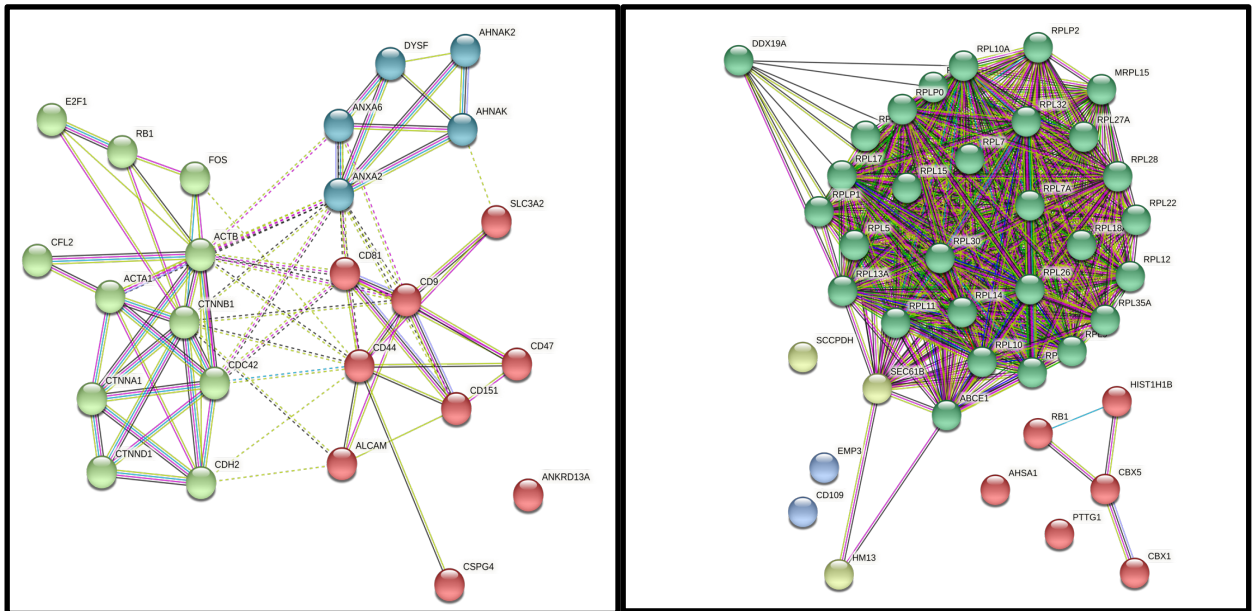

Figure 4.3. *Cluster 1 (left) and cluster 2 (right) interactomes for secondary cluster analysis. In reference to the original pRB interactome for S/G2/M phases shown in Figure 4.2, red is cluster 1 and yellow is cluster 2. Cluster 1 contained 3 secondary cluster and cluster 2 depicted 4 distinct subclusters.*

The secondary interactomes for clusters 3 and 4, color coded olive and green, respectively in Figure 4.2, are depicted in Figure 4.4. Cluster 3 contained 32 protein interactors, which were further grouped into 4 subclusters. As shown in Figure 4.4, pRB is present within the blue subcluster; since there are no connected protein interactors and pRB is the sole protein of the cluster, pRB does not have any statistically significant binding partners within this cluster. In addition, cluster 4 was comprised of 13 proteins that were subclustered for 4 secondary clusters. Cluster 4 shows pRB within the red secondary cluster, where it shows functional interactions with EFHD2, PTPLAD1, SYPL1, and PTH2 in regulation proliferative pathways.

