Establishment of a Novel Mechanism of MYC Acetylation by ATAC

By

Ryan Gates

University Honors

University of California, Riverside

Dr. Ernest Martinez

Department of Biochemistry

Dr. Richard Cardullo, Howard H Hayes Jr. Chair

University Honors

ABSTRACT

The recognition of modified amino acid residues by "reader" proteins is key to the intracellular mechanisms regulating many gene transcription processes integral to cancer disease progression. Here, I explain the possible new discovery of a novel intracellular interaction between the ATAC complex's "reader" protein, YEATS2, and the MYC oncoprotein which enables the acetylation of key MYC lysine (K) amino acid residues. The results within this capstone are preliminary results which may establish the groundwork for future research establishing the interaction between MYC and YEATS2. The groundwork for this mechanism was evaluated using a series of biochemical assays and bioinformatic analysis. My capstone research has uncovered possible gene networks regulated by MYC's lysine residues which may or may not be dependent on YEATS2 and MYC interaction.

ACKNOWLEDGMENTS

I would like to thank our collaborators at the City of Hope National Cancer Center for providing us with computational support and expertise. In addition to this, I would also like to thank the NIH MARC U STAR program at UCR for providing me with financial support to finish my honors capstone requirements and B.S. degree. I would also like to thank all my lab mates and mentors for helping me with my research throughout my four years at UCR. Lastly, I would like to thank my family and friends for all your love, support, and guidance. I love you all.

Table of Contents

INTRODUCTION

One of the most extensively studied oncoproteins, MYC, is essential for early embryonic development and its deregulation or overexpression is observed in up to 70% of human cancers where it is associated with the initiation, progression, and poor prognosis of most tumors (Dang, 2012). The MYC protein is a regulator of gene expression and modulates the transcription of genes associated with cancer cell proliferation and longevity. My study focuses on how the MYC protein is functionally stabilized by other coactivator proteins in cancer cells. The MYC protein activates gene transcription with help from specific coactivator proteins called Histone Acetyltransferases (HATs) (Faiola et al., 2005). In cells, HATs add an acetyl group to specific lysine (K) amino acid residues of nucleosome histones, affecting chromatin structure, and permitting gene activation. In addition, MYC itself is also directly acetylated by these coactivator HATs, which can lead to its stabilization in cancer cells (Wang et al., 2008) and regulates the ability of MYC to transform cells (our lab's unpublished data). Our lab established that MYC is acetylated by two key HATs: GCN5 and p300 (Patel et al. 2004). P300 acetylates three major MYC lysine residues, K149, K158, and K323, with a marked preference for K149 and K158 (Wang et al., 2008). In contrast, GCN5 specifically acetylates the K323 residue, which stabilizes MYC against degradation by the proteasome. In eukaryotic cells, GCN5 is known to function as a part of two multi-protein complexes called ATAC (Ada Two A Complex) and SAGA (Spt Ada GCN5 Acetyltransferase). While it is known that SAGA can functionally interact with MYC, it is still unknown if ATAC can interact with MYC (Wang et al., 2008). For my capstone project, I worked to understand how the ATAC complex physically and functionally interacts with MYC to regulate specific gene networks and biological processes involved in cancer. To test this, I worked to investigate two specific aims. The first was to

establish the mechanism of MYC acetylation by ATAC. The second was to identify key genetic pathways co-regulated by ATAC and MYC.

RESULTS

To establish the connection between the ATAC complex and MYC, I wanted to first focus on the YEATS2 protein (Figure 1a). The YEATS2 protein contains a "*YEATS acetyl-lysine reader domain*," a domain of the protein that specifically binds to acetylated lysine residues on histone proteins (Mi et al., 2017). Previous data from our lab suggested that MYC can interact with YEATS2 in Human Embryonic Kidney Cells (HEK293). In preliminary co-immunoprecipitation (co-IP) experiments we observed a weak interaction of MYC with YEATS2. My goal was to replicate these experiments done in HEK293 cells with the same materials and reagents. I used HEK293 cells at a low passage number and performed a transfection on these cells with a Flag-Tagged MYC expression vector in conjunction with the transfection reagent Lipofectamine 2000. The goal of this was to induce ectopic expression of FLAG-MYC in these cells to pulldown tagged MYC and associated proteins with an antibody specific to the FLAG epitope. After the pulldown, YEATS2 and protein controls (GCN5 and MYC) were analyzed for conjugation with the pulled-down FLAG-MYC via western blot analysis (Figure 1b). From this experiment, I was unable to determine a strong signal for YEATS2. However, I was able to see a signal for GCN5, which had been previously proven to interact with MYC. It should be noted that in figure 1a, there was a high concentration of MYC protein loaded into the western blot gel which accounts for this "bleeding effect" we see in the blot. Furthermore, all HEK293 samples were treated with HDAC inhibitors 2 hours prior to cell lysis to try and mitigate the possible transient presence of acetylated lysine residues within the sample. Due to the absence of YEATS2 in figure 1a, I decided to redo the experiment probing

for more proteins which may interact with MYC. However, although I was able to get a pulldown with MYC (as can be seen by the signal from MAX, an obligatory MYC binding partner) after treating the cells with HDAC inhibitors for two hours prior to lysis, YEATS2 was still notably absent within in the western blot analysis (Figure 1c). After trying these two CO-IPs with whole cell extracts by lysing the cells in an Eppendorf tube, I decided to try the experiment by lysing the cells in the plate to hopefully be able to see YEATS2 pulled down with MYC in the CO-IP. My reasoning for this is since it is possible to lose IP material during the process of lysing protein samples in a tube, I was hoping that by lysing the cells in the plate, I would be able to mitigate some of these loses. However, even after performing the CO-IP on the HEK293 samples which were lysed in the plate, I was still unable to see YEATS2 in the pulldown sample by western blot analysis (figure 1d).

