

ORIGINAL ARTICLE

Metagenomic characterization of lysine acetyltransferases in human cancer and their association with clinicopathologic features

Yuanyuan Jiang¹ | Xuhui Guo^{1,2} | Lanxin Liu¹ | Shomita Rode¹ | Rui Wang^{1,3} | Hui Liu^{1,4} | Zeng-Quan Yang^{1,5} 

¹Department of Oncology, Wayne State University School of Medicine, Detroit, MI, USA

²Department of Breast Surgery, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China

³Department of Diagnostics of Chinese Medicine, Hebei University of Chinese Medicine, Hebei, China

⁴The Second Affiliated Hospital, Zhejiang University School of Medicine, Zhejiang, China

⁵Molecular Therapeutics Program, Barbara Ann Karmanos Cancer Institute, Detroit, MI, USA

***Correspondence**

Zeng-Quan Yang, Barbara Ann Karmanos Cancer Institute, 4100 John R Street, HWCRC 815, Detroit, MI 48201, USA.
Email: yangz@karmanos.org

Funding information

DMC Foundation; DOD Breast Cancer Research Program, Grant/Award Number: BC161536; DoD Prostate Cancer Research Program, Grant/Award Number: PC130259

Abstract

Lysine acetyltransferases (KATs) are a highly diverse group of epigenetic enzymes that play important roles in various cellular processes including transcription, signal transduction, and cellular metabolism. However, our knowledge of the genomic and transcriptomic alterations of KAT genes and their clinical significance in human cancer remains incomplete. We undertook a metagenomic analysis of 37 KATs in more than 10 000 cancer samples across 33 tumor types, focusing on breast cancer. We identified associations among recurrent genetic alteration, gene expression, clinicopathologic features, and patient survival. Loss-of-function analysis was carried out to examine which KAT has important roles in growth and viability of breast cancer cells. We identified that a subset of KAT genes, including *NAA10*, *KAT6A*, and *CREBBP*, have high frequencies of genomic amplification or mutation in a spectrum of human cancers. Importantly, we found that 3 KATs, *NAA10*, *ACAT2*, and *BRD4*, were highly expressed in the aggressive basal-like subtype, and their expression was significantly associated with disease-free survival. Furthermore, we showed that depletion of *NAA10* inhibits basal-like breast cancer growth in vitro. Our findings provide a strong foundation for further mechanistic research and for developing therapies that target *NAA10* or other KATs in human cancer.

KEYWORDS

breast cancer, copy number alteration, gene amplification, lysine acetyltransferase, *NAA10*

1 | INTRODUCTION

The dynamic and reversible acetylation of protein, catalyzed by lysine acetyltransferases (KATs, also known as HATs) and deacetylases, plays critical roles in a wide range of biological processes including transcriptional regulation, DNA damage repair, signal

transduction, stem cell self-renewal, cellular metabolism, and biological rhythm. Lysine acetyltransferases are a highly diverse group of enzymes. Under the action of KATs, the acetyl group from acetyl coenzyme A is co- or posttranslationally attached to either the α -amino group of the N-terminus or the ϵ -amino group of lysine residues of proteins including both histone and nonhistone

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2020 The Authors. *Cancer Science* published by John Wiley & Sons Australia, Ltd on behalf of Japanese Cancer Association.

proteins.^{1,2} To date, more than 30 KATs and KAT candidates have been identified.²⁻⁴ Based on their cellular localization, KATs are classified into types A and B: type A is mainly localized in the nucleus, and type B is localized in the cytoplasm.^{2,5} The type A KATs can be further grouped into 5 major families: (i) the GNAT (the GCN5-related acetyltransferase) family that includes KAT2A, KAT2B, and NAT8; (ii) the CBP/p300 family that contains CREBBP and EP300; (iii) the MYST family that includes KAT5, KAT6A, KAT6B, KAT7, and KAT8; (iv) nuclear receptor coactivator family including NCOA1 and NCOA3; and (v) basal transcription factors such as TAF1 and TAF1L.²⁻⁴ In general, KATs and histone acetylation are functionally linked with open chromatin and transcriptional activation.⁴ In addition to histones, cytoskeletal proteins (eg actin and tubulin) are another crucial family targeted by KATs, including KAT2A and ATAT1. Furthermore, N-terminal acetylation, catalyzed by a family of N-terminal acetyltransferases, including NAA10, NAA50, and NAA60, can influence protein properties including folding, oligomerization, stabilization, and intermolecular interactions.^{6,7}

Alterations of several KATs due to gene mutations, amplifications, deletions, and translocations have been linked directly to devastating human diseases, notably developmental disorders and cancer.^{1,4-5,8,9} Mutations in *CREBBP* and *EP300* have been identified to cause Rubinstein-Taybi syndrome that is characterized by mental retardation, growth retardation, and a particular dysmorphism.^{10,11} Dominant mutations in *KAT6A* (also known as *MOZ* and *MYST3*) cause intellectual disability syndrome.¹² The X-linked lethal human genetic disorder Ogden syndrome is caused by *NAA10* mutations.^{13,14} Previous studies also document the existence of a myriad of alterations of KAT genes in both blood and solid tumors. For example, *KAT6A* is recurrently rearranged and fused to that of *CREBBP/EP300* and other partner genes in acute myeloid leukemia.^{15,16} Recurrent amplification of the *KAT6A* and *KAT6B* genes has been identified in various solid tumors, including breast cancer, ovarian cancer, uterine cervix cancer, lung adenocarcinoma, colon and rectal adenocarcinomas, and medulloblastoma.^{9,15,17} Nuclear receptor coactivators, including *NCOA1* and *NCOA3*, are overexpressed in breast, prostate, endometrial, and pancreatic cancers where they promote tumor growth, invasion, metastasis, and chemoresistance.¹⁸

The initiation and progression of hematological malignancies and solid tumors have been associated with dysregulation of several KATs. However, our knowledge of the genomic and transcriptomic alterations of KAT genes and the clinical significance of those alterations in human cancer remains incomplete. In the present study, we undertook a metagenomic analysis of KATs in more than 10 000 cancer samples across 33 tumor types. We then focused on human breast cancer, one of the most common cancers, resulting in more than 450 000 deaths each year worldwide. We investigated the associations between recurrent copy number alteration (CNA) and gene expression level of each KAT, clinicopathologic features, and disease-free survival of patients with breast cancer. Furthermore, loss-of-function assays

identified which KAT has important roles in cancer cell growth and survival in vitro. Our studies prioritize a subset of KATs for future research focused on understanding the molecular mechanisms and therapeutic potential.