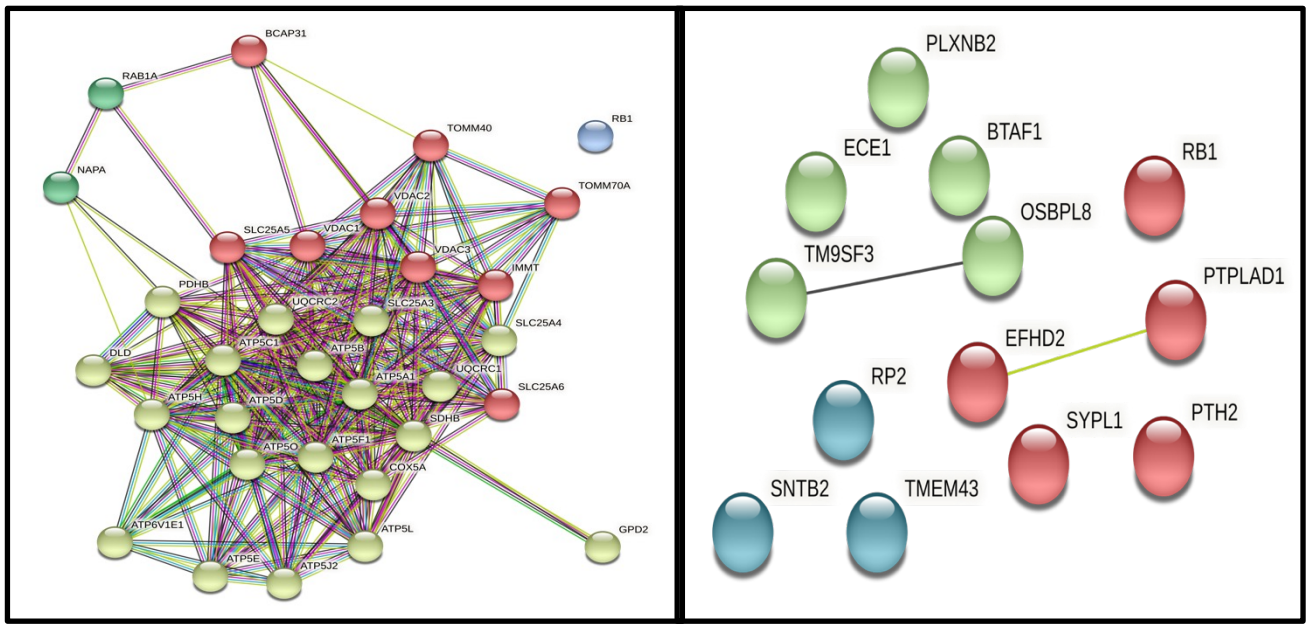

Figure 4.4. *Cluster 3 (left) and cluster 4 (right) interactomes for secondary cluster analysis. In reference to the original pRB interactome for S/G2/M phases shown in Figure 4.2 olive is cluster 3 and green is cluster 4. Cluster 3 contained 4 secondary cluster and cluster 4 showed 3 subclusters.*

The secondary interactomes for clusters 5 and 6, color coded lime green and sky blue, respectively in Figure 4.2, are shown in Figure 4.5. Cluster 5 contained 31 protein interactors, which were distributed into 4 subclusters. As shown in Figure 4.5, pRB is present within the green cluster. Although there are no physical edges present in connecting pRB to another interactor, the protein does belong to a cluster with MYOF, ATP1B3, RAB11B, HLA-B, PCSK1, PEF-1, APCH, and RAB2A, which indicates related function in protein transport. Additionally, cluster 6 contained 31 proteins, that were subclustered into 5 secondary clusters. As depicted in Figure 4.5, pRB is shown within the green subcluster. Although pRB does not have any edges within the interactome, the protein clusters with WDR1, ACSL3 and RALB, which is indicative of shared functions within KRAS-driven tumorigenesis.