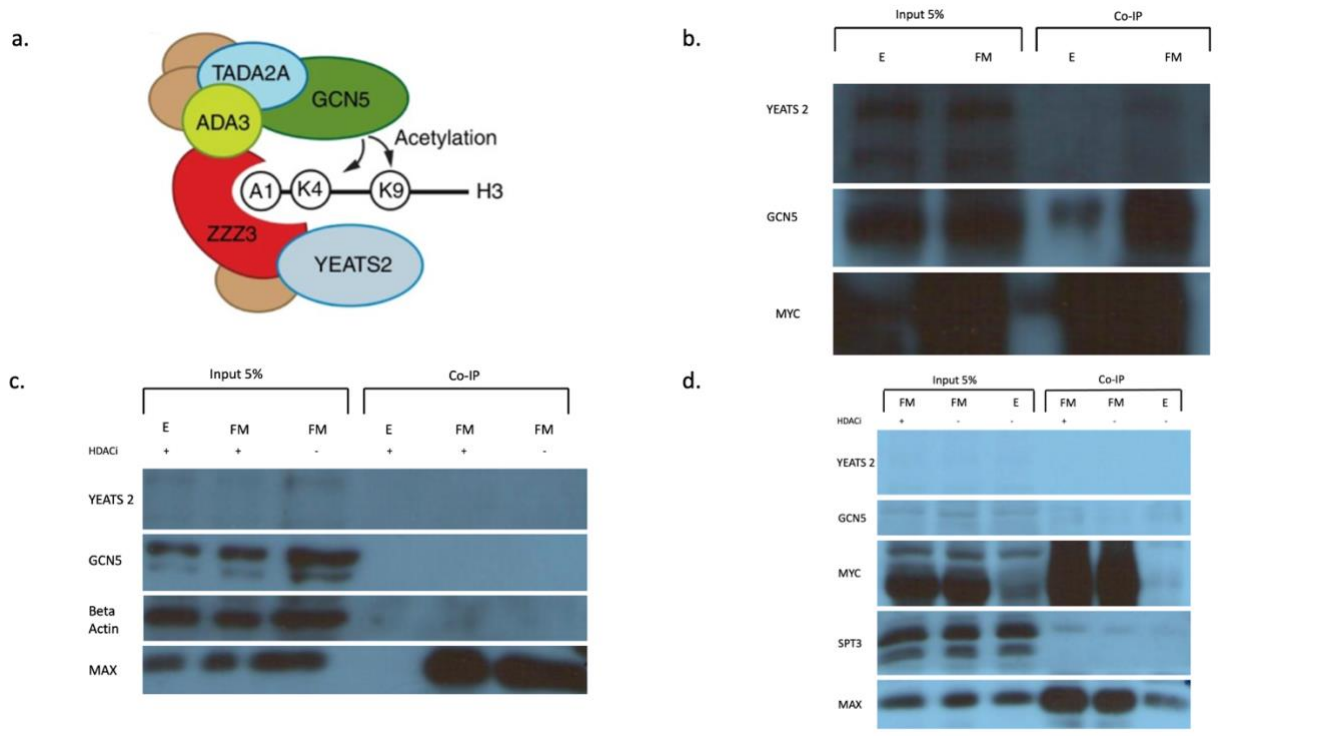

Fig 1. Co-Immunoprecipitation of Transfected FLAG MYC. a. Structure and Function of the YEATS2 Protein (Mi et al., 2017). b. YEATS2, GCN5 and MYC were all probed in the western blot with the addition of HDACi. MYC acts as the CO-IP control to verify the pulldown was successful. c. YEATS2, GCN5, Beta Actin, and MAX were all probed in this western blot analysis. Beta Actin acted as a loading control while MAX acted as the CO-IP control to confirm the pulldown of MYC was successful. It should be noted that MAX is an obligatory binding partner of MYC. Also, expression of MAX is directly proportional to expression of MYC within the cell. d. MAX and MYC both act as the CO-IP pulldown controls in this experiment. Also, SPT3 was probed in this experiment since it was previously shown to hybridize with MYC by our lab as a part of a different HDAC protein complex called SAGA.

To understand the genetic regulatory pathways which are regulated by YEATS2, and possibly MYC in human cells, a previous lab member decided to knockdown YEATS2 with siRNA in HEK293 cells and perform Bulk RNA-seq Analysis on transcribed mRNAs. It was my job to analyze this RNA-seq data and check for any possible MYC target genes which YEATS2 could play a role in regulating. Upon analyzing this data, I found that knockdown of YEATS2 within HEK293 cells caused the downregulation of ribosomal protein genes which are notably integral to cancer cell fitness (Figure 2a and 2b). In addition to deregulation of ribosomal protein genes, knockdown of YEATS2 in HEK293 cells also caused downregulation of MYC target genes (Figure 2c). Of all the MYC target genes found to be downregulated upon YEATS2 knockdown, there were two ribosomal protein genes (RPS10 and RPS6) which were found significantly downregulated from the list of differentially expressed genes. RPS10 specifically is notably important to ribosomal biogenies in cancer as well as in other human diseases (Sehrawat et al., 2019). However, these results provide limited evidence as to if the interaction of YEATS2 and MYC affect the transcription of these genes. Further bioinformatic studies would have to be carried out to determine if the interaction of MYC and YEATS2 was the true culprit of the downregulation of the ribosomal MYC target genes.