2 | MATERIALS AND METHODS

2.1 | Genomic and clinical data on TCGA and METABRIC cancer samples

Genetic and expression alteration data from 10 967 tumor samples spanning 33 tumor types in The Cancer Genome Atlas (TCGA) Pan-Cancer studies were obtained from the cBioPortal for Cancer Genomics¹⁹⁻²² (<http://www.cbioportal.org>). In the cBioPortal, the copy number for each KAT gene was generated by the Genomic Identification of Significant Targets in Cancer (GISTIC) algorithm and categorized as copy number level per gene: -2, possible homozygous deletion; -1, heterozygous deletion; 0, diploid; 1, low-level gain; and 2, high-level amplification. The relative expression of an individual gene and the gene's expression distribution in a reference population were analyzed in mRNA expression data. The reference population consists of tumors that are diploid for the gene in question. The Z score represents the number of standard deviations the expression of a gene is from the reference population gene expression. Somatic mutation data were obtained by exome sequencing.^{19,20} Breast cancer subtype and clinicopathologic information were obtained from a previous publication and extracted through cBioPortal.^{19,23} Among the 1084 breast cancer samples, 981 had intrinsic subtype data available, including 36 normal-like, 499 luminal A, 197 luminal B, 78 human epidermal growth factor receptor 2-enriched (HER2+), and 171 basal-like breast cancers.^{19,22} A detailed description of the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset can be found in the original publication.²⁴ The CNAs and normalized expression data from the METABRIC database were downloaded with permission from the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega>) under accession number EGAC00000000005 as well as from the cBioPortal for Cancer Genomics.¹⁹ In the METABRIC dataset, 1974 samples had subtype data available, including 199 normal-like, 718 luminal A, 488 luminal B, 240 HER2+, and 329 basal-like breast cancers.²⁴

2.2 | Semiquantitative PCR reactions

To assess gene expression at the mRNA level, RNA was prepared from human breast cancer cell lines and the MCF10A cell line by using an RNeasy Plus Mini Kit (Qiagen).²¹ RNA was mixed with qScript cDNA SuperMix (Quanta Biosciences) and then converted to cDNA through an RT reaction. The cDNA was then used for real-time PCR reactions. Primer sets were obtained from Life Technologies. A PUM1 primer set was used as a control. Semiquantitative RT-PCR

was carried out using the FastStart Universal SYBR Green Master (Roche Diagnostics) as described earlier.^{21,22}

2.3 | Cell culture and growth assays

The SUM cell lines were obtained from Dr Stephen P. Ethier, and the remaining cell lines were obtained from ATCC and German Collection of Microorganisms and Cell Cultures. All cell lines were tested routinely and authenticated using cell morphology, proliferation rate, a panel of genetic markers, and contamination checks. To determine the effect of NAA10 overexpression on the growth of human breast cancer in vitro, NAA10 expression was knocked down using siRNA in human breast cancer cell lines. The siRNAs were purchased from Sigma-Aldrich. We used a MISSION siRNA Universal Negative Control for the negative control group. For the transfection procedure, cells were seeded in appropriate cell culture plates and maintained overnight under standard conditions. Plate sizes, cell densities, and siRNA quantities were dependent on the cell line and the experimental setup; siRNA was transfected using the MISSION siRNA transfection reagent according to the manufacturer's protocol (Sigma-Aldrich). Expression levels of NAA10 mRNA and protein in knockdown cells were measured by quantitative (q)RT-PCR and western blot assays 48 and 72 hours after siRNA transfection, respectively. Five days after siRNA transfection, CellTiter-blue cell viability assays (Promega) were undertaken according to the manufacturer's guideline.

2.4 | Immunoblotting and Abs

Immunoblot assays were carried out as previously described.^{21,22} Whole-cell lysates were prepared by scraping cells from the dishes into cold RIPA lysis buffer. Centrifugation protein content was estimated by the Bradford method. A total of 20–50 µg of total cell lysate was resolved by SDS-PAGE and transferred onto a PVDF membrane. Antibodies used for the western blot in the study included anti-NAA10 (1:1000 GTX125971; GeneTex) and anti-β-actin (1:5000 A5441; Sigma-Aldrich).

2.5 | Statistical analysis

Statistical analyses were undertaken using R software (<http://www.r-project.org>) and GraphPad Prism (version 6.03).^{21,22} Statistical significance of the differences in mRNA expression level for each KAT among different subtypes, stages, and grades of breast cancer samples was determined using ANOVA and Welch's *t* test as described earlier.^{21,22} Spearman and Pearson correlation tests were used to correlate copy numbers and mRNA levels of each KAT from approximately 10 000 TCGA Pan-Cancer samples together, as well as 20 individual cancer types that contain at least 200 samples. We used the "cor" function in R for computation, specifying which type of test we wanted (Spearman or Pearson). Relationships between

KAT mRNA expression and disease-free survival in METABRIC breast cancer were analyzed by equally dividing 1980 samples into high and low expression groups for each KAT based on mRNA expression Z scores. Thus, there were 990 samples each with high or low expression for each KAT gene. A similar statistical method was also used for survival analysis of individual subtypes of METABRIC cancer patients.

3 | RESULTS

3.1 | Genomic alterations of KATs across 33 human tumor types

Genetic alterations, including CNA and somatic mutation, are a universal hallmark of cancer.^{25,26} We hypothesized that KATs with recurrent genetic alterations and mRNA dysregulations play important roles in different types of human tumor, and hence serve as diagnostic and therapeutic targets. Based on the current ChromoHub database, the human genome encodes 37 KAT proteins with demonstrated or predicted acetyltransferase activity (Figure S1 and Table S1).^{2,3} We first analyzed CNAs and mutations in 37 KAT genes compiled from 10 967 TCGA samples (Pan-Cancer cohort) across 33 divergent tumor types.^{19,20} These 33 tumor types include 3 lymphatic and hematologic tumors and 30 solid tumors (Table S2). The copy number for each KAT was generated by the copy number analysis algorithm GISTIC and categorized according to copy number level per gene as high-level amplification, low-level gain, diploid, shallow deletion (possibly a heterozygous deletion), and deep deletion (possibly a homozygous deletion). Somatic mutation data were obtained from exome sequencing.^{19,20} We will use the term "alteration" henceforth for gene amplification, deep deletion, and mutations, including missense mutation, gene truncation, and fusion. In the Pan-Cancer cohort, 42% of the tumors contained an alteration in at least 1 of the 37 KAT genes. Among genomic alterations, *KAT6A* (3.18%) and *NAA10* (1.78%) were the most frequently amplified, and *ESCO2* (2.68%) and *ELP3* (2.64%) were most frequently deeply deleted (Figure 1A,B). In mutations, 2 pairs of homologs *CREBBP* (4.65%) and *EP300* (4.22%), and *TAF1* (3.46%) and *TAF1L* (4.48%) were the most mutated KATs in the TCGA Pan-Cancer cohort (Figure 1C).

There is considerable variation in the KAT genes altered across different tumor types. We found that 3 KAT genes were amplified in more than 10% of individual TCGA solid tumors; *KAT6A* was amplified in 17.86% of uterine carcinosarcoma, *BRD4* was amplified in 12.06% of ovarian cancer (OV), and *ESCO1* was amplified in 10.38% of pancreatic cancer samples (Figure 1A and Table S3). The highest percentages of deep deletion were *ESCO2* in prostate cancer (6.95%) and OV (6.47%) samples (Figure 1B and Table S4). Seven KAT genes (*CREBBP*, *EP300*, *KAT6A*, *TAF1*, *TAF1L*, *MCM3AP*, and *CIITA*) had mutation frequencies greater than 10% in at least 1 tumor type (Figure 1C and Table S5). Notably, *CREBBP* was mutated in more than 10% of 4 tumor types: bladder cancer (BLCA, 13.41%), uterine corpus endometrial