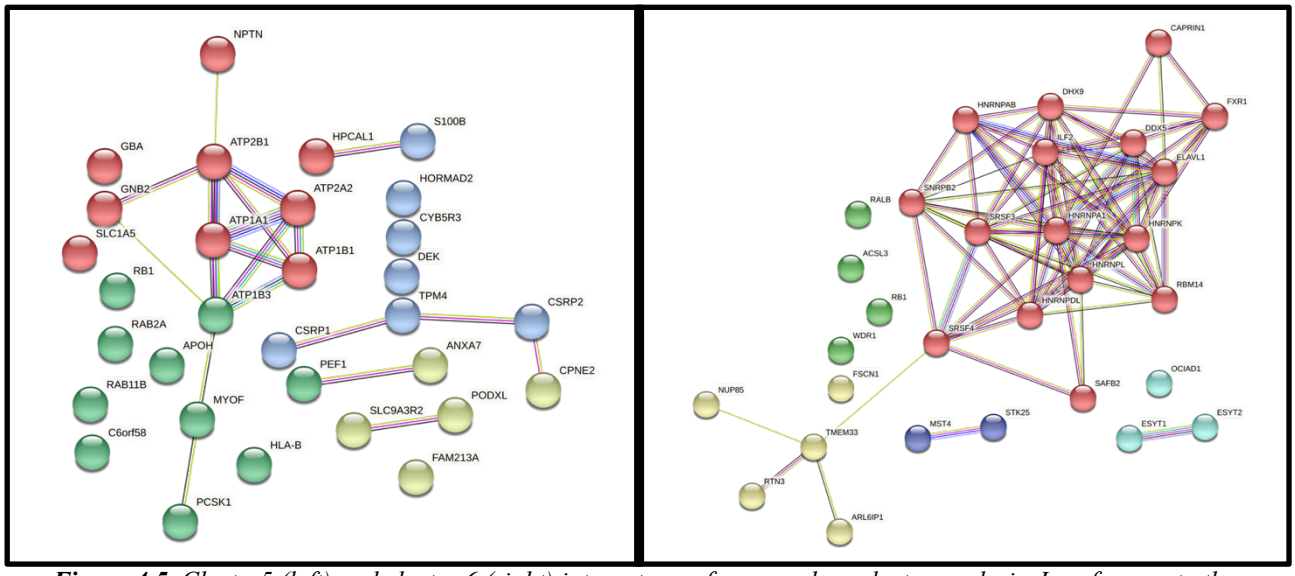

Figure 4.5. *Cluster5 (left) and cluster 6 (right) interactomes for secondary cluster analysis. In reference to the original pRB interactome for S/G2/M phases shown in Figure 4.2 lime green is cluster 5 and sky blue is cluster 6. Cluster 5 contained 4 secondary cluster and cluster 6 showed 5 subclusters.*

The secondary interactomes for clusters 7 and 8, color coded blue and purple in Figure 4.2, are shown in Figure 4.6. Cluster 7 contained 17 protein interactors, which were further distributed into 3 secondary clusters. As shown in Figure 4.6, pRB appears within the red subcluster, where it shares function with MLEC, RPN2, RPN1, TMED10, VAPA, VAPB, DDOST, DNAJCS, RAP1B, and COX4I1, in maintaining integrity of endoplasmic reticulum. In addition, cluster 8 contained 13 protein interactors, which were grouped into 3 subclusters. Within Figure 4.6, pRB is localized in the green subcluster along with 5 other proteins, where the protein appears to show strong binding to PHB. PHB functions in acting as a chaperone for respiration chain protein in the mitochondria.

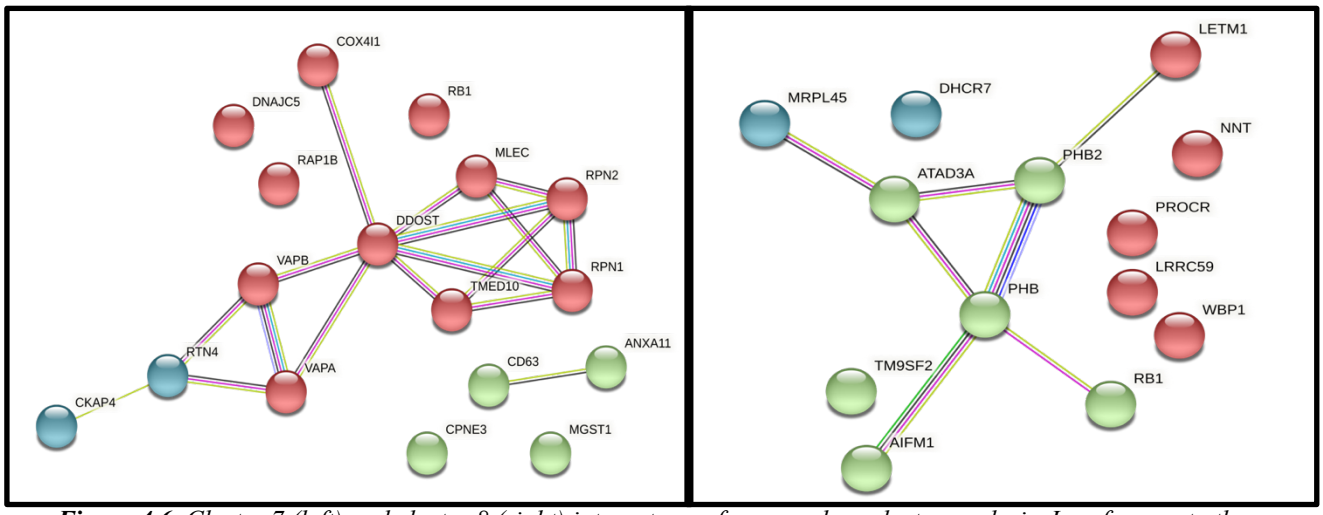

Figure 4.6. *Cluster 7 (left) and cluster 8 (right) interactomes for secondary cluster analysis. In reference to the original pRB interactome for S/G2/M phases shown in Figure 4.2 blue is cluster 5 and purple is cluster 6. Cluster 5 contained 3 secondary cluster and cluster 6 showed 3 subclusters.*