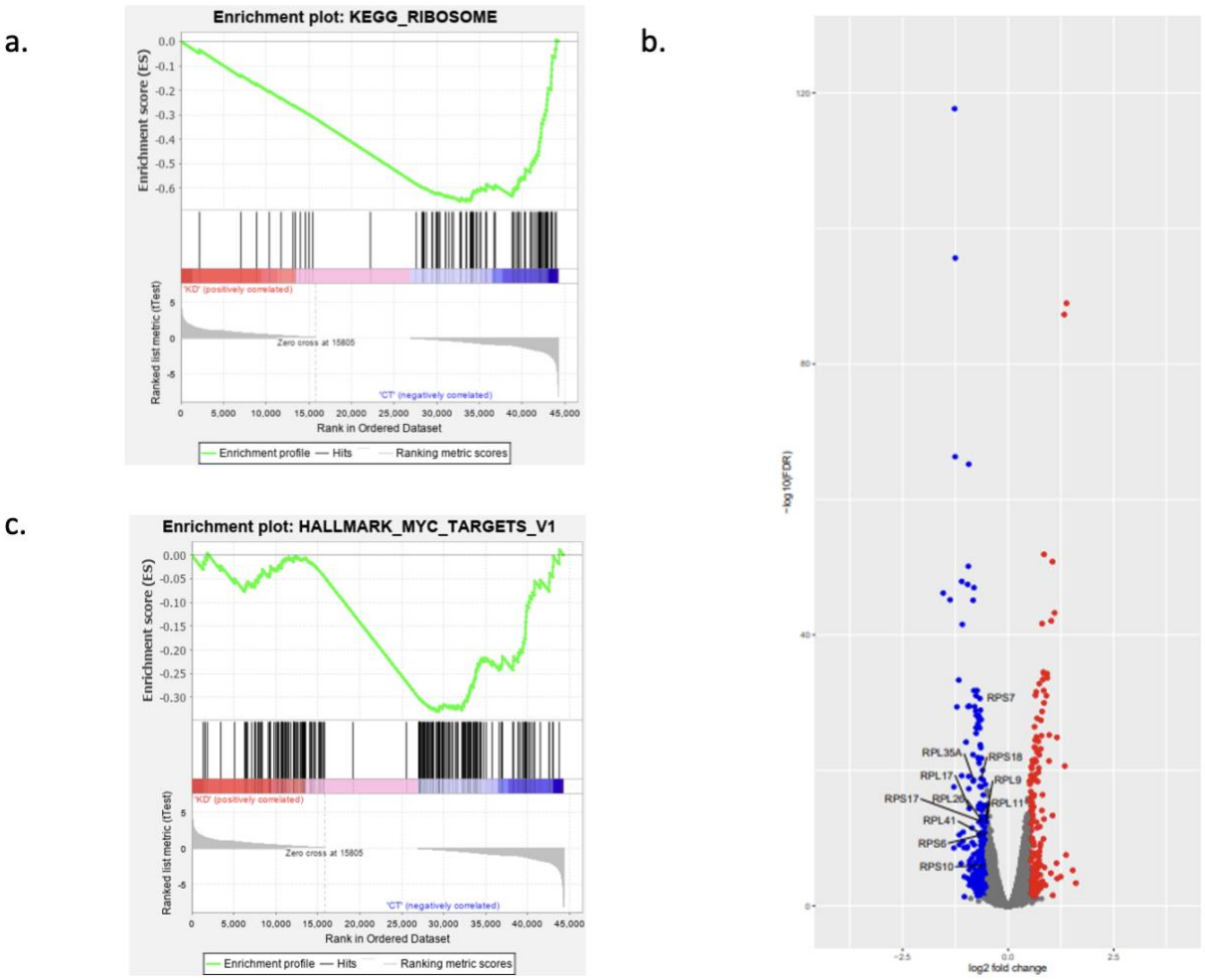

Fig 2. YEAST2 Knock Down RNA Seq Analysis. a. Deregulation of KEGG Ribosome Gene Set. The genes in the KEGG Ribosome gene set show greater enrichment in the control samples (cells not transfected with YEATS2 KD siRNA) which indicates more genes in this gene set are downregulated than upregulated when comparing the gene expression of the YEATS2 KD siRNA treated samples to the control. b. Volcano Plot of Ribosomal Protein Genes. The ribosomal protein genes from the KEGG Ribosome gene set which met the statistical cut off 0.5 log2FC and 0.05 FDR are annotated in the plot. c. Deregulation of Hallmark MYC Targets V1 Gene Set. The genes in the Hallmark MYC Targets V1 gene set show greater enrichment in the control samples which indicates more genes in this gene set are downregulated than upregulated when comparing the gene expression of Knockdown samples to the control.

Lastly, to understand the specific gene regulatory networks which are regulated by the three MYC lysine residues, K149, 158, and 323 a current lab member in the lab stably transformed the commercially available MCF10A cell line with a mouse MYC vector which had an amino acid mutation from lysine to arginine at one of the three MYC lysine residues. This lysine to arginine mutation helps to maintain the structure of the MYC protein within the cell due to the structure homology between lysine and arginine. However, by mutating the lysine residue

to an arginine, the mutated residue cannot be acetylated Our goal in this RNA-seq analysis was to understand how the stably transformed mutant MCF10A cell lines (R149, R158, and R323) gene expression profiles compared with the gene expression patterns of the wild type transformed MCF10A cell line. For the transformed MCF10A cell lines we found a variety of deregulated gene sets for each of the mutant cell lines when compared to the wildtype phenotype which met our cutoff of < 25% FDR q-value (figure 3). For the MCF10A R323 mutant, when compared to the wildtype, we found a total of four deregulated gene sets (figure 4). The genes in these sets were found to be involved with cellular digestion, cellular spliceosome function, and folate biosynthesis. Importantly, from these gene sets, a total of thirteen genes were found to be differentially regulated between the R323 mutant and the control which met are statistical cutoffs of +/- 0.5 log2FC and 0.05 FDR. Secondly, for the R158 MCF10A mutant we found only one statistically significant deregulated gene set when compared to the control. The genes found to be deregulated in this set were found to be involved with epithelial mesenchymal transition (figure 5). Importantly, from this set of genes we found a total of thirty-two gene which met our statistical cutoffs for differential expression. For the R149 MCF10A mutant, we found a total of seven deregulated gene sets when compared to the gene expression profile of the wildtype control sample. The genes from these sets were found to be involved with a variety of processes including the response of EIF2AK4 GCN2 to amino deficiencies, ribosomal protein genes, the P53 pathway in cancer, mitotic spindle assembly, Interferon Gamma and Alpha response and the apical surface pathway (figure 6). From these gene sets, there were one-hundred and thirty deregulated genes which met our cutoffs for statistical significance.

Fig 3. MCF10A GSEA Dot-plot. By using the legend on the side of the plot, we can determine how many genes are within the GSEA gene set by looking at the size of the dot. The significance of the findings can be distinguished by the color of the dot (FDR q value). The Normalized Enrichment Score (NES) signifies whether the genes in the gene set are upregulated or downregulated in the Mutant/WT samples vs. the control.