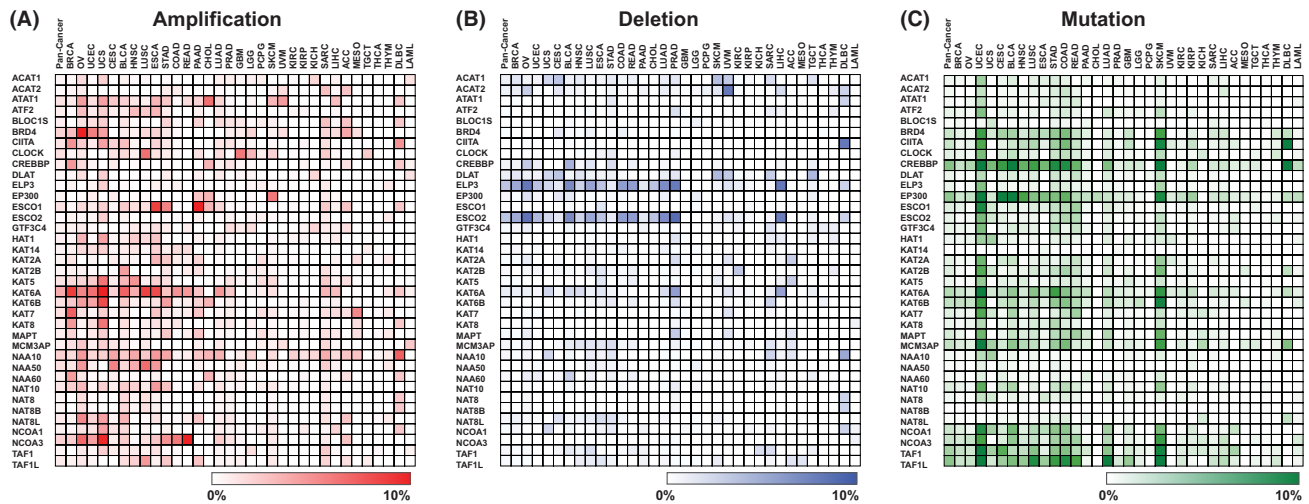


FIGURE 1 Heatmap showing the frequencies of lysine acetyltransferase (KAT) (A) amplification (red), (B) deep deletion (blue), and (C) mutations (green) across 33 tumor types from The Cancer Genome Atlas (TCGA). Heatmap was generated using Morpheus software from the Broad Institute (<https://software.broadinstitute.org/morpheus/>). The 33 TCGA tumor types are: breast invasive carcinoma (BRCA), cervical squamous cell carcinoma and endocervical adenocarcinoma (CESC), ovarian serous cystadenocarcinoma (OV), uterine carcinosarcoma (UCS), uterine corpus endometrial carcinoma (UCEC), bladder urothelial carcinoma (BLCA), esophageal carcinoma (ESCA), head and neck squamous cell carcinoma (HNSC), lung squamous cell carcinoma (LUSC), adrenocortical carcinoma (ACC), brain lower grade glioma (LGG), cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), glioblastoma multiforme (GBM), kidney chromophobe (KICH), kidney renal clear cell carcinoma (KIRC), kidney renal papillary cell carcinoma (KIRP), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), mesothelioma (MESO), pancreatic adenocarcinoma (PAAD), pheochromocytoma and paraganglioma (PCPG), prostate adenocarcinoma (PRAD), rectum adenocarcinoma (READ), sarcoma (SARC), skin cutaneous melanoma (SKCM), stomach adenocarcinoma (STAD), testicular germ cell tumors (TGCT), thyroid carcinoma (THCA), uveal melanoma (UVM), acute myeloid leukemia (LAML), lymphoid neoplasm diffuse large B-cell lymphoma (DLBC), thymoma (THYM)

cancer (UCEC, 12.77%), diffuse large B-cell lymphoma (12.20%), and melanoma (SKCM, 1 2.05%). The highest frequency of mutation in *TAF1*, *EP300*, and *TAF1L* is in UCEC (20.12%), BLCA (15.85%), and SKCM (15.45%), respectively. Taken together, among 37 KATs, several KAT genes, including *KAT6A*, *NAA10*, *CREBBP/EP300*, *TAF1/TAF1L*, *ESCO2*, and *ELP1* had relatively higher frequencies of genetic alterations in a spectrum of human tumors.

3.2 | Mutation profiling of CREBBP/EP300 and TAF1/TAF1L

The CBP/p300 (KAT3) subfamily of KATs is composed of 2 homologous enzymes, CREBBP and EP300, both containing a well-conserved acetyltransferase domain and a number of protein interaction domains (eg bromodomain and plant homeodomain) that facilitate binding with over 400 proteins.²⁷⁻²⁹ CREBBP and EP300 are key transcriptional coactivators that are essential for a multitude of cellular processes.²⁷ Both CREBBP and EP300 were highly mutated (more than 4%) in TCGA Pan-Cancer samples. In 10 436 TCGA samples with mutation data, there are 629 CREBBP mutations, including 451 missense, 148 truncating, 8 in-frame, and 22 fusion mutations. The most mutated site of CREBBP was R1446 (15 samples with missense mutation) which is located at the KAT enzymatic domain (Figure 2). For EP300, there are 556 mutations, including 378 missense, 159 truncating, 5 in-frame, and 14 fusion mutations. The

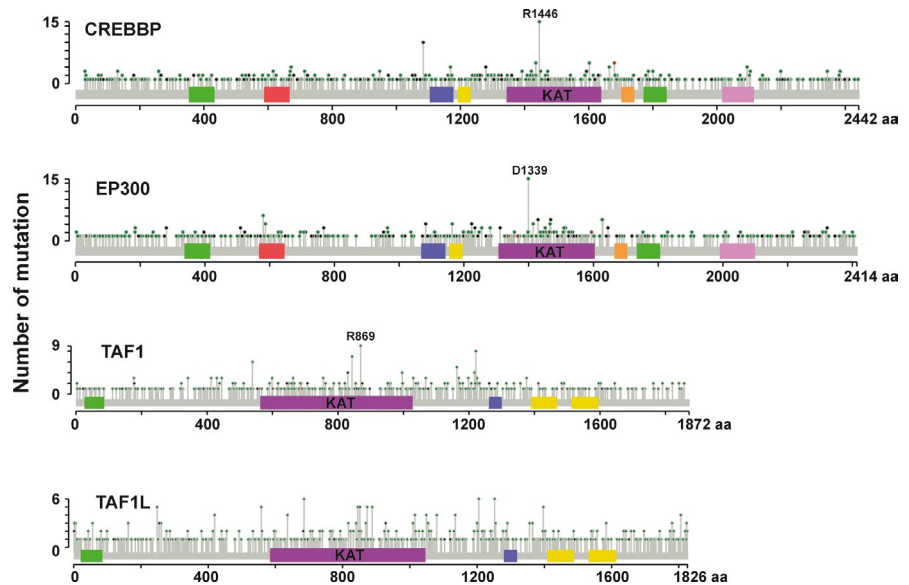
most mutated residue of EP300 was D1399, which also is located at the KAT enzymatic domain (Figure 2).

The *TAF1* gene encodes the largest subunit of the transcription factor II D complex, which promotes transcriptional initiation and activation.³⁰ The *TAF1* homologue *TAF1L* has 95% amino acid identity with *TAF1*.³¹ In TCGA Pan-Cancer, there are 477 *TAF1* mutations, including 418 missense, 54 truncating, 4 in-frame, and 1 fusion mutations. R869, which is located at the putative histone acetyltransferase domain of *TAF1*, had the highest frequency (9 mutations) of mutations (Figure 2). For *TAF1L*, there are 625 mutations, 565 missense, 58 truncating, 1 in-frame, and 1 fusion mutations. R658, R1205, and R1252 have mostly missense mutations (Figure 2). Together, 4 KAT genes (*CREBBP*, *EP300*, *TAF1*, and *TAF1L*) had higher frequencies of somatic mutation, and enriched mutation sites likely have an effect on their KAT activities.