4.3. ELM Protein Domain Analysis For LxCxE Binding Motif.

The Eukaryotic Linear Motif (ELM) software is a protein motif mapping database. ELM allows the user to input a given gene sequence, gene name, or protein code for which it analyzes the amino acid sequence and identifies short motifs that are characterized as being recognition motifs for regulatory proteins or binding partners.

The ELM software was used to analyze pRB binding partners in G1 phase, for the presence of the LxCxE domain. The LxCxE motif was chosen for study since RB contains the recognition domain for this sequence and many well-characterized pRB binding partners have already been demonstrated to contain the motif (Dahiya et al., 2000). This analysis indicated only five of the twenty-seven proteins that interact with hypophosphorylated pRB contain an LxCxE domain (CCND1, DST, HDAC1, HDAC2, and PSMD10). The remaining twenty-two proteins do not contain motifs that are known to moderate direct binding to pRB, suggesting they may interact with pRB through a novel, LxCxE-independent mechanism. Alternatively, these proteins may interact with pRB indirectly by associating with an LxCxE-containing protein

which in turn binds to pRB. Consistent with this possibility, three of the four functional subcomplexes defined by STRING analysis of the hypophosphorylated pRB (Figure 4.1) contain one or more LxCxE-domain containing protein.

Proteins that express an LxCxE motif and are included within the hypophosphorylated pRB interactome, fall into two main categories: proteins which regulate pRB function through phosphorylation and those that link pRB to chromatin modification (Table 4.1). The first category, regulation of pRB phosphorylation, contained the protein, Cyclin D (CCND1). This protein is a regulatory subunit of a holoenzyme, which phosphorylates and inactivates pRB, enabling cell cycle progression from G1 to S phase (Fu et al., 2004). The second category, proteins with function in chromatin modification, include HDAC1 and 2, PSMD10, and DST. HDAC1 and HDAC2 respond to DNA damage and interact directly with core histones (Miller et al., 2010). The two proteins deacetylate lysine residues on the N-terminal portion of histones, H2A, H2B, J3, and H4. Thus, both proteins are Class I deacetylases, since they have the ability to remove lysine-acetyl marks (Miller et al., 2010). PSMD10 is a subunit of the PA700/19S complex, which acts as a regulatory unit of the 26S proteasome (Dawson et al., 2002). This proteosome is essential for ubiquitin-dependent protein degradation within cells (Dawson et al., 2002). Finally, the LxCxE containing protein, DST, is a cytoskeletal linker protein; DST aids in integrating intermediate filaments to the actin cytoskeleton and regulating the overall stability of the microtubule network (Yang et al., 1999).

Protein Name	Function
CCND1	Encodes cyclin D1 protein, which phosphorylates & inactivates RB allowing progression from G1 to S phase
DST	Acts as integrator of filaments, actin & microtubule cytoskeleton networks
HDAC1	Deacetylates lysine residues on N-terminal part of cores histones
HDAC2	Deacetylates lysine residues on N-terminal part of cores histones (H2A, H2B, H3, H4), acts with HDAC1
PSMD10	Proteosome assembly chaperone

Table 4.1. *G1 interactors containing LxCxE domain. The 5 proteins reported to contain the motif were CCND1, DST, HDAC1, HDAC2, and PSMD10.*

I next performed the same ELM analysis on the S/G2/M phase pRB interactome. Interestingly, only 2 of the 243 proteins identified to interact with hyperphosphorylated pRB were identified as having an LxCxE motif (CBX5 and DLDH) (Table 4.2). The two LxCxE-containing proteins in the hyperphosphorylated pRB interactome, CBX5 and DLDH, are part of cluster 2 and 3, respectively.