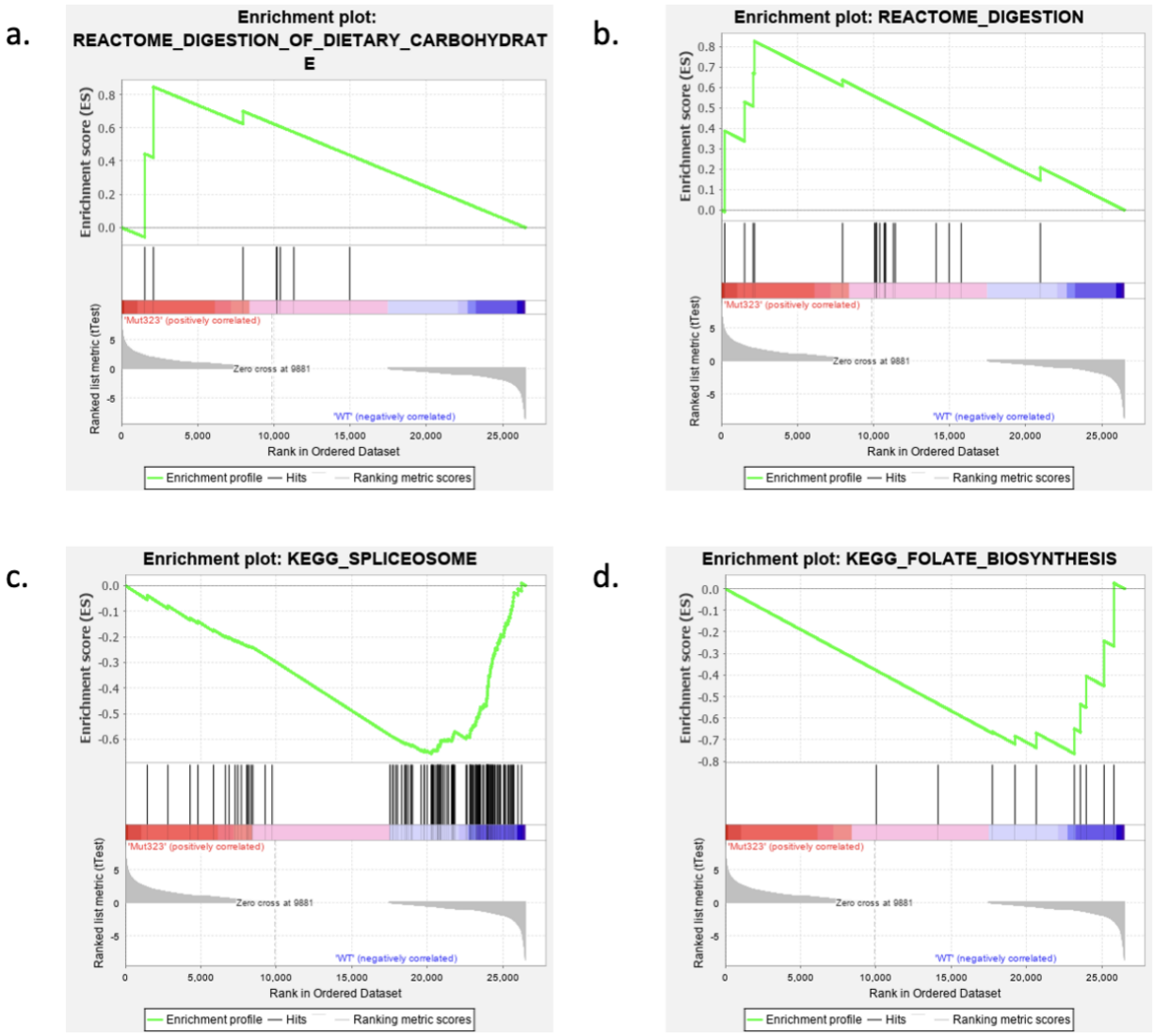

Fig 4. MCF10A R323 Mutant Enrichment Plots. a. Enrichment plot for Reactome Digestion of Dietary Carbohydrate. b. Enrichment plot for Reactome Digestion. c. Enrichment plot for KEGG Spliceosome. d. Enrichment plot for KEGG Folate Biosynthesis

