STUDIES OF THE MOLECULAR FEATURES OF BRD4-NUT AND P300 THAT CONTRIBUTE TO CONDENSATE FORMATION AND TRANSCRIPTIONAL REGULATION

APPROVED BY SUPERVISORY COMMITTEE

Michael Rosen, PhD

Jennifer Kohler, PhD

Benjamin Tu, PhD

Mike Henne, PhD

Glen Liszczak, PhD

Rolf Brekken, PhD

STUDIES OF THE MOLECULAR FEATURES OF BRD4-NUT AND P300 THAT CONTRIBUTE TO CONDENSATE FORMATION AND TRANSCRIPTIONAL REGULATION

by

MARTYNA OLGA KOSNO

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2023

ACKNOWLEDGMENTS

I would like to thank my PhD mentor, Dr. Michael Rosen, for allowing me to work towards my doctoral degree in his laboratory. Over the years spent in the Rosen lab, I have grown to become a more independent scientist, a rigorous, quantitative analyst and a critical thinker. Most importantly, Dr. Rosen has inspired me to be a better science communicator – in my writing as well as in poster and oral presentations. He taught me the valuse of clear and concise communication of scientific data – a skill that I hope to use in my future career steps. I am also tremendously grateful to him for making the laboratory a great place to do science and for his continuous support and guidance, even in the busiest of times.

I would also like to acknowledge my Thesis Committee members: Dr. Jennifer Kohler, Dr. Benjamin Tu, Dr. Mike Henne, Dr. Rolf Brekken and Dr. Glen Liszczak, for their great advice and input during our many committee meetings and work in progress seminars. I am very grateful for their guidance throughout the graduate school years, in my multiple and always changing scientific projects.

I thank the entire Rosen laboratory for making it a great work place. Everyday I wake up, I am excited to come to the lab and I always know there is a group of fantastic scientists by my side to learn from. I am especially grateful to Lynda Doolittle and Dorothee Staber for keeping the laboratory running smoothly, even in the difficult times of the COVID-19 pandemic. I would also like to thank

Dr. Simon Currie for being a great mentor throughout the years in graduate school, helping me shape my individual career development plan and teaching me how to write a scientific manuscript.

I have spent a very fun six and a half years in the Biological Chemistry Program. I would like to thank all the friends and colleagues, the Program Chair, Dr. Benjamin Tu, and the former Program Coordinator, Carla Childers for creating such a great program atmosphere, where I always felt supported and welcome.

I also thank my parents and my sister for their continuous support throughout the years in graduate school. I would not have been able to do this without their regular phone calls across the ocean and their visits in Dallas. I am tremendously grateful for their love and for teaching me the value of education from my childhood years back in Poland.

Finally, I would like to thank my husband, Dr. Anthony Vega. His love, patience, and willingness to share his broad knowledge – both in science and in life – have been absolutely instrumental in my successful completion of this dissertation and in staying sane in the process. Through all these years he has been my best friend, my greatest supporter and a fantastic scientific mentor. He has taught me the value of clear and respectful human communication, which I have been able to use in my personal and professional life. His support and love have contributed to my successes and my happiness and for that, I am extremely thankful.

STUDIES OF THE MOLECULAR FEATURES OF BRD4-NUT AND P300 THAT CONTRIBUTE TO CONDENSATE FORMATION AND TRANSCRIPTIONAL REGULATION

Publication No._____

Martyna Olga Kosno

The University of Texas Southwestern Medical Center at Dallas, 2023

Supervising Professor: Michael K. Rosen, PhD

Aberrant formation of biomolecular condensates has been proposed to play a role in several cancers. The oncogenic fusion protein Brd4-Nut drives aberrant gene expression and forms condensates in Nut Carcinoma (NC). It has not been clear how these condensates form and whether they modulate gene expression. Here, I dissected the molecular features of Brd4-Nut and a histone acetyltransferase (HAT), p300, and analyzed their contribution to condensate formation and transcriptional changes. I determined that a minimal fragment of Nut (MIN) in fusion with Brd4 is necessary and sufficient for binding to p300, and for condensate formation. A Brd4-p300 fusion protein also forms condensates and drives a transcriptional profile similar to Brd4-Nut(MIN). The intrinsically disordered regions, transcription factor – binding domains, and HAT activity of p300 all collectively contribute to condensate formation. Conversely, only HAT activity appears to be necessary to mimic the transcriptional profile of cells expressing

Brd4-Nut. My results suggest that interaction of Brd4-Nut with p300 is important for aberrant condensate formation, and that multiple, yet distinct, regions of p300 contribute to condensate formation and transcriptional regulation.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Kosno, M., Currie, S. L., Kumar, A., Xing, C., Rosen, M. K. 2023. 'Brd4-Nut interaction with p300 is critcal for condensate formation and transcriptional regulation'. *Manuscript in preparation, available upon request*

Michel, P., Owczarek, A., **Kosno, M.**, Gontarek, D., Matczak, M., Olszewska, M. A. 'Variation in polyphenolic profile and in vitro antioxidant activity of eastern teaberry (Gaultheria procumbens L.) leaves following foliar development'. 2017. *Phytochemistry Letters*, 20: 356-364.

Michel, P., Owczarek, A., Matczak, M., **Kosno, M.**, Szymański, P., Mikiciuk-Olasik, E., Kilanowicz, A., Wesołowski, W., Olszewska, M. A. 2017. 'Metabolite profiling of eastern teaberry (Gaultheria procumbens L.) lipophilic leaf extracts with hyaluronidase and lipoxygenase inhibitory activity'. *Molecules*, 22(3): 412.

Pak, C. W., **Kosno, M.**, Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V., Rosen, M. K. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein'. Molecular Cell, 63: 72-85.

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF DEFINITIONS AND ABBREVIATIONS

Biomolecular Condensates – micro-scale compartments in cells that lack a surrounding membrane

Aberrant Biomolecular Condensates – biomolecular condensates that form exclusively in the state of disease

Fusion Oncoprotein – a protein that is a result of a genetic translocation and is related with occurrence of a cancer. Examples include: Brd4-Nut, EWS-Fli1, Eml4-Alk and many others

Super Enhancer – a group of multiple enhancer DNA elements, gathered in close genomic proximity, with high levels of transcriptional activators bound to them

Epigenetic writers – molecules that introduce various chemical modifications on DNA and histones

Epigenetic readers – the specialized domain containing proteins that identify and interpret those modifications

Epigenetic erasers – the dedicated group of enzymes proficient in removing these chemical tags

Homotypic interactions – interactions between copies of the same biological molecule

Heterotypic interactions – interactions between different biological molecules

JQ1 – a potent inhibitor of BET family of proteins, which prevents the bromodomain from binding to acetylated histone tails

C646 – a selective inhibitor of p300 histone acetyltransferase.

Hi-C – high-throughput genomic analysis technique that detects genome-wide chromatin interactions, by combining next-generation sequencing with chromosome conformation capture

NC – Nut Carcinoma

HCC2429 – patient-derived Nut Carcinoma cell line

293TRex-FlpIn – human embryonic kidney-derived cell line, supplemented in a tetracycline operon and Flp-FRT recombination sites

HeLa – immortalized tumor cell lines derived from a cancer patient, Henrietta Lacks

U2OS – immortalized cell line derived from an osteosarcoma patient

CMV promoter – Cytomegalovirus-derived promoter

Flp-FRT – site-directed recombination technology utilizing recombination between FRT (flippase recognition target) sites, mediated by Flp (flippase) recombinase

TO – Tetracycline operon

BD1 – Bromodomain 1 of Brd4

BD2 – Bromodomain 2 of Brd4

ET – Extra terminal domain of Brd4

IDR – Intrinsically Disordered Region of a protein

TF – Transcription Factor

HAT – Histone Acetyltransferase

HDAC – Histone Deacetylase

BET – Bromodomain and Extra terminal domain protein family

TAD – transactivation domain

AIL – auto-inhibitory loop

TSS – transcription start site

JAK – Janus kinase

ChIP – Chromatin Immunoprecipitation

ChIPseq – Chromatin Immunoprecipitation with sequencing

RNAseq – RNA sequencing

PROTAC – proteolytic targeting chimera

Z-plane – refers to a single, in-focus microscopy image in the z dimension, taken per field of view

H3K27Ac – histone H3 acetyl mark on lysine 27

FL – Full Length

WT – Wild Type

PCA – principal component analysis

IPA – Ingenuity Pathway Analysis

NGS – next-generation sequencing

DCC – deleted in colorectal cancer receptor

°C – degrees Celsius

µL – microliter

µm – micrometer

µM – micromolar

min – minute

mL – milliliter

ng – nanogram

CHAPTER ONE

INTRODUCTION

Biomolecular Condensates

In order to control diverse biochemical and biophysical reactions, cells form different types of compartments. In addition to canonical, membrane-bound organelles, such as the nucleus, mitochondria, endoplasmic reticulum, lysosome and others, cells have also developed another way to compartmentalize cellular biochemistry: through the formation of biomolecular condensates (Banani et al. 2017; Hyman, Weber, and Julicher 2014). These condensates are micron-sized assemblies, concentrating collections of biological molecules, for example, proteins and nucleic acids. However, unlike the canonical organelles, condensates do not have a membrane separating them from the outside environment.