The gene CBX5 encodes a highly conserved protein, which is localized in the heterochromatin and associates with centromeres (Ligresti et al., 2019). The functional CBX5 protein binds to crucial kinetochore proteins, such as MIS12, and further aids in proper kinetochore formation. Additionally, CBX5 can cause epigenetic repression by binding to

methylated lysine 9 residue H3 (H3K9me) (Ligresti et al., 2019). When the protein binds to H3K9me, it forms a transcriptional repressor complex, which prevents transcription of downstream genes. The second protein found to contain LxCxE was DLDH. DLDH encodes the protein product, dihydrolipoamide dehydrogenase, which is a highly essential oxidoreductase. This enzyme forms the E3 component, which binds to various other enzymes (Babady et al., 2007). One prominent role of DLDH is to produce lipoic acid and NADH from dihydrolipic acid and NAD+ through a reduction reaction. NADH is essential in catabolic respiration reaction necessary to further produce energy for the cell (Babady et al., 2007). Thus, DLDH activity is necessary to form crucial complexes in maintaining proper metabolism, in terms of NADH levels.

Protein Name	Function
CBX5	Mediates Gene Silencing
DLDH	Converts dihydrolipic acid and NAD+ to lipoic acid and NADH

Table 4.2. *S/G2/M RB interactors containing LxCxE domain. The 2 proteins reported to contain the motif were CBX5 and DLDH.*

While the presence of CBX5 and DLDH in STRING clusters two and three could potentially serve to link pRB to the remaining proteins in each of these two clusters, they cannot explain the manner by which proteins in the remaining 6 clusters were able to interact with hyperphosphorylated pRB (Figure 4.2). Instead, the absence of an LxCxE motif in the remaining 241 proteins suggests many of these proteins may employ a yet unappreciated, LxCxE-independent manner of associating with pRB.

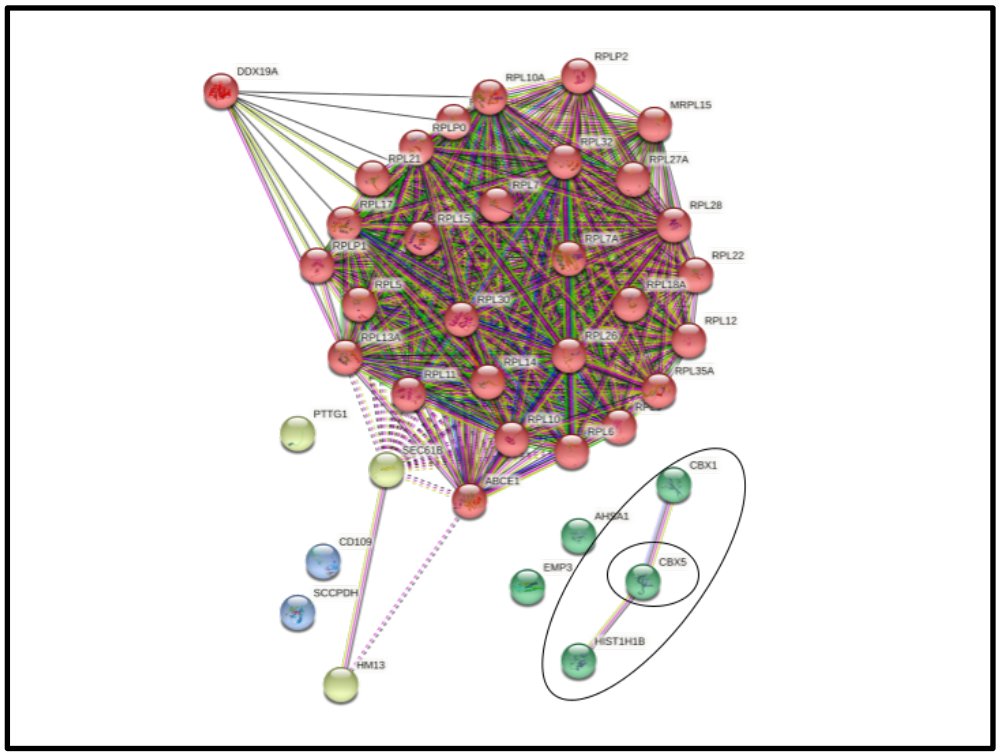

Figure 4.7. *STRING analysis for CBX5 protein. CBX5 is located in the green colored STRING cluster. It's predicted cluster, which contains proteins HISTH1B and CBX1, is denoted by the large oval.*

4.4 STRING Analysis For LxCxE Domain S/G2/M Proteins Found In ELM.

Since RB contains an LxCxE recognition site, I hypothesized that proteins containing LxCxE motifs could function as the link between pRB and proteins that lack an LxCxE motif. To assess this possibility, I performed further STRING analysis of the LxCxE containing proteins from the hyperphosphorylated pRB interactome. For this secondary STRING analysis, clusters function was set to 4 to allow for optimal clustering. This analysis indicated that CBX5 interacts strongly with HIST1H1B and CBX2 (Figure 4.7). Similar analysis of DLDH predicts strong interactions with ATP5A1, ATP5B, ATP5C1, ATP5D, ATP5E, ATP5F1, ATP5H, ATP5J2, ATP5L, ATP5O, ATP6V1E1, COX5A, PDHB, SDHB, SLC25A3, UQCRC1, and UQCRC2

(Figure 4.8). These data support the model that some interactions of proteins that lack an LxCxE motif with hyperphosphorylated pRB may be moderated by binding first to CBX5 or DLDH.