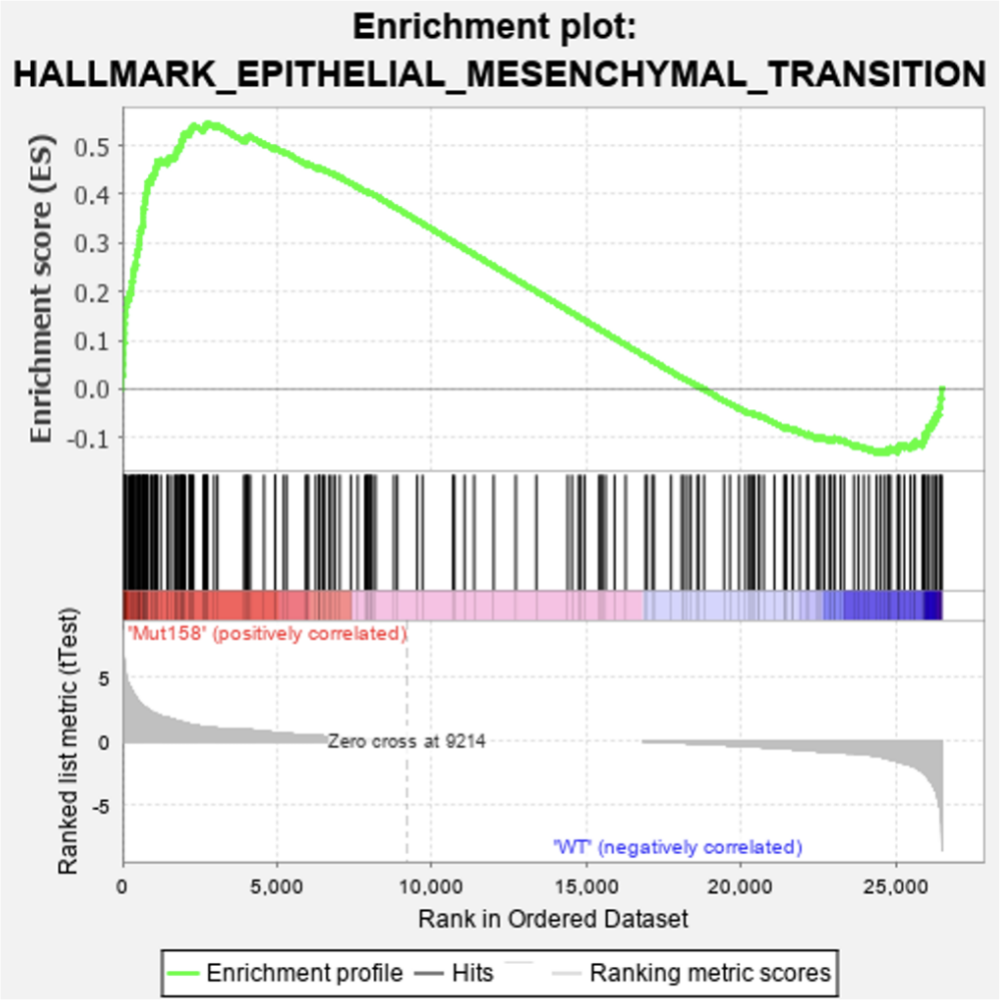

Fig 5. MCF10A R158 Mutant Enrichment Plot. Enrichment Plot for Hallmark Epithelial Mesenchymal Transition.

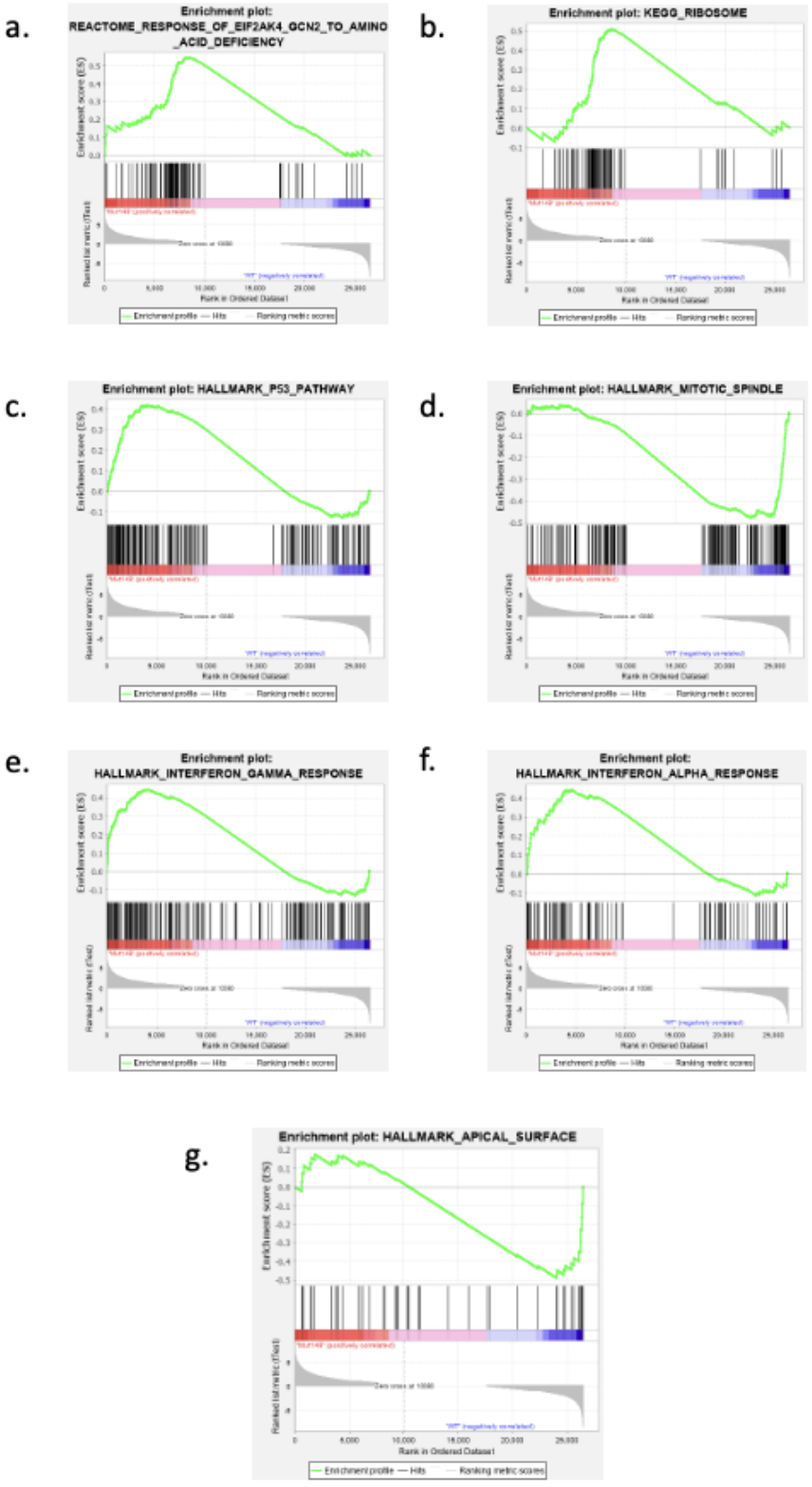

Fig 6. MCF10A R149 Mutant Enrichment Plot. a. Enrichment plot for Reactome response of EIF2AK4 GCN2 to Amino Acid Deficiency. b. Enrichment plot for KEGG ribosome. c. Enrichment plot for Hallmark p53 pathway. d. Enrichment plot for Hallmark mitotic spindle. e. Enrichment plot for Hallmark interferon gamma response. f. Enrichment plot for Hallmark interferon alpha response. g. Enrichment plot for Hallmark apical surface.