3.3 | Copy number and expression profiling of KATs in different subtypes of breast cancer

Breast cancer is the most common cancer and one of the leading causes of cancer death among women. Using gene expression profiling, breast cancer has been classified into 5 intrinsic subtypes with distinct risks and underlying biology; these 5 subtypes are luminal A, luminal B, HER2+, basal-like, and normal-like breast cancers.^{23,32} One of the KATs, *MAPT*, is in the PAM50 (prediction analysis for microarray

FIGURE 2 Mutational spectra of *CREBBP*, *EP300*, *TAF1*, and *TAF1L* in The Cancer Genome Atlas (TCGA) Pan-Cancers. Images show protein domains and the positions of somatic mutations in *CREBBP*, *EP300*, *TAF1*, and *TAF1L* in TCGA Pan-Cancers. Red dot, nonsense mutation, frameshift deletion, insertion, or splice; green dot, missense mutation; black dot, in-frame insertion or deletion



50) gene panel for classifying the 5 subtypes of breast cancer. To determine whether the genetic alteration or mRNA expression of other KATs is specific to a breast cancer subtype, we analyzed CNA and mRNA expression independently across 5 subtypes of breast cancer samples.²⁰ The frequencies of high-level amplification, low-level gain, diploid, heterozygous deletions, homozygous deletions, and somatic mutation of 37 KAT genes in 5 TCGA breast cancer subtypes are shown in Table S6. Notably, *KAT6A* was amplified in 7.62% of luminal A, 9.64% of luminal B, 5.13% of HER2+, and 11.11% of basal subtypes of breast cancer. Additionally, *ELP3* and *ESCO2* also showed the highest frequency of homologous deletion in luminal B and basal-like subtypes (Table S6). We also noted that *CREBBP* had the highest frequency (7.02%) of mutation in basal-like breast cancer (Table S6).

Next, we analyzed the correlation between copy number and mRNA level of 37 KATs from TCGA samples (Table S7). In the Pan-Cancer analysis, 36 of 37 KATs have positive correlation between mRNA and DNA copy number in both Spearman and Pearson correlation, with 23 KATs having both Spearman and Pearson coefficients greater than 0.3. Based on the means of Spearman coefficient of all KATs, we found that *BRCA*, *OV*, lung squamous cell carcinoma, and cervical squamous cell carcinoma had relatively higher positive correlation coefficient, whereas thyroid carcinoma and acute myeloid leukemia had lower positive correlation coefficient (Table S7 and Figure S2). These data suggested that CNA could be a major factor in the dysregulation of several KAT mRNAs in human cancer, including breast cancer. We also found that different subtypes of breast cancer had different patterns of expression for each KAT gene. Consistently with previous studies, we revealed that *MAPT* was significantly underexpressed, whereas *BRD4* was significantly overexpressed in estrogen receptor-negative (ER-), basal-like breast cancer (Table S8). In addition to *BRD4*, we found that 8 KATs, including *NAA10* and *ACAT2*, were also significantly higher expressed ($P < .05$, mean Z score difference greater than 0.5) in basal-like compared with luminal subtypes, and 11 genes including *KAT6A*, *KAT6B*, *KAT7*, and *NCOA3* showed lower expression in basal-like subtype breast cancer (Figure 3A and Table S8).

To validate our findings from TCGA breast cancer dataset regarding KAT genetic alterations, we undertook an independent analysis using the METABRIC dataset, which contains 2509 breast cancer samples with long-term clinical follow-up data.²⁴ We found that *KAT6A* had the highest frequency (10.81%) of high-level amplification in the METABRIC breast cancer samples (Table S9). Among 37 KAT genes, 3 genes, *EP300*, *TAF1*, and *NCOA3*, had been selected for the mutation analysis with the frequencies of 2.75%, 2.07%, and 1.63%, respectively, in METABRIC breast cancer samples.³³ For mRNA expression, data for 33 KATs were available (mRNA data of *BLOC1S1*, *CIITA*, *NAT8B*, and *NAT8L* were not available) in the METABRIC dataset. Again, we found that 3 KATs (*NAA10*, *ACAT2*, and *DLAT*) had significantly higher expression ($P < .05$, mean Z score difference greater than 0.5), whereas 5 KATs (*MAPT*, *CREBBP*, *EP300*, *KAT6B*, *KAT8*, and *NAA60*) were underexpressed in METABRIC basal-like subtypes compared with luminal subtypes (Figure 3B and Table S10).

3.4 | Association of KAT gene expression with clinical features and survival of breast cancer patients

We next investigated the clinical relevance of KAT expressions in the METABRIC cohort, as the METABRIC breast cancers have long-term clinical follow-up data. We first examined expression levels of each KAT gene at different grades of METABRIC breast cancer samples. The means of Z score and P value for each KAT gene across grades 1-3 are shown in Table S11. We found that 3 genes, *NAA10*, *ACAT2*, and *NCOA3* were significantly highly expressed in higher-grade breast cancers (t test: grade 3 vs 1 + 2; $P < .001$ and means difference greater than 0.4; Figure 3C and Table S11). The Nottingham prognostic index (NPI), a clinicopathologic classification system based on tumor size, histological grade, and lymph node status that is widely used in Europe for breast cancer prognostication, was also available in the METABRIC cohort.³⁴ Thus, we compared expression levels of KATs between patients with high NPI (greater than 3.4) vs those with low NPI (3.4 or less). As shown