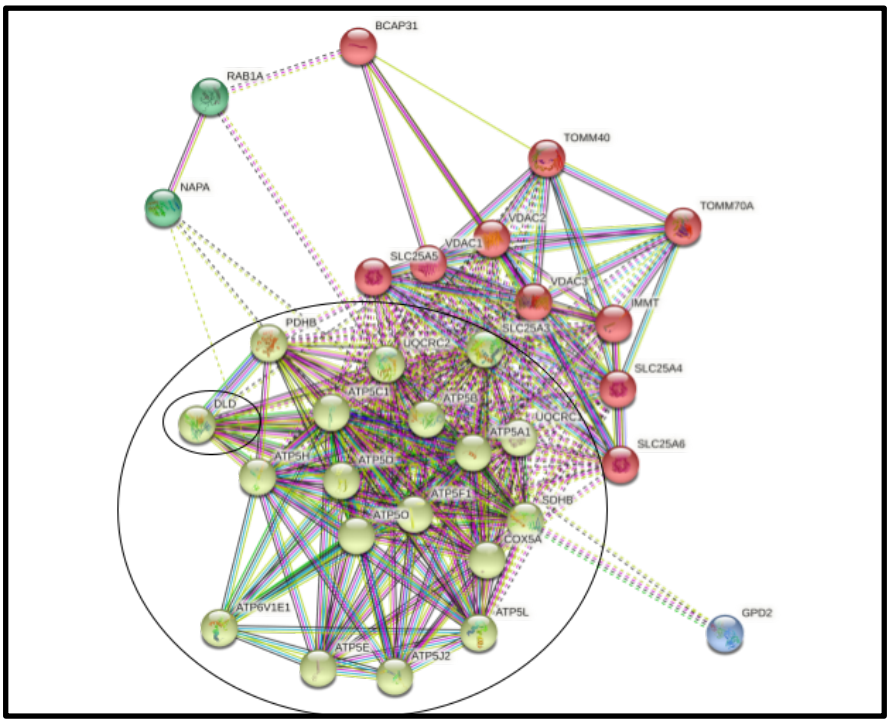

Figure 4.8. *STRING analysis for DLDH protein. DLDH is recognized by its alias DLD within STRING; it is located in the yellow-colored set of nodes. It's predicted cluster, is denoted by the large oval, which contains all nodes within the yellow cluster.*

5. Conclusion

In conclusion, the presence of an extensive protein interactome specific to hyperphosphorylated pRB indicates that pRB contains abundant cellular functions aside from regulating the G1 checkpoint. Through utilizing STRING and ELM, the data showed that pRB's LxCxE recognition domain is prominent in mediating protein interactions during G1 of the cell cycle but not in S/G2/M phases; only 2/243 proteins within the hyperphosphorylated pRB interactome were found to contain LxCxE motifs, while 5/27 of the proteins within the hyperphosphorylated pRB interactome contained the motif.

Amongst the proteins which contained LxCxE motifs in the hyperphosphorylated pRB interactome, CBX5 and DLDH showed high binding capacity to numerous binding partners within respective interactomes. CBX5 appears to form a complex with CBX1 and HISTH1B, which could further interact with pRB. Although DLDH contains a large network of binding partners, the strength of edges is apparent throughout interactions within the subcluster, which could indicate presence of various complexes that interact with each other. Through the strong interactions represented within the data, it is suggested that some interactions with pRB are moderated by DLDH and CBX5. Nevertheless, the vast majority of proteins in the hyperphosphorylated pRB interactome have neither an LxCxE motif, nor demonstrate a strong affinity for an LxCxE motif-containing protein that could otherwise explain their interaction with pRB. These ~252 proteins must instead interact with pRB directly or indirectly through a yet unknown mechanism. To better understand the function of pRB at later stages of the cell cycle, and the role each of these interacting proteins play in those functions, it will be important to determine the mechanisms of interaction for each, and the manner by which these interactions are regulated to enable complex formation during S/G2/M, but not during G1.