Some biomolecular condensates were first discovered over a century ago in the cytoplasm of insect germ cells and referred to as “granules” (Anderson and Kedersha 2006; Ritter 1890). Since then, a plethora of different kinds of condensates has been described in cells, which have been proposed to possess diverse functions important for the cell physiology. Some of the other biomolecular condensates include: Cajal bodies – nuclear condensates found in neurons, involved in snRNP biogenesis (Lafarga et al. 2009), processing bodies (or P bodies) – cytoplasmic condensates that play important roles in RNA degradation and processing (Parker and Sheth 2007), nuclear paraspeckles – likely playing

a role in mRNA processing (Fox and Lamond 2010), PML bodies – nuclear condensates proposed to be involved in transcriptional regulation, DNA replication and epigenetic silencing (Carracedo, Ito, and Pandolfi 2011), stress granules – condensates found in the cytoplasm, which form upon stress and promote cell survival likely by condensing mRNAs, ribosomal components and other molecules (Collier and Schlesinger 1986). In addition to a growing body of evidence about the various condensates, the mechanisms of condensate formation have also been extensively studied and over the years became better understood.

Mechanisms of Biomolecular Condensates Assembly

It has been shown that biomolecular condensates are often enriched in multivalent molecules as their scaffolds (Li et al. 2012; Banani et al. 2016). Furthermore, polymer chemistry – based studies indicate that such multivalent molecules can form larger oligomers, leading to a decrease in their solubility, as long as the interactions between these multivalent molecules are stronger than the interactions between these molecules and a surrounding medium (Ruff, Dar, and Pappu 2021; Zumbro and Alexander-Katz 2021; Flory 1941; Huggins 1941). Such solubility decrease can in turn cause physical separation of phases and result in a higher volume dilute phase, with a low concentration of the multivalent molecule, and a lower volume dense phase, with a higher concentration of the molecule. It is also worth noting that condensates frequently exhibit liquid-like properties, such as rapid exchange of their components with the surrounding medium, fission

and fusion with other condensates and return to a round shape upon fusion (Banani et al. 2017; Brangwynne, Mitchison, and Hyman 2011). All of the observations and theory above suggest that multivalency-based interactions can lead to formation of some types of biomolecular condensates based on liquid-liquid phase separation model (Li et al. 2012; Case et al. 2019; Case et al. 2022; Mondal et al. 2022; Cabral, Otis, and Mowry 2022). The multivalency-based phase separation mechanism has been shown to be important in the formation of condensates involving folded protein domains (Banjade and Rosen 2014), intrinsically disordered regions (IDRs) (Pak et al. 2016; Kato et al. 2012; Lyons et al. 2023), nucleic acids (Jain and Vale 2017) and chromatin (Gibson et al. 2019).

The multivalency – based phase separation model does not account for all the molecular details in the formation mechanism of nuclear condensates involved in transcription. In the nucleus, in addition to multivalency-based interactions, it has been proposed that a positive feedback loop between transcriptional machinery elements could sometimes drive condensate formation, too (Wei et al. 2020; Henninger et al. 2021). Such condensates are often associated with super enhancers, which in turn are bound by master transcription factors, as well as epigenetic readers, writers, and erasers (Sabari et al. 2018; Ibrahim et al. 2022). Higher concentration and activity of epigenetic regulators can cause increased levels of acetylation or methylation of chromatin, in turn leading to additional recruitment of acetylated-histone or methylated-histone readers. These readers can then also recruit other elements of transcriptional machinery,

thus forming a growing nucleated condensate, which could lead to amplified gene expression. This mechanism has been shown to be responsible for condensate formation in the case of some fusion onco-proteins (Shirnekhi, Chandra, and Kriwacki 2023; Wang, Gan, et al. 2021) and FET-family transcriptional regulators (Wei et al. 2020), described in more detail below, in the “Phase Separation in Cancer” section.

Biomolecular Condensates in Transcription

The relatively new field of biology studying biomolecular condensates has been growing exponentially over the past years, with more and more types of condensates being discovered, possessing more and more novel functions. The proposal that transcription might be in part organized through biomolecular condensation attracted particular interest, of both supporters and opponents of this idea. Biomolecular condensates were first proposed to play a role in transcription in a conceptual model published in 2017 (Hnisz et al. 2017). Here, the authors provided a general biomolecular condensation-based mechanism, which accounts for formation of super enhancers and their sensitivity to perturbations as well as transcriptional patterns of bursting of enhancers and the fact that some enhancers can produce simultaneous activation of several genes at the same time. In this model, transcriptional elements would come together in close proximity and form a condensate via the physical mechanism of phase separation, involving multivalent interactions between molecules (Fig.1.1). In this way, multiple

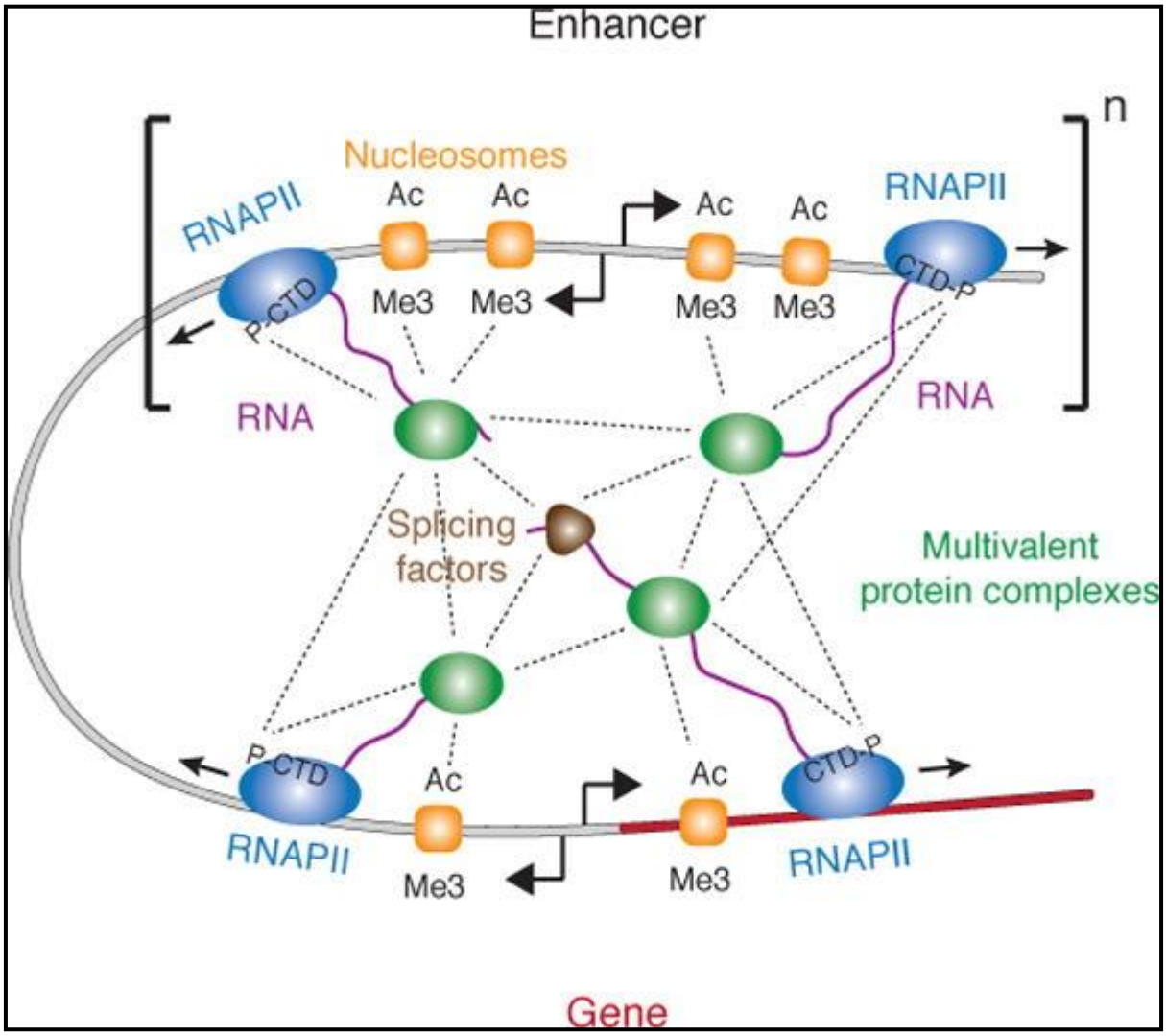

Fig.1.1: Proposed model of transcriptional control via biomolecular condensate formation. *Model from (Hnisz et al. 2017)*