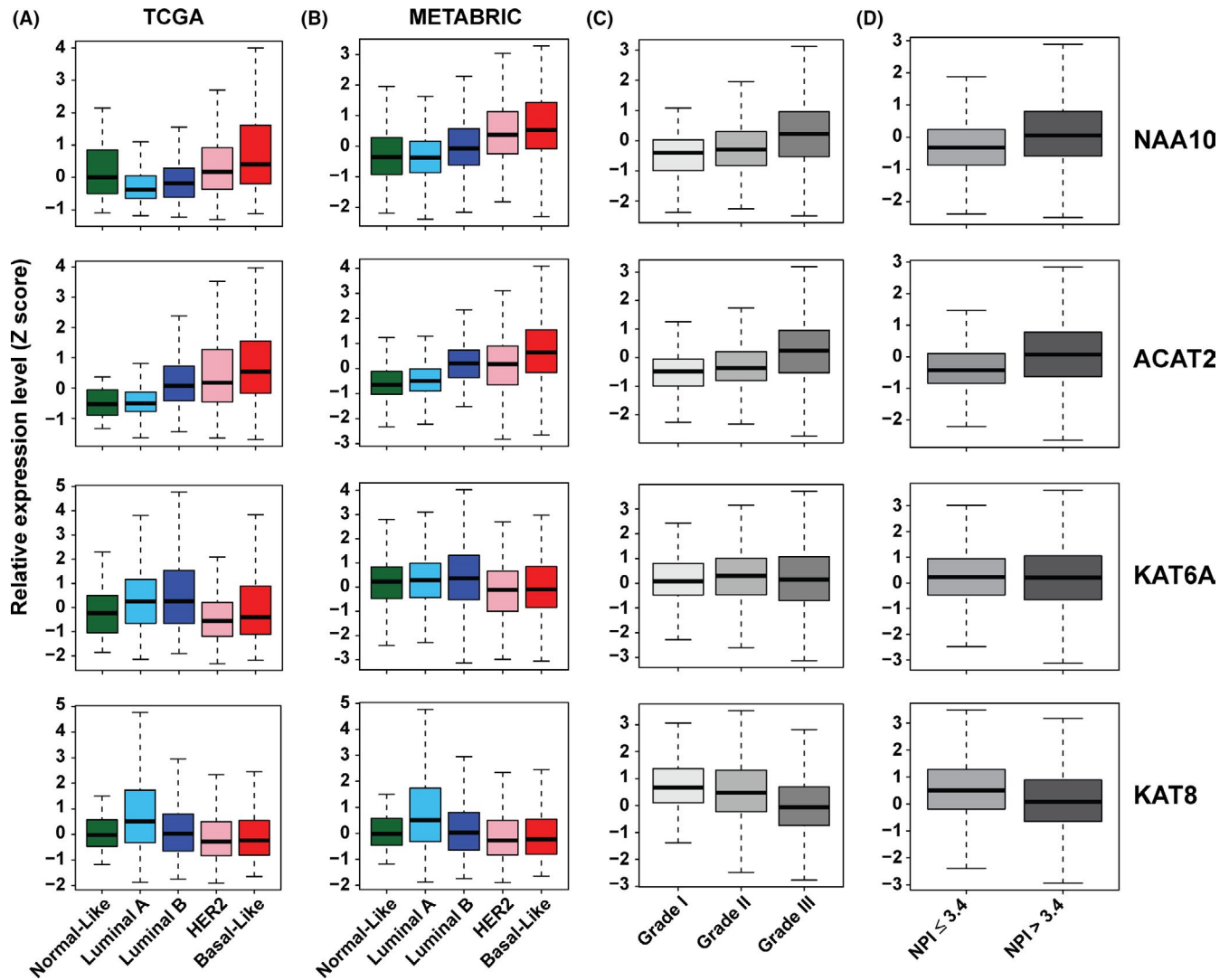


FIGURE 3 Expression levels of 4 lysine acetyltransferases (KATs), NAA10, ACAT2, KAT6A, and KAT8, in breast cancers grouped by subtype, grade, and Nottingham prognostic index (NPI). A, Expression levels of 4 KATs across 5 subtypes of The Cancer Genome Atlas (TCGA) breast cancer samples. B, Expression levels of 4 KATs across 5 subtypes of Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) breast cancer samples. NAA10 and ACAT2 were expressed at significantly higher levels ($P < .05$) in basal-like compared with luminal subtypes. C, Expression levels of 4 KATs in 3 grades of METABRIC breast cancer samples. D, Expression levels of 4 KATs in METABRIC patients with a poor prognosis (NPI > 3.4) and good prognosis (NPI \leq 3.4) scores. High expression of NAA10 and ACAT2, but not KAT6A and KAT8, were significantly ($P < .001$) associated with high grade and poor prognosis of METABRIC breast cancers

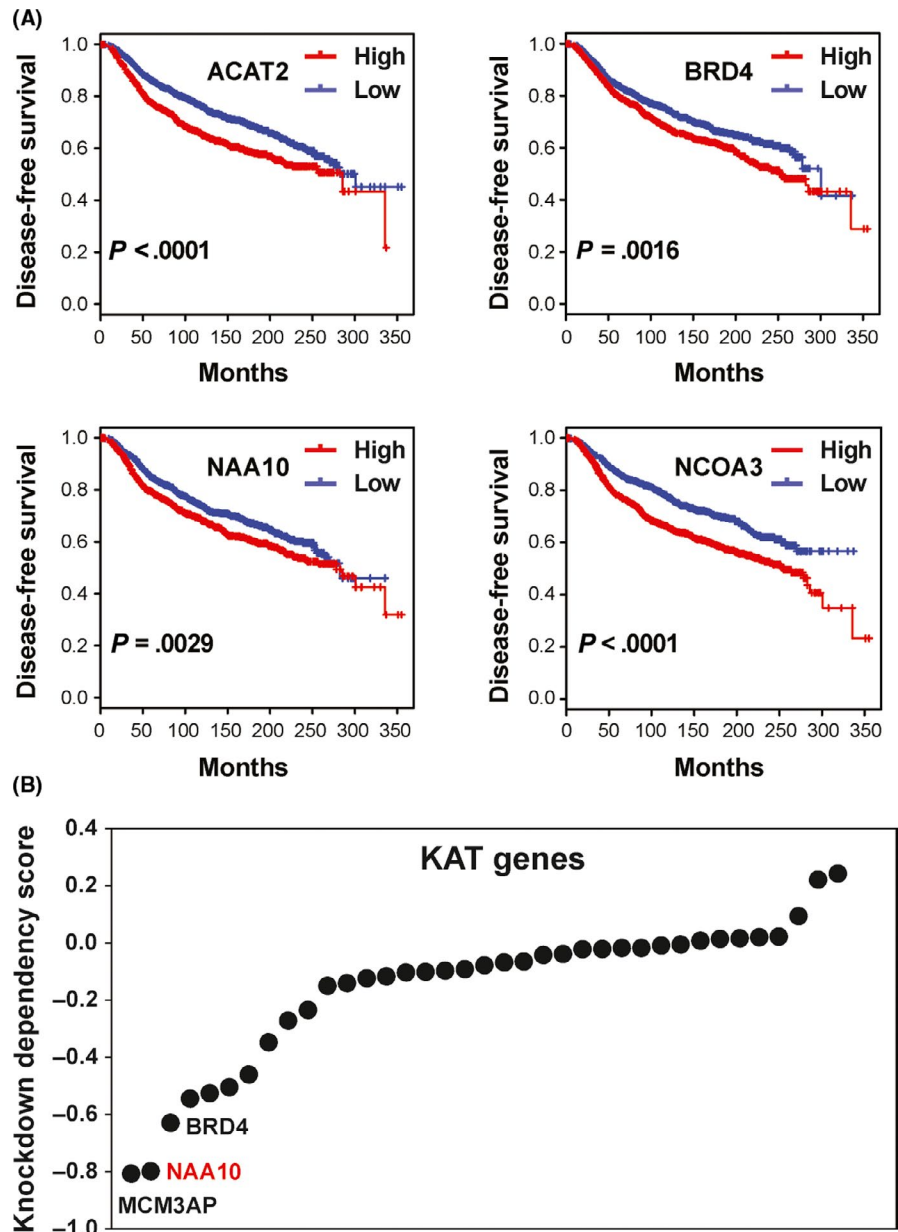
in Figure 3D and Table S11, 3 KATs, NAA10, ACAT2, and NCOA3, were significantly overexpressed, whereas *ELP3* and *MAPT* were underexpressed in samples with higher NPI ($P < .001$ and means Z score difference greater than 0.4) compared with that in lower NPI samples.³⁵ Next, we analyzed the relationship between KAT mRNA expression and disease-free survival of METABRIC breast cancer patients. We found that higher mRNA levels of 4 KATs (*ACAT2*, *BRD4*, *NAA10*, and *NCOA3*) and lower expression of 5 KATs (*MAPT*, *ACAT1*, *ELP3*, *KAT2A*, and *KAT8*) were significantly associated with shorter disease-free survival in METABRIC breast cancer patients ($P < .005$) (Figures 4A and S3, and Table S12).³⁵ We also undertook survival analysis of KATs in different subtypes of breast cancer. As shown in Table S12, we found that higher expression of *BRD4* was significantly associated with shorter disease-free survival in basal-like and luminal B subtypes, whereas lower expression of *MAPT*

and *ACAT1* was significantly associated with shorter disease-free survival in luminal A patients ($P < .05$). In summary, we found that three KATs, NAA10, ACAT2, and *BRD4*, were highly expressed in the aggressive basal-like subtype, and *NCOA3* was highly expressed in luminal B subtype, and several KATs (eg *BRD4* and *NAA10*) were significantly associated with tumor grade and disease-free survival of breast cancers.