References

Babady, N. E., Elpeleg, O., Isaya, G., & Pang, Y.-P. (2007). Cryptic proteolytic activity of dihydrolipoamide dehydrogenase. *Proceedings of the National Academy of Sciences*, *104*(15), 6158–6163. https://doi.org/10.1073/pnas.0610618104

Bertoli, C., Skotheim, J. M., & de Bruin, R. A. (2013). Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology*, *14*(8), 518–528. https://doi.org/10.1038/nrm3629

Blagosklonny, M. V. & Pardee, A. B. (2002). The Restriction Point of the Cell Cycle. *Cell Cyle,* 1(2), 102-109. https://doi.org/10.4161/cc.1.2.108

Caporarello, N., Jones, D. L., Ligresti, G., Meridew, J. A., & Tan, Q. (2019). CBX5/G9A/h3k9me-mediated gene repression is essential to fibroblast activation during lung fibrosis. *JCI Insight*, *4*(12). https://doi.org/10.1172/jci.insight.127111

Chan, H., Smith, L. & La Thangue, N. (2001). Role of LXCXE motif-dependent interactions in the activity of the retinoblastoma protein. *Oncogene* **20,** 6152–6163. https://doi.org/10.1038/sj.onc.1204793

Chatterjee, S. J., George, B., Goebell, P. J., Alavi-Tafreshi, M., Shi, S. R., Fung, Y. K., Jones, P. A., Cordon-Cardo, C., Datar, R. H., & Cote, R. J. (2004). Hyperphosphorylation of pRb: a mechanism for RB tumour suppressor pathway inactivation in bladder cancer. *The Journal of pathology*, *203*(3), 762–770. https://doi.org/10.1002/path.1567

Dahiya, A., Gavin, M. R., Luo, R. X., & Dean, D. C. (2000). Role of the LXCXE binding site in Rb function. *Molecular and cellular biology*, *20*(18), 6799–6805. https://doi.org/10.1128/MCB.20.18.6799-6805.2000

Dawson, S., Apcher, S., Mee, M., Mayer, R. J., & Higashitsuji, H. (2002). Gankyrin is an ankyrin-repeat oncoprotein that interacts with CDK4 kinase and the S6 ATPase of the 26 S proteasome. *Journal of Biological Chemistry*, *277*(13), 10893–10902. https://doi.org/10.1074/jbc.m107313200

Dick, F.A. (2007). Structure-function analysis of the retinoblastoma tumor suppressor protein – is the whole a sum of its parts?. *Cell Div* **2,** 26. https://doi.org/10.1186/1747-1028-2-26

Dick, F. A., & Rubin, S. M. (2013). Molecular mechanisms underlying RB protein function. *Nature Reviews Molecular Cell Biology*, *14*(5), 297–306. https://doi.org/10.1038/nrm3567

Foster, D. A., Yellen, P., Xu, L., & Saqcena, M. (2010). Regulation of G1 cell cycle progression: Distinguishing the restriction point from a nutrient-sensing cell growth

checkpoint(s). *Genes & Cancer*, *1*(11), 1124–1131. https://doi.org/10.1177/1947601910392989

Fu, M., Li, Z., Pestell, R. G., Sakamaki, T., & Wang, C. (2004). Minireview: Cyclin D1: Normal and abnormal functions. *Endocrinology*, *145*(12), 5439–5447. https://doi.org/10.1210/en.2004-0959

Giacinti, C., & Giordano, A. (2006). RB and cell cycle progression. *Oncogene*, *25*(38), 5220–5227. https://doi.org/10.1038/sj.onc.1209615

Giarrè, M., Caldeira, S., Malanchi, I., Ciccolini, F., Leão, M. J., & Tommasino, M. (2001). Induction of pRb degradation by the human papillomavirus type 16 E7 protein is essential to efficiently overcome p16INK4a-imposed G1 cell cycle Arrest. *Journal of virology*, *75*(10), 4705–4712. https://doi.org/10.1128/JVI.75.10.4705-4712.2001

Goodrich, D. W. (2006). The retinoblastoma tumor-suppressor gene, the exception that proves the rule. *Oncogene*, *25*(38), 5233–5243. https://doi.org/10.1038/sj.onc.1209616