components of transcriptional machinery, including RNA polymerase II, transcriptional activators, enhancers and even several genes could be involved at the same time (Fig.1.1). While attractive, this model provided no direct evidence. A breakthrough came in two studies in 2018, published in the same issue in Science (Sabari et al. 2018; Cho et al. 2018). Both groups examined condensates observed

to concentrate subunits of the Mediator complex, Brd4 and RNA polymerase II. Sabari et al. discovered that transcriptional coactivators in fact form condensates at super enhancers, thus linking biomolecular condensates and control of gene expression. In this study, the authors found that MED1 and Brd4 transcriptional coactivators both form condensates inside the nuclei of embryonic stem cells and that these condensates localize at super enhancers in DNA. The proposed mechanism here involved the general observation that many transcription factors often possess IDRs in their sequence. The authors here postulated that these IDRs might mediate weak multivalent interactions, thus assisting in condensate formation. This process in turn would help create an environment with high local concentrations of transcriptional machinery elements, facilitating gene expression, possibly at several promoter sites at a time (Sabari et al. 2018). Cho et al. also found that Mediator components and RNA polymerase II can form condensates, but they went a step further and classified these condensates into two distinct groups: small transient, and large stable condensates. Rather than focusing on the concentration of all necessary transcriptional machinery elements in these condensates, Cho et al. proposed a mechanism, where the large condensates formed at enhancer regions transiently “kiss” the transcriptional apparatus at promoter regions. This suggested ability of Mediator condensates to easily move to a new RNA polymerase II might be important for genes that are being transcribed quickly. This leads to a conclusion that it is not only the high local concentration of transcriptional factors that is important, but also for the condensate to be able to engage with multiple

transcriptional sites in a rapid manner. The fact that Mediator condensates found by the authors are relatively large (~300nm in size) might provide means for these condensates to contact transcription apparatus at multiple different gene promoters simultaneously, without the need to concentrate the multiple promoters within the condensate (Cho et al. 2018). Following these studies, many other publications have proposed similar mechanisms for different transcriptional components (Lu et al. 2018; Shrinivas et al. 2019; Guo et al. 2019; Nair et al. 2019).

All these studies provide very attractive models to increase our understanding of how transcription works at a molecular level and what are the roles of intrinsically disordered regions of transcription factors. However, the small size of transcriptional condensates (only up to 100s of nanometers) and their highly dynamic nature likely make these condensates difficult to study. These considerations have made some scientists question the reliability of data provided in studies of biomolecular condensation in transcription (McSwiggen et al. 2019; Musacchio 2022). Additional criticism has emerged due to some of the experimental methods used in studies of transcriptional condensates and extrapolation of sometimes not fully conclusive results. These include among others:

- Measurements of roundness, fusion, and fission, which might prove particularly challenging depending on the resolution of used microscope and the size of observed condensates (Cole, Jinadasa, and Brown 2011). Again, due to the particularly small size of transcriptional condensates, these

measurements would require very high-resolution techniques to produce reliable data (McSwiggen et al. 2019).

- Potential bias and inconsistencies in the process of image acquisition and image post-processing and representation, especially important in the case of cell imaging, but also in some *in vitro* systems (Gibson 2021; McSwiggen et al. 2019).
- Using 1,6-hexanediol to perturb condensates, as a proof of phase separation. This assay is in fact extremely non-specific and the chemical – especially when used in cells – can perturb functions of many enzymatic pathways that are not related with the condensates at all. Additionally, 1,6-hexanedial has been shown to perturb membrane permeability (Duster et al. 2021).
- Using fluorescence recovery after photobleaching (FRAP) as a measure of phase separation properties of condensates. While FRAP can be a useful measure of how dynamically a condensate exchanges components with the surrounding medium, it is important to note that there are several important variables that must be taken into account, which might change the outcome of a FRAP experiment. These include: the size of a condensate bleached, size of a bleached spot, bleaching an entire condensate vs. partial bleaching and analyzing the shape of a FRAP curve in context of the nature of underlying interactions driving the condensate formation (McSwiggen et al. 2019; Musacchio 2022).

- Only studying condensates *in vitro*, without additional assays in cells or organisms (Musacchio 2022). *In vitro* studies of reconstituted molecules are a very valuable tool to understand complex assemblies and processes. However, it is important to note the exact constructs used, the protein purification protocols and the relevance of the reconstitution to more complex systems, when put in the context of a cell. This is especially if the reconstituted system only consists of parts of molecules involved (such as only IDRs of transcription factors) and/or if the buffer components might be artificially increasing the propensity of the molecules to form condensates (Alberti et al. 2018).

While condensate formation has been observed in transcription, it is important to carefully examine available data to draw correct conclusions on the potential functions of such transcriptional condensates. Some of the transcriptional condensates that have been shown to likely possess important functions include condensates formed by fusion oncoproteins, in several different types of cancer.

Biomolecular Condensates in Cancer

Biomolecular condensates have been extensively studied in many different cellular contexts and have been found to play important roles in transcription, RNA processing, a variety of signaling pathways, response to stress and many other cellular processes (Hyman, Weber, and Julicher 2014; Banani et al. 2017; Baumann 2022). In addition to normal cellular functions, biomolecular condensates have also

been implicated in diseases, such as neurodegeneration and cancer. In these cases, the condensates are often referred to as “aberrant”, to reflect their pathological nature (Alberti and Dormann 2019; Wang, Zhang, et al. 2021; Taniue and Akimitsu 2022; Zbinden et al. 2020).

Some of the cancer types where aberrant condensate formation is thought to be important in cell transformation are characterized by genetic translocations, resulting in expression of fusion oncoproteins (Pavlaki 2021; Quiroga et al. 2022; Shirnekhi, Chandra, and Kriwacki 2023). There are already several noteworthy examples of such oncogenic fusion and aberrant condensate – driven cancers. These include: pediatric AML, which involves NUP98 fusion proteins (Chandra et al. 2022; Terlecki-Zaniewicz et al. 2021; Ahn et al. 2021), Ewing sarcoma, which involves FET fusion proteins, including EWS-Fli1 (Boulay et al. 2017; Ahmed et al. 2021) and myxoid liposarcoma, caused by another type of a FET fusion protein, specifically FUS-CHOP (Owen et al. 2021; Davis et al. 2021), renal cell carcinoma, which involves a NONO-TFE3 fusion protein (Wang, Gan, et al. 2021) and some forms of lung cancer, for example involving Eml4-Alk (Qin et al. 2021; Sampson et al. 2021; Tulpule et al. 2021).

The NUP98-HOXA9 fusion protein forms condensates based on multivalent interactions, mediated by its FG-rich IDR. These condensates are formed *in vitro*, using a purified protein, as well as in cell nuclei, where they associate with chromatin. The nuclear NUP98-HOXA9 condensates have been established

to form the condensates based on a mixture of homotypic and heterotypic interactions (Ahn et al. 2021; Chandra et al. 2022). FET fusion oncoproteins, for example EWS-Fli1, have been seen to form condensates for many years, but the mechanism of condensate formation has been elucidated just recently. In order to form condensates, EWS-Fli1 interacts with chromatin remodeler complex SWI/SNF and recruits it to GGAA enhancer-like DNA microsatellites. There is a specific IDR sequence within EWSR1 part of the fusion that is required for this effect (Boulay et al. 2017). On the other hand, the FUS-CHOP fusion protein does not require a specific set of residues in the IDR of FUS, but a specific IDR length is needed for it to form condensates. Similarly to other oncogenic nuclear fusion proteins, also FUS-CHOP recruits additional elements of transcriptional machinery, including Brd4 and SWI/SNF remodeler complex, suggesting that the positive feedback loop mechanism of condensate formation might be involved (Owen et al. 2021; Davis et al. 2021). NONO-TFE3 fusion oncoprotein forms nuclear condensates through a positive feedback loop mechanism that involves one of the IDRs of NONO and the NRF1 transcription factor (Wang, Gan, et al. 2021). Finally, Eml4-Alk causes cytoplasmic condensates formation, where a trimerization domain and hydrophobic HELP motif of Eml4 are required, in addition to the kinase domain of Alk. Previous research has shown that the multivalent nature of these molecules causes recruitment of additional elements such as GRB2, GAB1, SOS1 and RAS-GTP, suggesting that the Eml4-Alk condensates might play a role of “signaling hubs” (Tulpule et al. 2021).

In the examples above, the studies have shown that formation of fusion protein – based aberrant condensates is correlated with altered transcription and in some cases, oncogenic transformation. Nut Carcinoma is another example of a cancer that is triggered by expression of a fusion protein: Brd4-Nut. And similarly to other fusion protein – driven cancers, Brd4-Nut can also form large, aberrant condensates in cells. In my dissertation project, I set out to study potential links between aberrant condensate formation and transcriptional changes in cells expressing the Brd4-Nut fusion protein.