3.5 | Potential role of NAA10 in breast cancer proliferation

The availability of multiple datasets comprising genome-scale RNAi viability screens in hundreds of diverse cancer cell lines presents new opportunities for understanding cancer vulnerabilities.^{36,37}

FIGURE 4 Lysine acetyltransferases (KATs) associated with disease-free survival of breast cancer and contributed to growth and viability of tumor cells in vitro. A, Kaplan-Meier plots of disease-free survival associated with mRNA expression levels of 4 KATs (ACAT2, BRD4, NAA10, and NCOA3) in Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) breast cancers. For the survival analysis, we equally divided 1980 METABRIC samples into high and low expression groups for each KAT based on mRNA expression Z scores. B, Scatter plot showing mean of each KAT dependency score in genome-scale loss-of-function screens of 712 tumor lines



To determine how endogenous KATs contribute to growth and viability of tumor cells, we analyzed a genome-wide loss-of-function shRNA screen in 712 tumor lines (<https://depmap.org/portal/>).^{36,38,39} We ranked average dependency (DEMETER2) scores of all 37 KATs and revealed that 6 of them, notably MCM3AP, NAA10 and BRD4, have average DEMETER2 scores of less than -0.5 (lower score means more dependency on target) (Figure 4B).⁴⁰ Given that NAA10 was overexpressed in advanced stage, higher grade, and basal-like breast cancer, we undertook a more detailed analysis of NAA10 shRNA screen data in breast cancer cell lines. Among 80 breast cancer lines, 63 lines have DEMETER2 scores of less than -0.5 , and among them, 10 lines have scores of less than -1.0 (Table S13).

Next, we carried out qRT-PCR and western blot assays to measure NAA10 in 22 breast cancer cell lines. Compared with MCF10A, an immortalized but nontumorigenic breast epithelial cell line, mRNA

levels of NAA10 were highly expressed in 13 lines, particularly the basal-like subtype lines including SUM149, MDA-MB-468, and SUM1315 (Figure 5A). Expression levels of the NAA10 protein were also higher in basal-like breast cancer lines (Figure 5B). To confirm the results of genome-wide RNAi screens in cancer cell lines, we examined the effects of knocking down NAA10 in SUM149, MDA-MB-468, and SUM1315 basal-like breast cancer lines. We obtained 3 siRNAs targeting different regions of NAA10 gene. Quantitative RT-PCR and western blot assays revealed that 2 siRNAs significantly decreased the expression of NAA10 at mRNA and protein levels (Figures 5C and S4). As shown in Figure 5C, NAA10 knockdown dramatically slowed SUM149 cell growth to approximately 40% of the growth of the nonsilenced control. NAA10 knockdown also modularly, but statistically significantly, slowed MDA-MB-468 and SUM1315 growth (Figure 5C). Our genomic, clinical, and in vitro cellular data suggest that NAA10 has a potential role in promoting breast cancer growth and progression.

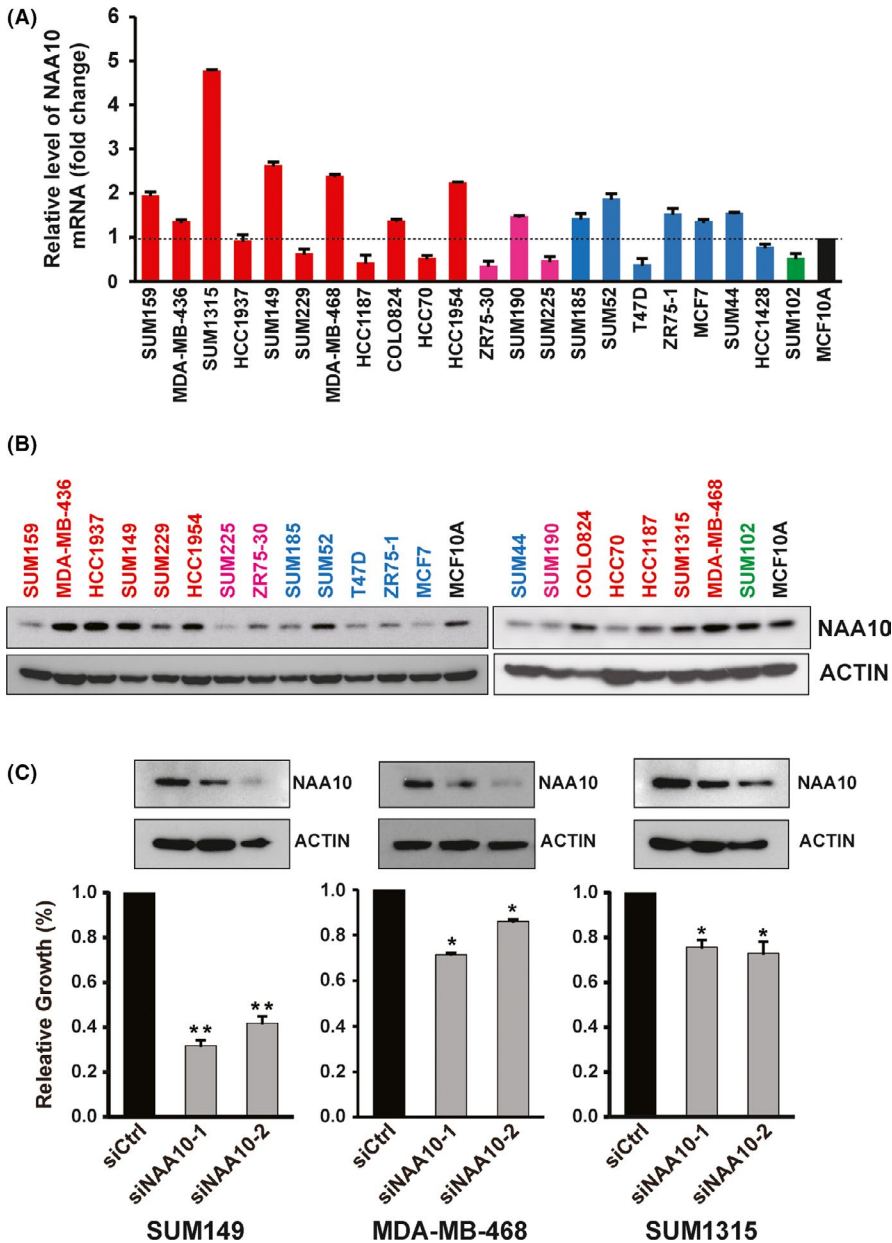


FIGURE 5 Knockdown of NAA10 inhibits cell proliferation and survival in breast cancer cell lines. A, mRNA expression levels of NAA10, measured by quantitative RT-PCR, in a panel of 21 breast cancer cell lines. mRNA expression levels in the immortalized but nontumorigenic breast epithelial cell line MCF10A cells were arbitrarily set as 1. Cell lines: black, MCF10A; green, normal-like breast cancer line; blue, luminal breast cancer cell lines; pink, HER2+ breast cancer cell lines; and red, basal-like breast cancer cell lines. B, Immunoblot analysis of NAA10 expression in a panel of breast cancer cell lines as well as MCF10A line. C, top panel: western blots showing successful knockdown NAA10 protein levels with 2 different siRNAs. Bottom panel: bar graph shows relative cell growth after knocking down NAA10 in SUM149, MDA-MB-468, and SUM1315 breast cancer cells (* $P < .05$, ** $P < .01$). Data are expressed as mean \pm SD

4 | DISCUSSION

Here, we describe the genomic and functional analysis of 37 KATs in a large cohort of human cancer primary samples and cell lines. Our Pan-Cancer analysis showed that several KAT genes, including *KAT6A*, *NAA10*, *CREBBP/EP300*, and *TAF1/TAF1L* have relatively higher frequencies of genetic alterations in various human cancers. Next, we undertook integrated genomic and transcriptomic analyses of KAT genes in breast cancer with different molecular subtypes and clinicopathologic features. We identified that a subset of KAT genes, notably *KAT6A* and *NAA10*, are commonly amplified and overexpressed in breast cancer. Different subtypes of breast cancer had different patterns of copy number and expression of each KAT. Importantly, we found that 3 KATs, *NAA10*, *ACAT2*, and *BRD4*, were highly expressed in the aggressive basal-like subtype, and their expression was significantly associated with disease-free survival.