DISCUSSION

Starting off with the FLAG-MYC Co-IP data, the results are inconclusive if YEATS2 does truly hybridize with MYC after transfection and treating cells with or without an HDACi. However, the pulldown was successful as could be seen by the presence of signal in our pulldown control lanes (MAX or MYC). There was however some evidence to show that the Histone Acetyl Transferase, GCN5 does interact with MYC which confirms previous studies conducted by our lab. To confirm once and for all if YEATS2 and associated ATAC complex proteins interact with MYC, we would either need to further optimize the Co-IP conditions or perform a Mass Spectrometry analysis on the pulldown product to confirm if the ATAC complex is found to interact with MYC. The later of these experiments might have the biggest impact if it were done with patient derived tumor samples of multiple different cancer types. Doing this would determine in which cancer types the MYC-ATAC interaction is most prevalent and provide clinically significant insights into which cancers the interaction may be important to driving tumorigenesis.

As mentioned previously, siRNA mediated knockdown of YEATS2 in HEK293 cells produced statically significant downregulation of a distinct set of ribosomal protein genes. Downregulation of ribosomal protein genes are of importance in cancer since tumorigenesis requires upregulation of this same set of genes (El Khoury et al., 2021). This finding could elucidate the ATAC complex's function in regulating cancer cell fitness by acting either to acylate MYC or genomic histones, therefore affecting gene transcription. Again, to determine if these results are truly due to the ATAC complex's acetylation of MYC, we would have to perform a ChIP-seq analysis to determine the exact genomic binding site of both MYC and the ATAC complex.

For the RNA-seq analysis in the MCF10A cell line the most interesting gene that I found statistically significantly downregulated in the R323 mutant samples compared to the control samples was HSPA1A. The logFC value for this protein when comparing the gene expression levels of the mutant to the control was -1.1 with an FDR of 2.27 x 10^-11. This means that in the MCF10A WT cells this RNA transcript was twice as prevalent when compared to the mutant R323 cell line. Transcription of HSPA1A is induced downstream of the MAPK/ERK signaling cascade in cancer cells to enhance tumorigenesis suppress cell death and reduce the effects of stress-induced apoptosis (Balogi et al., 2019). This gene is a part of the KEGG spliceosome gene set, and it provides evidence to suggest that presence of acetylation at the K323 residue of MYC may aide in some way with cancer cell fitness by way of the MAPK signaling cascade. As for the R158 mutant cell line, although there was only found to be one deregulated gene set (Hallmark Epithelial Mesenchymal Transition) this gene set is of particular interest to our lab since in breast cancers normal breast carcinoma cells can change their cell surface from epithelial like to mesenchymal. This can allow for breast tumors to become metastatic further allowing the cancer to spread through the body (Wang et al., 2011). One particularly interesting gene that was found to be upregulated in the R158 mutant samples when compared to the controls was DKK1. This gene had a logFC value of 1.71 and an FDR of 3.17 x 10^-6. Intriguingly, this gene was found to inhibit breast cancer metastasis via the suppression of the Beta-catenin/MMP7 signaling pathway (Niu et al., 2019). This means that absence of acetylation at the MYC K158 site may actually prevent breast cancer metastasis and tumorigenesis possibly resulting in more positive patient prognosis. Another interesting gene that was found to be upregulated in the R158 mutant samples compared to the control was POSTN. This gene had an enormous logFC of 4.56 and an FDR of 4.40 x 10^-20. It should be noted that this gene had one of the highest levels of

deregulation between the R158 mutant and control in the entire dataset. POSTN is important for many of processes in breast cancer tumorigenesis. Most notably it has been found to be upregulation of this gene has been associated with a far worse prognosis for patients. It is also believed that the expression of this gene may be regulated by a crosstalk between the FGFR signaling cascade and the TGFbeta/PI3K/AKT pathways (Labrèche et al., 2021). Due to the magnitude of POSTN upregulation in the R158 mutant samples compared to the controls, it is possible that the upregulation of the DKK1 was in fact due to a secondary effect of upregulation of POSTN instead of the direct affect cause by the R158 MYC mutant. It is also possible that the upregulation of POSTN could make for a worse patient prognosis and more aggressive breast cancer due to absence of MYC K158 Acetylation despite the apparent upregulation of DKK1. To better understand the effects of the absence of acetylation on the tumor growth and cancer metastasis a possible experiment would be to take patient derived breast cancer cells and use CRISPR/CAS9 to create MYC R158 mutations prior to injecting them into the immuno-deficient laboratory mice to test for in-vivo cancer cell growth and metastasis. For the MYC R149 mutant cell line there were a total of seven deregulated gene sets when comparing the gene expression profile to the control. These gene sets were, the Reactome response of EIF2AK4 GCN2 to amino acid deficiency, KEGG ribosome, hallmark P53 pathway, hallmark mitotic spindle, hallmark interferon gamma response, hallmark interferon alpha response, and the gene set for hallmark apical surface. The sets that contained genes which were found to be downregulated in the MYC R149 mutant samples were the hallmark mitotic spindle, and hallmark apical surface. The sets which contained genes which were found to be upregulated in the MYC R149 mutant samples were the Reactome response of EIF2AK4 GCN2 to amino acid deficiency, KEGG Ribosome, Hallmark P53 pathway, Hallmark interferon gamma response, and Hallmark interferon alpha