Brd4-Nut Fusion in Nut Carcinoma

Nut Carcinoma (NC) is an aggressive and poorly differentiated type of cancer. It is a rare disease, with a total of only about 20-30 new cases every year in the United States (Liu and Ferzli 2018). It used to be mostly observed in the midline organs (head, neck, and trunk), but with improvements in diagnostic technologies as well as increased awareness of the disease, Nut Carcinoma has been identified more accurately over the past years. As a result, the disease has been diagnosed more frequently in kidneys, endometrium, ovaries, and many other organs of the human body (Huang et al. 2019). The disease can develop at any age and has a very poor prognosis, with most patients surviving less than one year from the time of diagnosis (Huang et al. 2019). Because of lack of deep understanding of the molecular mechanism of Nut Carcinoma, the field still does not offer any targeted treatment options for patients, most likely due to lack of thorough

understanding of how Brd4-Nut fusion leads to tumorigenesis (Virarkar et al. 2021). The treatments proposed to date involve mostly non-specific options such as surgery, chemotherapy, and radiotherapy. In addition, there are a few small molecule inhibitors currently in clinical trials, including BET inhibitors (Bromodomain and Extra-Terminal domain protein family inhibitors, e.g. ZEN3694 compound in trial NCT05372640 or GSK525762 compound in trial NCT01587703; (Piha-Paul et al. 2020)) and HDAC inhibitors (Histone Deacetylase inhibitors, e.g. CUDC-907 compound in trial NCT02307240; (Jung et al. 2019)); more detailed molecular mechanisms explained below. Some of these molecules show promising results in treating Nut Carcinoma, but unfortunately they are also often highly toxic and have a potential for patients to develop delayed drug resistance (Wang et al. 2020).

On a molecular level Nut Carcinoma is characterized by chromosomal translocations that involve the gene of *NUT* on chromosome 15 and a partner gene. In about 75% of known cases that partner gene has been reported to be *BRD4* on chromosome 19 (French et al. 2003). Interestingly, most of the remaining 25% of patients harbor either a *BRD3-NUT* or *NSD3-NUT* fusion. Brd3 is another member of the same Bromodomain and Extra-Terminal domain protein family (BET) as Brd4. Nsd3, on the other hand, is a direct interactor of Brd4 (Wang and You 2015). Therefore, it is predicted that regardless of the heterogeneity of fusion oncoproteins involved, the still unknown molecular mechanism through which all the Nut-fusion oncoproteins cause Nut Carcinoma, might be related in most patients. Additionally, existing data suggest that the Nut-fusion proteins are a single driver

of the carcinoma, with none or few additional oncogenic mutations observed in patients (Lee et al. 2017). This makes the Nut-fusion proteins a potentially attractive therapeutic target. Because of the higher prevalence of the Brd4-Nut fusion over the other types of fusion oncoproteins observed in Nut Carcinoma, my studies described in this dissertation are focused on Brd4-Nut as a model fusion oncoprotein.

Brd4 is a transcriptional regulator, expressed in healthy human cells in two isoforms: long and short. Both isoforms have identical N-terminal regions consisting of two bromodomains and extra-terminal domain (Brd4 belongs to the BET family of proteins). The long isoform additionally includes a long, disordered C-terminal tail. The two bromodomains of Brd4 bind to acetylated histone tails and the extra-terminal domain interacts with additional elements of transcriptional machinery, such as NSD3, JMJD6, CHD4, ATAD5 or GLTSCR1 (Wang et al. 2012; Rahman et al. 2011). This allows Brd4 to maintain transcriptional activation through direct recruitment of these transcriptional modulators but also through indirect recruitment of several larger multiprotein complexes. In this multi-modal way, through both the bromodomains and the extra terminal domain activity, Brd4 assists in transcriptional regulation (Rahman et al. 2011; Wang et al. 2012). Through its role in transcription, Brd4 has been shown to play important roles in processes such as development, cancer progression and virus-host pathogenesis (Rahman et al. 2011; Donati, Lorenzini, and Ciarrocchi 2018). The high level of Brd4's transcriptional involvement can, however, also be deleterious to human

health. Brd4 has been implicated in many types of cancer, where its dysregulated expression can contribute to aberrant chromatin remodeling and transcriptional changes that mediate tumor – associated inflammation, tumor initiation, progression and metastasis (Marazzi et al. 2018; White, Fenger, and Carson 2019).

It has been previously shown that Brd4 can form intranuclear condensates, which are associated with super enhancers (Sabari et al. 2018). The protein

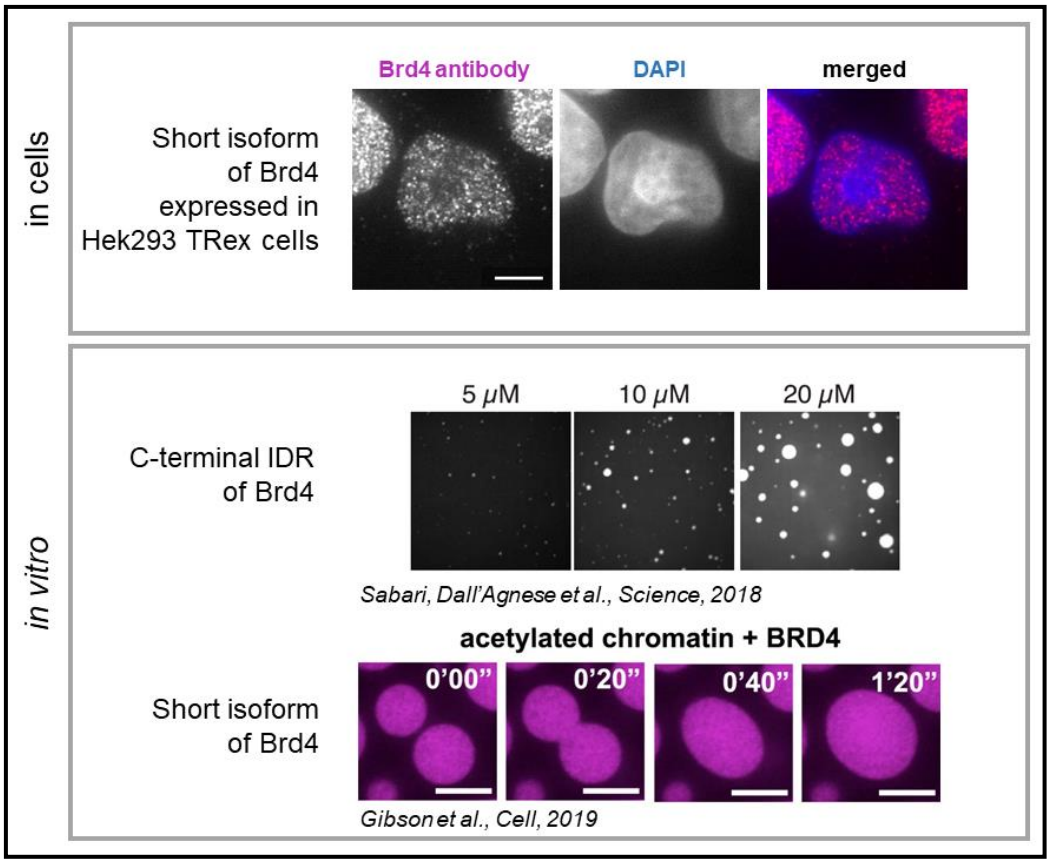

Fig.1.2: Brd4 forms small condensates when expressed in mammalian cells and phase separates in vitro: (in cells) *Brd4 protein was expressed in Hek293 TRex FlpIn cells and then immunostained with an anti-Brd4 antibody. The protein forms small condensates which are localized to the cell's nucleus.* ***(in vitro, top)*** *the C-terminal IDR of the long isoform of Brd4 was purified and showed to form condensates at different protein concentrations in a test tube (Sabari et al. 2018).* ***(in vitro, bottom)*** *the short isoform of Brd4 was purified and showed to form condensates when mixed with acetylated chromatin. The condensates observed can merge over time, showing that they exhibit liquid-like properties (Gibson et al. 2019).*

has also been shown to form liquid-like phase separated droplets *in vitro* (Fig.1.2) (Sabari et al. 2018; Gibson et al. 2019). In the chromosomal fusion observed in Nut Carcinoma, the translocation breakpoint makes the gene of *BRD4* fuse with *NUT*, such that the amino acids 1-719 of Brd4 are retained (approximately equivalent to the short isoform of Brd4). In this way both bromodomains and the extra-terminal domain of Brd4 remain intact and functional in the fusion oncoprotein (Fig.1.3). The function of bromodomains of Brd4 can be affected by small molecule bromodomain and extra terminal domain (BET) inhibitors. BET inhibitors act as competitive binders of the bromodomains, in the exact binding pocket of acetylated histone tails. In Nut Carcinoma, a BET inhibitor molecule can disrupt Brd4-Nut binding to acetylated histone tail and thus, cause detachment of the fusion protein from acetylated chromatin.