Furthermore, we found that knockdown of NAA10 inhibits basal-like breast cancer growth, suggesting NAA10 has oncogenic potential in breast cancer.

Among 37 KATs, the *KAT6A* gene (located at 8p11.21) had the highest frequency of amplification in both TCGA Pan-Cancer and METABRIC breast cohorts. Notably, more than 5% of breast, uterine, lung squamous, esophageal, and bladder cancers have *KAT6A* amplification. These data agree with and consolidate prior reports on the genetic alterations and oncogenic potentials of *KAT6A* in human cancer.^{17,41-43} For example, Turner-Ivey et al suggested that *KAT6A* is a candidate oncogene in luminal breast cancer.⁴¹ Furthermore, Yu *et al* found that *KAT6A* activates ER α promoter through its histone acetyltransferase function.⁴⁴ In this study, we also found that *KAT6A* was highly amplified and overexpressed in a subset of ER-, basal-like breast cancers. Very recently, Baell and colleagues identified novel competitive *KAT6A* inhibitors, WM-8014 and WM-1119, which induce cell cycle

exit and senescence and are effective in preventing the progression of lymphoma in mice.⁴⁵ It will be interesting to test KAT6A inhibitors in human cancer models with KAT6A amplification in the future.

CREBBP and EP300 (CBP/p300) are the most mutated KATs in human cancer. It is possible that compensatory mechanisms exist among different KATs in different cell types and functional states. However, previous studies on KAT-null and heterozygous mice have revealed highly specific functions of individual enzymes in development, physiology, and disease.^{5,46} Studies using CBP/p300 conditional knockout mice reveal a distinct role for CREBBP and EP300 in defined cell lineages.⁴⁷ Many studies indicated that CBP/p300 function as tumor suppressors.²⁷ Furthermore, CBP/p300 might contribute to DNA repair through histone acetylation and facilitate the recruitment of DNA repair factors to the site of damage.⁴⁸ Our finding, together with others, indicate that the KAT domains of CREBBP and EP300 are hotspots for mutation in human cancer, suggesting the critical roles of KAT function of CREBBP/EP300 in regulating tumor growth and progression.²⁷

Genome-scale shRNA viability screens in human cancer cell lines revealed that, among KATs, MCM3AP, NAA10, and BRD4 have the strongest preferential dependency (Figure 4B). The *MCM3AP* gene is mutated in 2.36% of TCGA samples, with the highest frequency of mutation (10.06%) in UCEC. The majority of *MCM3AP* mutations in UCEC tumor were missense. The *MCM3AP* acetylates MCM3, a highly conserved minichromosome maintenance protein, which is involved in the initiation of genome replication.⁴⁹ Interestingly, 1 UCEC line, called EN, with the *MCM3AP* mutation has the lowest DEMETER2 score in genome-scale shRNA viability screens of human cancer cell lines (<https://depmap.org/portal/>). Therefore, it is worth investigating the functional roles of *MCM3AP* mutation in human cancer, particularly UCEC, and whether it is a novel therapeutic target for UCEC in the future. The functional roles of BRD4 in cancer have been intensively studied; BRD4 plays critical roles in superenhancer organization and oncogene expression regulation.⁵⁰ It was recently reported that BRD4 has intrinsic acetyltransferase activity, although the functional significance of this activity to the function of BRD4 remains to be further investigated.⁵¹ Nevertheless, BRD4 is currently regarded as one of the most promising new drug targets for both hematological and solid tumors.^{50,52}

N-terminal acetylation impacts the modulations of protein-protein interactions, protein subcellular targeting, folding, stabilization, and degradation.⁷ NAA10 is the catalytic subunit of the N-terminal acetyltransferase NatA complex, which includes NAA10, NAA15, and NAA50.⁷ N-terminal acetylation activity by NAA10 was found to relate strongly to human genetic disease, notably Ogden syndrome.^{13,14} Studies revealed that Naa10-null mice display partial embryonic lethality, growth retardation, brain disorders, and maternal effect lethality.⁵³ In addition, NAA10 can act independently to posttranslationally acetylate a distinct set of substrates, including histone and actin.⁷ Previous studies also showed that NAA10 regulates cell proliferation and cancer formation, DNA damage response, hypoxia, bone formation, and neuronal development in enzymatic activity-dependent and -independent manners.⁵⁴ Our functional

study revealed that NAA10 is critical for cancer cell growth and survival, particularly aggressive, basal-like breast cancer. It is important to deeply investigate NAA10 function in different subtypes of breast cancer by overexpressing NAA10 in model cells in the future.

In summary, we undertook a large-scale genomic analysis of 37 KATs in human cancer, focusing on breast cancer. We found that *KAT6A* and *NAA10* were the most commonly amplified and *CREBBP/EP300* and *TAF1/TAF1L* were most often mutated in a spectrum of human cancers. Integrated genomic, transcriptomic, and clinicopathologic data in breast cancers revealed that several KATs, including *NAA10*, *ACAT2*, and *BRD4*, were significantly associated with aggressiveness and poor prognosis of patients. Loss-of-function analysis revealed that *NAA10* had important roles in promoting cancer cell growth and survival. Our findings provide a strong foundation for further mechanistic research and for developing therapies that target *NAA10* or other KATs in human cancer. Given that many cellular factors can be regulated by KATs, and KATs act in a highly context-specific and cell type-dependent manner, care is required when choosing KAT inhibitors against a specific type of cancer.

ACKNOWLEDGMENTS

This work was partially supported by grants from the Department of Defense (DoD) Breast Cancer Program BC161536, DoD Prostate Cancer Program PC130259, and the DMC Foundation to Dr Zeng-Quan Yang. We thank Dr Stephen P. Ethier for providing the SUM breast cancer cell lines. We thank Qianhui Huang for technical contributions, and Madison Bonahoom for proofreading.

DISCLOSURE

The authors have no conflict of interest.

ORCID

Zeng-Quan Yang  <https://orcid.org/0000-0002-3658-0081>

REFERENCES

1. Drazic A, Myklebust LM, Ree R, Arnesen T. The world of protein acetylation. *Biochim Biophys Acta*. 2016;1864:1372-1401.
2. Sheikh BN, Akhtar A. The many lives of KATs - detectors, integrators and modulators of the cellular environment. *Nat Rev Genet*. 2019;20:7-23.
3. Liu L, Zhen XT, Denton E, Marsden BD, Schapira M. ChromoHub: a data hub for navigators of chromatin-mediated signalling. *Bioinformatics*. 2012;28:2205-2206.
4. Hirsch CL, Wrana JL, Dent SYR. KATapulting toward pluripotency and cancer. *J Mol Biol*. 2017;429:1958-1977.
5. Trisciuglio D, Di Martile M, Del Bufalo D. Emerging role of histone acetyltransferase in stem cells and cancer. *Stem Cells Int*. 2018;2018:8908751.
6. Aksnes H, Drazic A, Marie M, Arnesen T. First things first: vital protein marks by N-terminal acetyltransferases. *Trends Biochem Sci*. 2016;41:746-760.
7. Aksnes H, Ree R, Arnesen T. Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases. *Mol Cell*. 2019;73:1097-1114.
8. Di Martile M, Del Bufalo D, Trisciuglio D. The multifaceted role of lysine acetylation in cancer: prognostic biomarker and therapeutic target. *Oncotarget*. 2016;7:55789-55810.