response. Of the upregulated genes in the EIF2AK4 gene set, the most interesting was DDIT3. DDIT3 is a pro-apoptotic transcription factor which has been shown to suppress breast cancer cell growth when it is upregulated within the cell (Block et al., 2019). The differential expression of this gene in the R149 samples compared to the controls was a logFC value of 1.6 and an FDR of 6.55 x 10^-6. This may signify that absence of acetylation at MYC lysine residue 149 may have anti-tumor effects by inducing apoptosis and lead to better patient outcomes. For the KEGG ribosome gene set, there were no genes within the gene set found to be deregulated in the DEGs which met our cutoffs for statistical significance. This was quite disheartening to see since we assumed that there may be some cross-over between the ribosomal protein genes found to be downregulated in the YEATS2 knockdown experiments and the R149 mutants. Although the same gene set was found to be deregulated, more experiments will have to be performed to confirm if YEATS2 and the ATAC complex are responsible for the acetylation of MYC at the lysine 149 residue which may cause the subsequent deregulation of gene expression that we saw in this data set. As mentioned previously, the experiments which could be done to confirm this interaction and subsequent deregulation of gene expression would be to perform a mass-spec analysis and ChIP-seq. For the Hallmark P53 pathway, we were encouraged to see the upregulation of this gene set in our mutant R149 samples since the P53 pathway is known to be so import in regulating cancer cell division and apoptosis. There was a total of thirty-seven genes found to be statistically significantly upregulated within our R149 mutant samples. Of these thirty-seven genes, one gene which stuck out to me was the upregulation of the proto-oncogene c-FOS within our mutant samples. This gene had a logFC value of 0.62 and a FDR of 0.0003. This gene exists to dimerize with another protein named JUN to regulate gene transcription in cancer cells which can cause anchorage independent growth and metastasis (Portal et al., 2006).

However, it should also be noted that interestingly JUN was statically significantly downregulated within out R149 mutant samples (logFC -0.53, FDR 2.29x10^-9). This may indicate the apparent upregulation of c-FOS due to the absence of acetylation at Lysine residue 149 on MYC may not be significant for cancer metastasis due to the reciprocal downregulation of the JUN protein. However, to determine if this is the case, one would still need to determine if spheroidal cancer cells can still grow in an anchorage independent manor if the MYC lysine residue 149 is mutated to arginine. An experiment like this was done by a current lab member and the results from this may shed some light on this mechanism. For the hallmark mitotic spindle gene set, there were a total of forty-five downregulated genes in the R149 samples which met our statical cutoffs. Most notable from this list was KIF23 (logFC -0.57, FDR 1.28 x 10^-9). KIF23 was found to be important to promote triple negative breast cancer via the Wnt/beta-catenin pathway (Zhang et al., 2021). This may mean that absence of acetylation at MYC lysine residue 149 may be beneficial to preventing tumorigenesis in breast cancer. In conclusion, our data may suggest that the MYC K323 lysine residue is involved with the regulation of HSPA1A and subsequent function of the spliceosome within cancer cells. Our data may also suggest that the MYC K158 lysine residue may be involved with the regulation of DKK1 and POSTN genes with a greater impact on regulating the expression of POSTN. Lastly, our data may suggest that the MYC K149 residue is important to regulating DDIT3, ribosomal protein genes, c-FOS, JUN, and KIF23. This data may prove influential if the interaction between the ATAC complex and MYC is confirmed by using the experiments explained previously. Overall, this study does prove novel in the fact that for the first time the MYC oncoprotein lysine residues (K149, 158, and 323) have distinct and important roles in gene transcription within human cancers, specifically breast cancer. For future studies, it will be important to elucidate which specific histone

acetyltransferases are responsible for acetylating which residues on MYC. It will be important to perform a variety of different ChIP-seq experiments to determine of these acetyltransferases, which are absolutely required for regulating the transcription of the genes and gene sets established in this bioinformatic analysis. Therefore, it may be the case that one or more of these lysine residues are required to be acetylated (or de-acetylated) for the transcription of these genes as well. It would be important to test this possibility by creating double or triple mutant MYC in cancer cell lines and re-running some of these same bioinformatic analyses.

METHODS

FLAG-MYC Co-IP:

The FLAG-MYC expression vector was incubated with lipofectamine 2000 according to the manufactures protocol. After incubation, HEK293 cells were transfected with FLAG-MYC expression vector at 70% confluency. After transfection the HEK293 cells were left in the cell incubator overnight. After the overnight incubation period, the cell media for the cells was changed (DMEM + 10% FBS) after washing the cells with PBS. If the cells were treated with HDACi the media supplemented contained (1M NAM, 1M NaBut, and 4mM TSA). After changing media, the cells were left in the incubator for 2 hours prior to cell lysis. The cell lysis buffer contained 150mM NaCl, 1% IGEPAL, 50mM HEPEs, 0.2 mM EDTA, 2mM B-Me, and 0.2 mM PMSF. M2 FLAG beads were used to pull-down FLAG-MYC from the lysate solution. Samples were incubated overnight with M2 beads @ 4 degrees Celsius. After incubation, samples were treated with 2x SDS and 1M DTT to deconjugate M2 FLAG beads from the FLAG-MYC protein. After pull-down FLAG-MYC and associated proteins were probed for by western blot using specific antibodies.

YEATS2 KD RNA-Seq Analysis:

72hrs after transfection, RNA was extracted, purified, and checked for quality control through RTq-PCR. The library prep kit used for NGS was NEBNext Ultra Directional RNA for illumina. The cDNA was sequenced at the UCR Genomics Core. The sequencer type was NextSeq, and the read length was High Output 1x75 (single end). At the genomics core, adaptor sequences were trimmed as a preliminary data analysis step. The reference genome used for RNA-seq analysis was the human genome, hg38, which was downloaded from the NCBI genome database. For genome indexing, HISAT2 was used. This program was also used to align the sequenced reads to the reference genome. After this, we used the summarizeOverlaps function to quantify aligned reads which overlapped with annotated ranges of interest. The read counting was performed for exonic gene regions in a strand-specific manner while ignoring overlaps among different genes. However, since the illumina single-end RNAseq protocol generates antisense reads, we made sure to customize our analysis to perform strand-specific read counting for the antisense strand. After the read counting was completed, we then normalized the counts with DESeq2 and used this data frame as the input for Gene Set Enrichment Analysis (GSEA). For the GSEA, we used the Java based GSEA program created by the Broad Institute. We ran our normalized read counts against the Oncogenic Signatures, KEGG, REACTOME, GO, and HALLMARKS Gene Sets. In addition to performing a GSEA, we also used our raw read counts to complete a Differentially Expressed Gene Analysis. To do this, we used DESeq2 and edgeR, but ended up using the edgeR output for downstream analysis. After this, we filtered the edgeR differentially expressed genes (DEGs), with a 0.5 log2FC and 5% FDR cutoff to find statistically significant DEGs.