Nut is a much less extensively studied protein. In a healthy human body, Nut is expressed exclusively in male testes, in a tightly regulated manner. There, it plays an important role in spermatogenesis (Shiota, Barral, et al. 2018). The precise mechanism of its involvement in this complex process is still not fully understood, but it has been previously shown that Nut can interact with a histone acetyltransferase (HAT), p300 (Reynoird et al. 2010). It was later proposed that through this interaction, Nut might have an indirect role in the regulation of hyperacetylation of histone tails in male germ cells. This process in turn causes recruitment of Brdt, a testes-specific member of the BET protein family, which in turn orchestrates the histone-to-protamine transition – a process during which histone

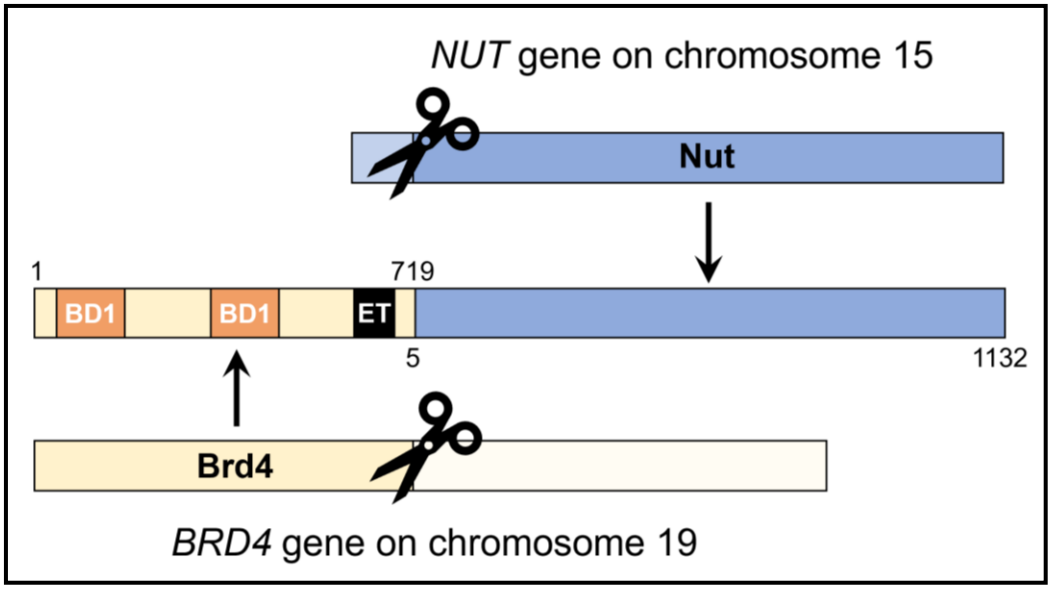

Fig.1.3: Chromosomal translocation in Nut Carcinoma: *Illustration of the chromosomal translocation between NUT and BRD4 and the resulting fusion oncoprotein product. The numbers show the retained amino acids in each of the fusion partners. BD1 – bromodomain 1, BD2 – bromodomain 2, ET – extra-terminal domain*.

variants and specific histone modifications as well as chromatin remodelers help modulate chromatin compaction. Histone-to-protamine transition is necessary to prevent mutagenesis or damage in the paternal genome. In this transition, the majority of histones are first replaced by histone variants that are specific to testis. Next, transition proteins replace these histone variants and finally, protamines replace the transition proteins, to assist in packing the genome into a highly condensed form (Wang et al. 2019). This transition is necessary for proper spermatogenesis. Knockout or inactivation of either Nut or Brdt results in male sterility due to spermatogenesis arrest (Shiota, Barral, et al. 2018; Wang et al. 2019). Nut is not known to have any additional functions in other cell types, and its expression is silenced elsewhere.

The genetic fusion of *NUT* with *BRD4* allows to overcome the silencing and upon translation, the fusion gene leads to expression of a large fusion protein. The most common oncogenic fusion of Brd4-Nut includes amino acids 5-1132 of Nut (Fig.1.3). The secondary structure of Nut is unknown, but secondary structure prediction algorithms indicate that the protein lacks a defined structure throughout most of its sequence, with short stretches of predicted α-helices in the N-terminal half (Fig.1.4). Nevertheless, none of the predicted folded parts of Nut resemble any known protein domains. Within the predicted α-helical folded parts, there is a p300 interaction motif (residues 355-505 of Nut) (Reynoird et al. 2010; Ibrahim et al. 2022). Furthermore, recent studies have determined that within this 150-residue long fragment of Nut, there are two transactivation domains (TADs), that can directly bind to p300: one spanning residues 403-418 of Nut (TAD1) and the other with residues

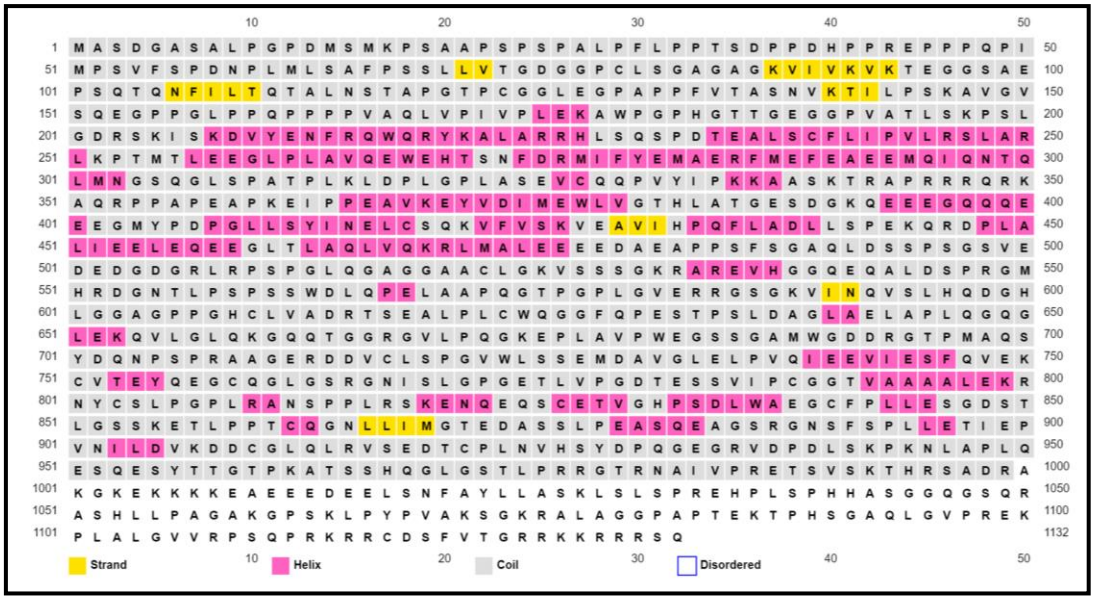

Fig.1.4.: Secondary structure prediction of Nut protein, PsiPred.

419-470 of Nut (TAD2) (Yu et al. 2023). The structural elements of p300 are described in more detail below, in the section “Potential roles of p300 in Nut Carcinoma”. Here I would like to focus on the TAZ2 domain of p300, which has been shown to interact with TAD1 and TAD2 of Nut. The TAZ2 domain of p300 contains four α helices (Fig.1.5A, green). TAD1 fragment of Nut within the p300-interaction motif also retains a short α-helical structure, where residues Y405, L410, Y413 and I414 interact with the hydrophobic patch of TAZ2 (Fig. 1.5A). TAD2 of Nut forms two anti-parallel α helices (α1’, residues 433-440 and α2’, residues 448-460). Helix α1’ binds to TAZ2 in a similar way to TAD1, with side chains of residues F435, L436 and