Gubern, A., Joaquin, M., Marquès, M., Maseres, P., Garcia-Garcia, J., Amat, R., González-Nuñez, D., Oliva, B., Real, F. X., de Nadal, E., & Posas, F. (2016). The N-Terminal Phosphorylation of RB by p38 Bypasses Its Inactivation by CDKs and Prevents Proliferation in Cancer Cells. *Molecular cell*, *64*(1), 25–36. https://doi.org/10.1016/j.molcel.2016.08.015

Harbour, J. W. (2000). The RB/E2F pathway: Expanding roles and emerging paradigms. *Genes & Development*, *14*(19), 2393–2409. https://doi.org/10.1101/gad.813200

Harbour, J. W., & Dean, D. C. (2000). Rb function in cell-cycle regulation and apoptosis. *Nature Cell Biology*, *2*(4). https://doi.org/10.1038/35008695

Henley, S.A. (2012). The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle. *Cell Div* **7,** 10. https://doi.org/10.1186/1747-1028-7-10

Johnson, D. G. (1998). Role of E2F in cell cycle control and cancer. *Frontiers in Bioscience*, *3*(4), d447–458. https://doi.org/10.2741/a291

Kato, J., Matsushime, H., Hiebert, S. W., Ewen, M. E., & Sherr, C. J. (1993). Direct binding of cyclin D to the retinoblastoma gene product (pRb) and pRb phosphorylation by the cyclin D-dependent kinase CDK4. *Genes & development*, *7*(3), 331–342. https://doi.org/10.1101/gad.7.3.331

Knudsen, E. S., & Wang, J. Y. (1996). Differential regulation of retinoblastoma protein function by specific Cdk phosphorylation sites. *The Journal of biological chemistry*, *271*(14), 8313–8320. https://doi.org/10.1074/jbc.271.14.8313

Komori, H., Iwanaga, R., Bradford, A. P. , & Ohtani, K. A. K. (2018). Distinct E2F-Mediated Transcriptional Mechanisms in Cell Proliferation, Endoreplication and Apoptosis. In (Ed.), Gene Regulation. *IntechOpen*. https://doi.org/10.5772/intechopen.82448

Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B., & Gibson, T. J. (2020). ELM-the eukaryotic linear motif resource in 2020. *Nucleic acids research*, *48*(D1), D296–D306. https://doi.org/10.1093/nar/gkz1030

Miller, K. M., Tjeertes, J. V., Coates, J., & Legube, G. (2010). Human HDAC1 and HDAC2 function in the DNA-damage response to promote DNA nonhomologous end-joining. *Nature Structural & Molecular Biology*, *17*(9), 1144–1151. https://doi.org/10.1038/nsmb.1899

Nicolas, E., Ait-Si-Ali, S., & Trouche, D. (2001). The histone deacetylase HDAC3 targets RbAp48 to the retinoblastoma protein. *Nucleic acids research*, *29*(15), 3131–3136. https://doi.org/10.1093/nar/29.15.3131

Sanidas, I., Morris, R., Fella, K. A., Rumde, P. H., Boukhali, M., Tai, E. C., Ting, D. T., Lawrence, M. S., Haas, W., & Dyson, N. J. (2019). A Code of Mono-phosphorylation Modulates the Function of RB. *Molecular cell*, *73*(5), 985–1000.e6. https://doi.org/10.1016/j.molcel.2019.01.004

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Tamrakar, S., Rubin, E., & Ludlow, J. W. (2000). Role of pRB dephosphorylation in cell cycle regulation. *Frontiers in bioscience : a journal and virtual library*, *5*, D121–D137. https://doi.org/10.2741/tamrakar

Wade P. A. (2001). Transcriptional control at regulatory checkpoints by histone deacetylases: molecular connections between cancer and chromatin. *Human molecular genetics*, *10*(7), 693–698. https://doi.org/10.1093/hmg/10.7.693

Yang, Y., Bauer, C., Strasser, G., Wollman, R., & Julien, J.-P. (1999). Integrators of the cytoskeleton that stabilize microtubules. *Cell*, *98*(2), 229–238. https://doi.org/10.1016/s0092-8674(00)81017-x

Zhang, H. S., Postigo, A. A., & Dean, D. C. (1999). Active transcriptional repression by the RB–E2F complex mediates G1 arrest triggered by p16ink4a, TGFB, and contact inhibition. *Cell*, *97*(1), 53–61. https://doi.org/10.1016/s0092-8674(00)80714-x