9. Hu Z, Zhou J, Jiang J, et al. Genomic characterization of genes encoding histone acetylation modulator proteins identifies therapeutic targets for cancer treatment. *Nat Commun.* 2019;10:733.
10. Korzus E. Rubinstein-taybi syndrome and epigenetic alterations. *Adv Exp Med Biol.* 2017;978:39-62.
11. Roelfsema JH, Peters DJ. Rubinstein-Taybi syndrome: clinical and molecular overview. *Expert Rev Mol Med.* 2007;9:1-16.
12. Tham E, Lindstrand A, Santani A, et al. Dominant mutations in KAT6A cause intellectual disability with recognizable syndromic features. *Am J Hum Genet.* 2015;96:507-513.
13. Rope AF, Wang K, Evjenth R, et al. Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *Am J Hum Genet.* 2011;89:28-43.
14. Dorfel MJ, Lyon GJ. The biological functions of Naa10 - From amino-terminal acetylation to human disease. *Gene.* 2015;567:103-131.
15. Yang XJ. MOZ and MORF acetyltransferases: Molecular interaction, animal development and human disease. *Biochim Biophys Acta.* 2015;1853:1818-1826.
16. Borrow J, Stanton VP Jr, Andresen JM, et al. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. *Nat Genet.* 1996;14:33-41.
17. Zack TI, Schumacher SE, Carter SL, et al. Pan-cancer patterns of somatic copy number alteration. *Nat Genet.* 2013;45:1134-1140.
18. Rohira AD, Lonard DM. Steroid receptor coactivators present a unique opportunity for drug development in hormone-dependent cancers. *Biochem Pharmacol.* 2017;140:1-7.
19. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012;2:401-404.
20. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal.* 2013;6:pl1.
21. Jiang Y, Liu L, Shan W, Yang ZQ. An integrated genomic analysis of Tudor domain-containing proteins identifies PHD finger protein 20-like 1 (PHF20L1) as a candidate oncogene in breast cancer. *Mol Oncol.* 2016;10:292-302.
22. Liu H, Liu L, Holowatyj A, Jiang Y, Yang ZQ. Integrated genomic and functional analyses of histone demethylases identify oncogenic KDM2A isoform in breast cancer. *Mol Carcinogene.* 2016;55:977-990.
23. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490:61-70.
24. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature.* 2012;486:346-352.
25. Beroukhim R, Mermel CH, Porter D, et al. The landscape of somatic copy-number alteration across human cancers. *Nature.* 2010;463:899-905.
26. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet.* 2013;45:1127-1133.
27. Attar N, Kurdistani SK. Exploitation of EP300 and CREBBP Lysine Acetyltransferases by Cancer. *Cold Spring Harb Perspect Med.* 2017;7(3):a026534.
28. Dancy BM, Cole PA. Protein lysine acetylation by p300/CBP. *Chem Rev.* 2015;115:2419-2452.
29. Yu H, Jiang Y, Liu L, et al. Integrative genomic and transcriptomic analysis for pinpointing recurrent alterations of plant homeodomain genes and their clinical significance in breast cancer. *Oncotarget.* 2017;8:13099-13115.
30. Filippakopoulos P, Picaud S, Mangos M, et al. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell.* 2012;149:214-231.
31. Wang PJ, Page DC. Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Human Mol Genet.* 2002;11:2341-2346.
32. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406:747-752.
33. Pereira B, Chin SF, Rueda OM, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nat Commun.* 2016;7:11479.
34. Galea MH, Blamey RW, Elston CE, Ellis IO. The Nottingham Prognostic Index in primary breast cancer. *Breast Cancer Res Treat.* 1992;22:207-219.
35. Manning M, Jiang Y, Wang R, et al. Pan-cancer analysis of RNA methyltransferases identifies FTSJ3 as a potential regulator of breast cancer progression. *RNA Biol.* 2020;17(4):474-486.
36. Tsherniak A, Vazquez F, Montgomery PG, et al. Defining a Cancer Dependency Map. *Cell.* 2017;170:564-576.e16.
37. McDonald ER 3rd, de Weck A, Schlabach MR, et al. Project DRIVE: A compendium of cancer dependencies and synthetic lethal relationships uncovered by large-scale, deep RNAi screening. *Cell.* 2017;170: 577-592.e10.
38. Marcotte R, Sayad A, Brown KR, et al. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell.* 2016;164:293-309.
39. Meyers RM, Bryan JG, McFarland JM, et al. Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat Genet.* 2017;49:1779-1784.
40. McFarland JM, Ho ZV, Kugener G, et al. Improved estimation of cancer dependencies from large-scale RNAi screens using model-based normalization and data integration. *Nat Commun.* 2018;9:4610.
41. Turner-Ivey B, Guest ST, Irish JC, et al. KAT6A, a chromatin modifier from the 8p11-p12 amplicon is a candidate oncogene in luminal breast cancer. *Neoplasia.* 2014;16:644-655.
42. Huang F, Abmayr SM, Workman JL. Regulation of KAT6 acetyltransferases and their roles in cell cycle progression, stem cell maintenance, and human disease. *Mol Cell Biol.* 2016;36:1900-1907.
43. Yu L, Liang Y, Cao X, et al. Identification of MYST3 as a novel epigenetic activator of ER alpha frequently amplified in breast cancer. *Oncogene.* 2017;36:2910-2918.
44. Yu L, Liang Y, Cao X, et al. Identification of MYST3 as a novel epigenetic activator of ERalpha frequently amplified in breast cancer. *Oncogene.* 2017;36:2910-2918.
45. Baell JB, Leaver DJ, Hermans SJ, et al. Inhibitors of histone acetyltransferases KAT6A/B induce senescence and arrest tumour growth. *Nature.* 2018;560:253-257.
46. Yamauchi T, Yamauchi J, Kuwata T, et al. Distinct but overlapping roles of histone acetylase PCAF and of the closely related PCAF-B/GCN5 in mouse embryogenesis. *Proc Natl Acad Sci U S A.* 2000;97:11303-11306.
47. Kasper LH, Lerach S, Wang J, Wu S, Jeevan T, Brindle PK. CBP/p300 double null cells reveal effect of coactivator level and diversity on CREB transactivation. *EMBO J.* 2010;29:3660-3672.
48. Dutto I, Scalera C, Prosperi E. CREBBP and p300 lysine acetyl transferases in the DNA damage response. *Cell Mol Life Sci.* 2018;75:1325-1338.
49. Takei Y, Swietlik M, Tanoue A, Tsujimoto G, Kouzarides T, Laskey R. MCM3AP, a novel acetyltransferase that acetylates replication protein MCM3. *EMBO Rep.* 2001;2:119-123.
50. Donati B, Lorenzini E, Ciarrocchi A. BRD4 and Cancer: going beyond transcriptional regulation. *Mol Cancer.* 2018;17:164.
51. Devaiah BN, Case-Borden C, Geggion A, et al. BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat Struct Mol Biol.* 2016;23:540-548.
52. Shu S, Polyak K. BET bromodomain proteins as cancer therapeutic targets. *Cold Spring Harbor Symposia Quantit Biol.* 2016;81:123-129.

53. Lee CC, Peng SH, Shen L, et al. The Role of N-alpha-acetyltransferase 10 Protein in DNA Methylation and Genomic Imprinting. *Mol Cell*. 2017;68:89-103.e7.
54. Wild J, Schmiedel BJ, Maurer A, et al. Neutralization of (NK-cell-derived) B-cell activating factor by Belimumab restores sensitivity of chronic lymphoid leukemia cells to direct and Rituximab-induced NK lysis. *Leukemia*. 2015;29:1676-1683.

How to cite this article: Jiang Y, Guo X, Liu L, et al. Metagenomic characterization of lysine acetyltransferases in human cancer and their association with clinicopathologic features. *Cancer Sci*. 2020;111:1829-1839. <https://doi.org/10.1111/cas.14385>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.