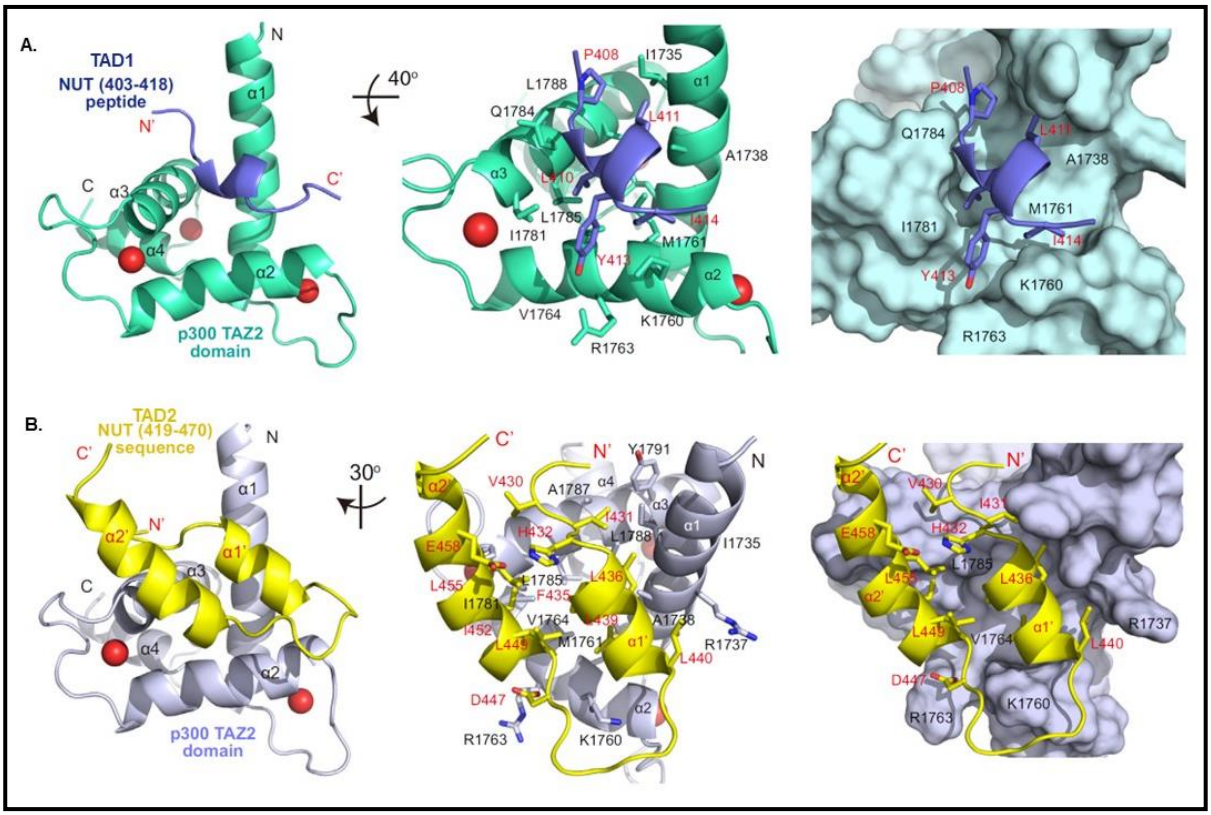

Fig.1.5.: Structures of p300 TAZ2 domain in complex with Nut TAD1 (A) and TAD2 (B). A. *Figure adapted (Yu et al. 2023)*

L440 located in the binding pocket of TAZ2. Helix α2' interacts with two other helices of TAZ2 as well as with α1' helix of TAD2. In this way, both helices of TAD2 establish a more extensive contact with p300's TAZ2 domain (Fig.1.5B).

While several different isoforms of the Brd4-Nut fusion protein can be found in some patients, all the isoforms discovered to date retain the part of Nut that has been shown to interact with p300 (Reynoird et al. 2010; Thompson-Wicking et al. 2013). It has been observed in the past that both Brd4 and Brd4-Nut can form liquid-like intranuclear condensates (Fig.1.1), (Sabari et al. 2018; Gibson et al. 2019). These condensates are sometimes larger than 2 megabase in size and have been often referred to as "megadomains" in the past studies (Alekseyenko et al. 2015). The megadomains have been proposed to be a result of an aberrant feed forward loop mechanism, involving high levels of acetylation and binding to acetylated histone tails (Alekseyenko et al. 2015; Alekseyenko et al. 2017). Additionally, proteomic studies have uncovered a plethora of transcription – related proteins to be recruited into the Brd4-Nut condensates, potentially strengthening the positive feedback loop responsible for the condensate formation. These transcriptional machinery elements include among others: ZNF532, ZMYND8, ZNF687, ZNF592, Brd4, NSD3 (Alekseyenko et al. 2017; Shiota, Elya, et al. 2018). Interestingly, many of these proteins have also been observed in fusion with Nut in some patients, indicating that the general mechanism of how Nut Carcinoma arises might be related in most clinical cases (Shiota, Elya, et al. 2018; French et al. 2014). Furthermore, while the genomic locations of the megadomains seem to be specific to a cell line

used, by majority in Nut Carcinoma – derived cell lines they associate with cancer – related genes, such as *MYC*, *TP63* and *SOX2* (Moreno, Saluja, and Pina-Oviedo 2022; Wang et al. 2014; Eagen and French 2021). In addition to these oncogenes being abnormally activated in Nut Carcinoma cells, other genes, responsible for cell differentiation are suppressed, leading to formation of a largely undifferentiated tumor tissue (Schwartz et al. 2011; Yan et al. 2011). Recent studies using Hi-C method, which allows to map distant chromatin contacts, uncovered that the Nut Carcinoma – associated megadomains contain hyperacetylated chromatin and allow chromatin interaction over megabase distances, even between separate chromosomes (Rosencrance et al. 2020). These same scientists also discovered that Brd4-Nut degradation using a small molecule proteolytic targeting chimera (PROTACs) abrogates these contacts (Rosencrance et al. 2020). It has also been shown that targeting the Brd4-Nut-p300 axis in Nut Carcinoma cells leads to dissolution of the condensates and cell differentiation (Morrison-Smith et al. 2020). All these results together indicate that in addition to Brd4-Nut, p300 acetyltransferase activity is also indispensable for both condensate formation and tumor-related transcriptional changes in Nut Carcinoma cells.

Potential Roles of p300 in Nut Carcinoma

Brd4-Nut condensates are enriched in histone marks that have been associated with active transcription, including H3K27Ac (Rosencrance et al. 2020; Alekseyenko et al. 2015). Additionally, the condensates have previously been

correlated with driving the expression of some oncogenic targets, including *MYC, TP63* and *SOX2* in Nut Carcinoma cell lines (Alekseyenko et al. 2015). The high level of acetylation within the condensates is a result of the recruitment of p300 (Rousseaux, Reynoird, and Khochbin 2022; Morrison-Smith et al. 2020; Eagen and French 2021), a histone acetyltransferase protein responsible for about 30% of all nuclear acetylation (Weinert et al. 2018).

The architecture of p300 comprises a HAT domain, a series of different transcription factor and protein interaction motifs (among others: bromodomain,

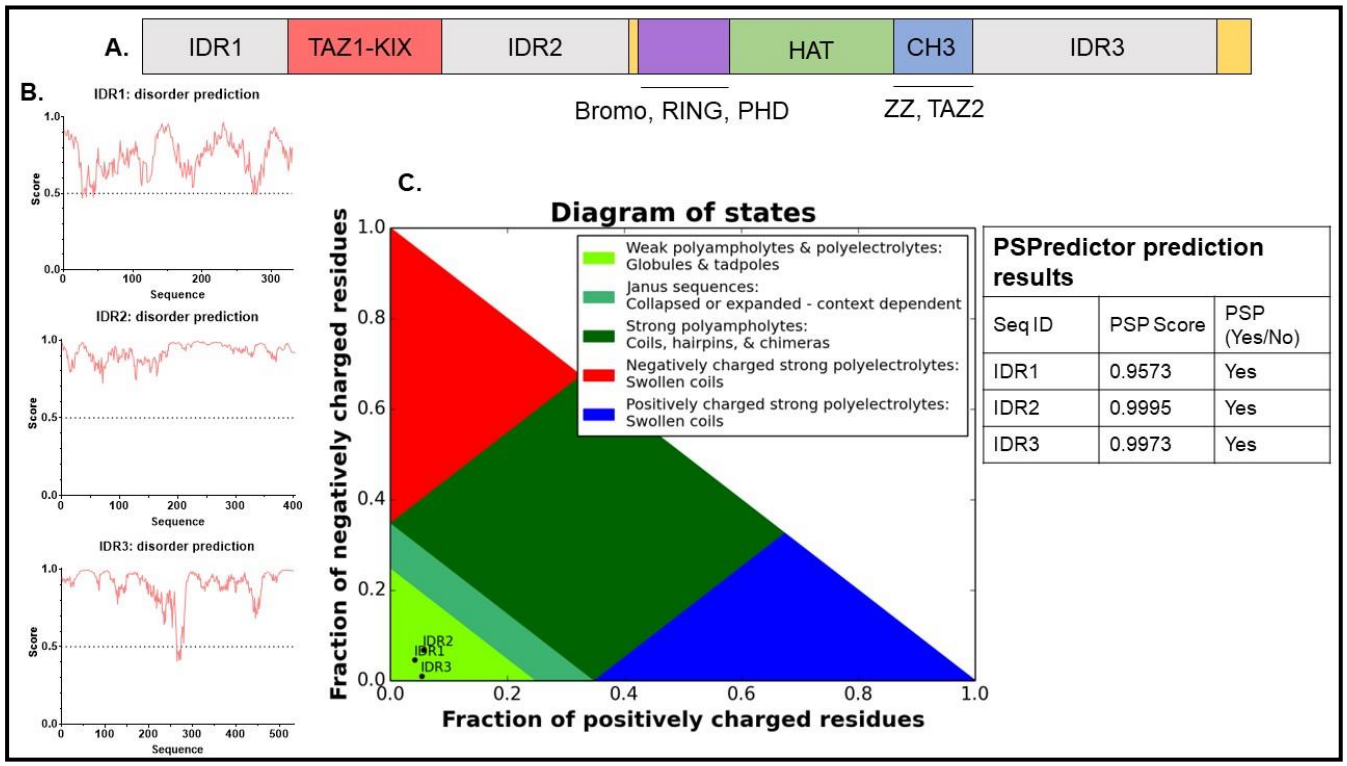

Fig.1.6.: p300 protein architecture and the description of its IDRs: A. *Cartoon of Brd4-p300 with all its domains and IDRs described.* ***B.*** *IUPRED2 predictions of disorder for the IDR1, IDR2 and IDR3 fragments shown in A.* ***C.*** *Left: CIDER diagram showing that all three: IDR2, IDR2 and IDR3 are classified as weak polyampholytes and polyelectrolytes. Right: PSPredictor results showing that all three: IDR2, IDR2 and IDR3 are predicted to phase separate.*