MCF10A RNA-seq Analysis:

RNA sequencing libraries were prepared with the Kapa RNA mRNA HyperPrep kit according to the manufacturer's protocol. Briefly, 250 ng of total RNA from each sample was used for polyA RNA enrichment. The enriched mRNA underwent fragmentation and first strand cDNA synthesis. The combined 2nd cDNA synthesis with dUTP and A-tailing reaction generated the resulting ds cDNA with dAMP to the 3' ends. The barcoded adaptors were ligated to the ds cDNA fragments. A 10-cycle of PCR was performed to produce the final sequencing library. The libraries were validated with the Agilent Bioanalyzer DNA High Sensitivity Kit and quantified with Qubit. Sequencing was performed on Illumina Novaseq with S4 flow cell, 200 cycle of v1.5 reagents kit with the sequencing of 2x101. Real-time analysis (RTA) v3.4.4 software was used to process the image analysis. For RNA-seq data analysis, sequence reads that passed quality check were subjected to adaptor and polyA removal, followed by mapped using STAR v.2.7.9a with an index built from human reference genome hg38 with the inclusion of mouse c-MYC cDNA sequence and transcript/gene annotations. Strand-specific read counts of each gene annotated in RefGene, along with mouse c-MYC, were calculated by featureCounts v.1.6.4. To promote the accuracy of the gene expression estimates, genes with counts per million (CPM) > 1 in at least three samples were required for downstream analysis. Bioconductor package edgeR v.3.32.1 was used to assess the statistical significance of differentially expressed genes between compared groups based on quasi-likelihood pipeline. P-values were adjusted by the Benjamini and Hochberg method. Genes below the false discovery rate of 0.05 were considered significant. Pre-ranked Gene Set Enrichment Analysis was conducted for functional enrichment analysis with gene sets obtained from MSigDB v7.4. Genes from edgeR results were ranked by the sign of logFC in combination with -log10 of the P-value for this analysis. After the Pre-ranked Gene Set Enrichment Analysis was completed for all the MCF10a cell lines, a list of gene sets for the cell

lines was created which met the threshold of < 25% FDR q-val. Once these lists were created, they were plotted with ggplot2 v.3.3.6.

REFERENCES

Balogi Z, Multhoff G, Jensen TK, et al. HSP70 interactions with membrane lipids regulate cellular functions in health and disease. Progress in Lipid Research. 2019;74:18-30. doi:10.1016/j.plipres.2019.01.004

Block I, Müller C, Sdogati D, et al. CFP suppresses breast cancer cell growth by TES-mediated upregulation of the transcription factor DDIT3. Oncogene. 2019;38(23):4560-4573. doi:10.1038/s41388-019-0739-0

Dang CV. MYC on the path to cancer. Cell. 2012;149(1):22–35. doi:10.1016/j.cell.2012.03.003

El Khoury W, Nasr Z. Deregulation of ribosomal proteins in human cancers. Bioscience Reports. 2021;41(12). doi:10.1042/bsr20211577

Faiola F, Liu X, Lo S, et al. Dual Regulation of c-Myc by p300 via Acetylation-Dependent Control of Myc Protein Turnover and Coactivation of Myc-Induced Transcription. Molecular and Cellular Biology. 2005;25(23):10220-10234. doi:10.1128/mcb.25.23.10220-10234.2005.

Labrèche C, Cook DP, Abou-Hamad J, et al. Periostin gene expression in Neu-positive breast cancer cells is regulated by a FGFR signaling cross talk with TGFB/PI3K/AKT PATHWAYS. Breast Cancer Research. 2021;23(1). doi:10.1186/s13058-021-01487-8

Mi W, Guan H, Lyu J, et al. YEATS2 links histone acetylation to tumorigenesis of non-small cell lung cancer. Nature Communications. 2017;8(1):1088. doi:10.1038/s41467-017-01173-4

Niu J, Li X-M, Wang X, et al. DKK1 inhibits breast cancer cell migration and invasion through suppression of β-catenin/MMP7 signaling pathway. Cancer Cell International. 2019;19(1). doi:10.1186/s12935-019-0883-1

Patel JH, Du Y, Ard PG, et al. The c-MYC Oncoprotein Is a Substrate of the Acetyltransferases hGCN5/PCAF and TIP60. Molecular and Cellular Biology. 2004;24(24):10826-10834. doi:10.1128/MCB.24.24.10826-10834.2004

Portal MM, Ferrero GO, Caputto BL. N-terminal c-fos tyrosine phosphorylation regulates c-fos/ER association and c-fos-dependent phospholipid synthesis activation. Oncogene. 2006;26(24):3551-3558. doi:10.1038/sj.onc.1210137

Sehrawat U, Koning F, Ashkenazi S, Stelzer G, Leshkowitz D, Dikstein R. Cancer-associated eukaryotic translation initiation factor 1A mutants impair RPS3 and RPS10 binding and enhance scanning of cell cycle genes. Molecular and Cellular Biology. 2019;39(3). doi:10.1128/mcb.00441-18

Wang Y-L, Faiola F, Xu M, Pan S, Martinez E. Human ATAC Is a GCN5/PCAF-containing Acetylase Complex with a Novel NC2-like Histone Fold Module That Interacts with the TATA-binding Protein. Journal of Biological Chemistry. 2008;283(49):33808-33815. doi:10.1074/jbc.m806936200.

Wang Y, Zhou BP. Epithelial-mesenchymal transition in breast cancer progression and metastasis. Chinese Journal of Cancer. 2011;30(9):603-611. doi:10.5732/cjc.011.10226

Zhang C, Yang Y, Yi L, Paizula X, Xu W, Wu X. Hoxd antisense growth-associated long noncoding RNA promotes triple-negative breast cancer progression by activating wnt signaling pathway. Journal of Breast Cancer. 2021;24(3):315. doi:10.4048/jbc.2021.24.e24