RING domain, PHD domain, ZZ, TAZ1, TAZ2 and others), here collectively termed “TF-binding domains”, as well as three long stretches of residues that are predicted to lack a defined secondary structure, as shown by the disorder prediction software, IUPRED2 (Fig.1.6A,B). These intrinsically disordered (IDR) fragments of p300 are predicted to behave as polyampholytes or polyelectrolytes and to be prone to phase separate. This is shown via CIDER – a tool that helps analyze sequences of intrinsically disordered proteins. Based on the fractions of positively and negatively charged residues, as well as the patterning of oppositely charged residues, CIDER categorized IDPs into five different categories, depicted in the graph in figure 1.6C. In this way, I was able to categorize IDR1, 2 and 3 of p300 as polyampholytes or polyelectrolytes (Holehouse et al. 2017). Additionally, I used a machine learning – based online tool, called PSPredictor, to assess whether the three IDRs of p300 are prone to phase separate. PSPredictor was trained based on a large database of protein sequences that are prone to form condensates. The tool available online makes a prediction on whether a given sequence can phase separate, offering a simple yes or no answer, along with a score from 0 to 1, establishing the more quantitative measure of probability of phase separation. All three IDRs of p300 are predicted to phase separate and are assigned a high score of 0.9573 to 0.9995 (Fig.1.6C).

Finally, it has been shown that p300 histone acetyltransferase must be trans-autoacetylated within its autoinhibitory loop (AIL) to be activated and to perform its function (Fig.1.7). This is possible when two molecules of p300 come in close

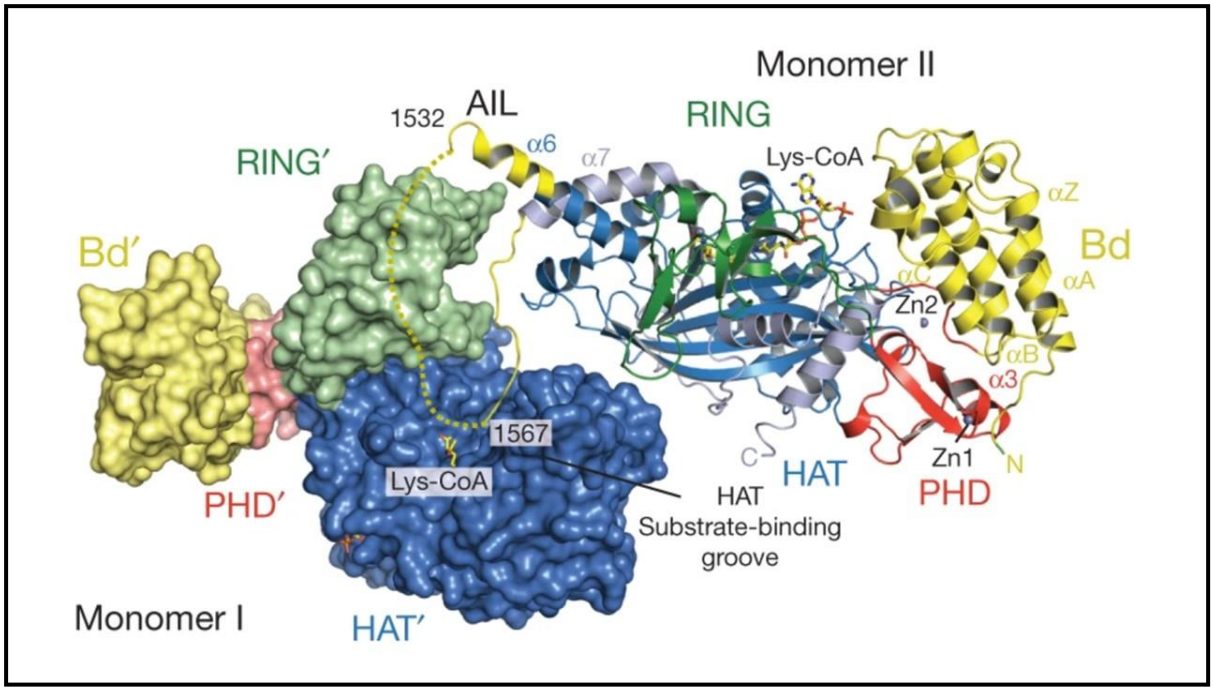

Fig.1.7.: Partial structure of p300, with the conformation of its AIL in a formed dimer: Surface of *Monomer I and cartoon of Monomer II. The AIL loop of monomer II is in yellow and is located in proximity to the HAT substrate-binding pocket of monomer I. Figure adapted from (Ortega et al. 2018).*

proximity (Fig.1.7), (Ortega et al. 2018). When two p300 molecules are close to one another, their HAT domains are close to one another (blue domains in Fig.1.7). Then, AIL residues 1520-1532 can adopt a helical structure, locating itself closely to the RING domain of the second p300 molecule (Fig.1.6). At the same time, residues 1566-1581 of the other monomer can move away from its own HAT domain and get accommodated within the first monomer's HAT binding pocket (Fig.1.7) (Ortega et al. 2018). This complex interaction between two p300 monomers can be achieved through an interaction with a dimerizing transcription factor (Ortega et al. 2018). An example of such dimerization would be STAT1 transcriptional

activator, which directly interacts with CBP/p300. STAT1 becomes activated upon Janus Kinase (JAK) – mediated phosphorylation, which induces STAT1's dimerization (Levy and Darnell 2002). The structure of STAT1 bound to CBP has been previously solved and scientists found that STAT1 binds to CBP TAZ2 domain, similarly to the interaction found between Nut and p300 (Fig.1.5), (Wojciak et al. 2009). In a more recent study, a detailed mechanism of p300 activation upon binding to STAT1 has been elucidated. Like explained above, in order to become active, p300 needs to be trans-autoacetylated and this autoacetylation can only be achieved when STAT1 is found in its dimeric form. Only STAT1 binding to p300 TAZ2 is not sufficient to result in p300 activation and the same proved right for another dimerizing transcription factor, too (Ortega et al. 2018). Without a transcription factor dimerization, p300 trans-autoacetylation within its AIL (and thus, activation) is physically hindered by its RING domain when the protein is found in a monomeric state. When the RING domain is removed, p300 becomes hyperactivated and starts forming nuclear condensates when transiently overexpressed (Fig.1.8). If, in addition to the RING domain deletion, the protein also harbors a HAT-inactivating point mutation (D1399Y) (Ito et al. 2001), it no longer forms condensates (Fig.1.8). These findings indicate that activation of HAT, with lack of structural inhibition, is required for p300 condensate formation (Fig.1.8) (Ortega et al. 2018). Additionally, treatment with a HAT inhibitor or a bromodomain inhibitor results in a decreased p300 condensate formation, pointing to the importance

of hyperacetylation and bromodomain–acetylated histone interaction in driving the p300 condensate formation.

Given that Nut binds to p300 through TAZ2, similarly to other transcription factors, it is possible that Brd4-Nut can also induce p300 trans-autoacetylation. While it is not known if Brd4-Nut can dimerize to physically bring two molecules of p300 in close proximity, an alternative mechanism could be employed here. A positive feedback loop mechanism, engaging Brd4 bromodomain binding to acetylated histone tails, Nut interacting with p300 and finally, p300 causing hyperacetylation of nearby histones, could result in bringing multiple molecules of p300 in physical proximity, resulting in the trans-autoacetylation and thus, their activation. Furthermore, this mechanism could also be at least partly responsible

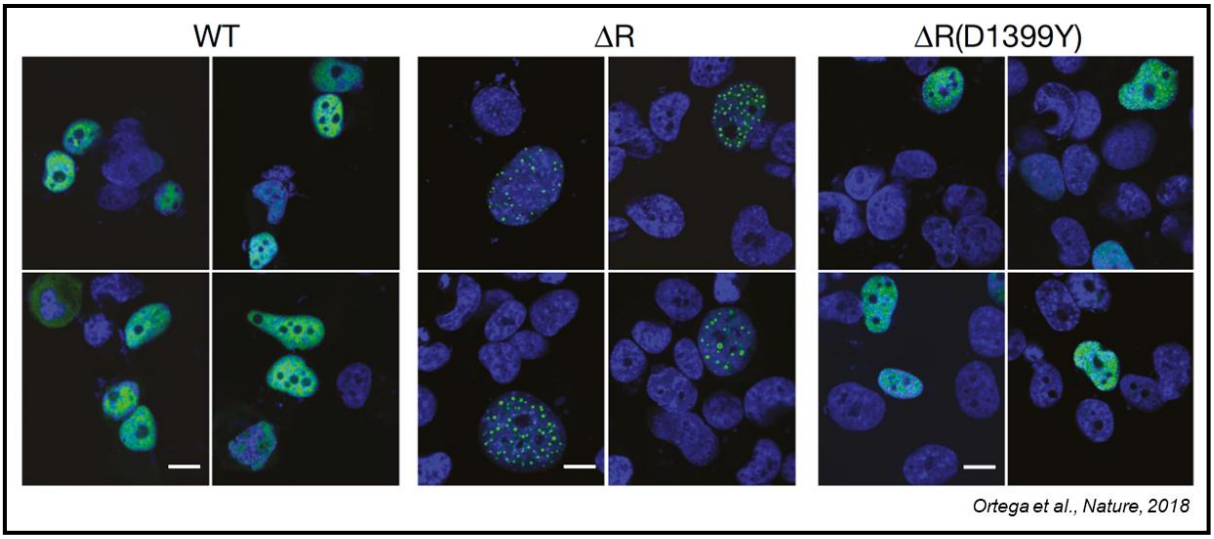

Fig.1.8.: RING domain of p300 hinders the protein's autoacetylation and therefore condensate formation. *Cells stained for p300; constructs expressed: left: wild-type p300, center: p300(ΔRING), right: p300(ΔRING)(D1399Y). The removal of RING domain causes hyperactivation of p300, resulting in condensates formation. Additional point mutation disactivating HAT domain removes this effect (Ortega et al. 2018)*

for Brd4-Nut condensate formation, as it has been shown that hyperactivated p300 can also form large nuclear condensates in mammalian cells.

Here, I dissected the molecular features of proteins involved in aberrant Brd4-Nut condensate formation and studied their relation to transcriptional changes in cells. I found that the p300 interaction motif of Nut is sufficient, in fusion with Brd4, to form condensates. Relatedly, the small molecule p300 HAT inhibitor (C646) decreases condensates formation. Thus, I focused on the roles that different regions of p300 contribute to Brd4 condensate formation and transcriptional changes. I developed a series of stable cell lines inducibly expressing different mutants of Brd4-Nut and Brd4-p300 and examined their capacity to form condensates via spinning disc confocal microscopy and in-depth image analyses. I also performed RNAseq as well as Nut and histone H3K27Ac – based ChIPseq to study the link between condensate formation and transcription. I found that the ability to form condensates and to cause changes in transcription were distinguishable. While the histone acetyltransferase activity of p300 is critical for changes in transcription, multiple molecular features of p300 collectively contribute to condensate formation.

CHAPTER TWO

QUANTITATIVE CELLULAR MICROSCOPY

Introduction

It has been observed in the past that both Brd4 and Brd4-Nut (Fig.1.1B) can form liquid-like intranuclear condensates (Sabari et al. 2018; Alekseyenko et al. 2015; Yan et al. 2011; Zee et al. 2016; Reynoird et al. 2010). Line profiles across a representative nucleus demonstrate that the fluorescence intensity peaks observed upon expression of Brd4-Nut are higher and broader than the ones seen with either Brd4 or Nut alone (Fig.2.1A). This indicates that condensates formed by the fusion protein are more concentrated and larger than the ones formed by either Brd4 or Nut. I hypothesized that the Brd4-Nut condensates might represent aberrant condensation of Brd4, which could alter transcription, ultimately leading to cancerous cell transformation. In my dissertation project, I studied the molecular mechanisms of Brd4-Nut condensate formation and how the formation of these condensates relates to transcription. To address these questions, I expressed a series of fusion proteins containing different molecular elements of Brd4, Nut and p300 in human cells. Based on this model system, I was able to measure the ability of the different fusion mutant proteins to form condensates and study if and how these proteins might cause transcriptional changes in cells.

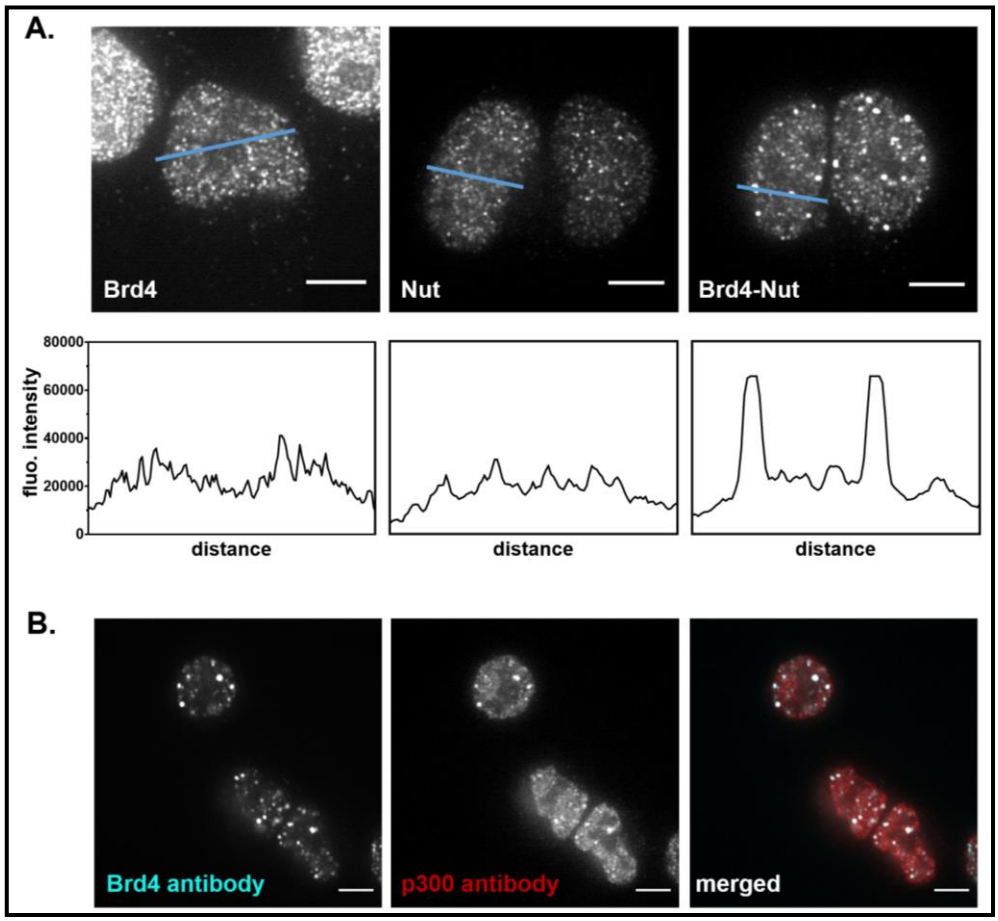

Fig.2.1.: Brd4-Nut fusion protein forms nuclear condensates that recruit p300 histone acetyltransferase: A. *Brd4, Nut or Brd4-Nut expressed in Hek293 TRex-FlpIn cells, immuno-stained for mNeonGreen tag. Blue line indicates where the line profile was measured, shown as a graph below each of the images. Line profiles show the fluorescence intensity along the blue line.* ***B.*** *Patient-derived Nut Carcinoma cell line, HCC2429, was immuno-stained for Brd4 and p300 proteins, showing that the resulting large nuclear condensates concentrate both Brd4-Nut and p300. All scale bars = 10 µm.*

I developed a rigorous way to quantify the differences between the condensates formed by Brd4 alone and by the Brd4-Nut fusion. In order to accurately compare cells expressing different proteins, I first developed multiple stable cell lines using an inducible protein expression system based on 293TRex-FlpIn cells. Next, I tested a series of different expression induction regimes,

to best approximate protein expression level and condensate formation in a NC patient-derived cell line, HCC2429. Then I built a microscopy analysis pipeline, incorporating tools from Cellpose (Stringer et al. 2021) and CellProfiler (Stirling et al. 2021).

Stable Cell Lines and Protein Expression Tuning

To be able to accurately compare condensate formation abilities as well as transcriptional profiles between cells expressing different fusion protein mutants, I needed to develop a series of stable cell lines. The choice of system was nontrivial, as expression of many fusion oncoproteins, including Brd4-Nut and its mutants, is toxic to cells. I initially tested expressing Brd4-Nut fusion protein via a transient transfection in HeLa cells and U2OS cells. Both these cell lines showed

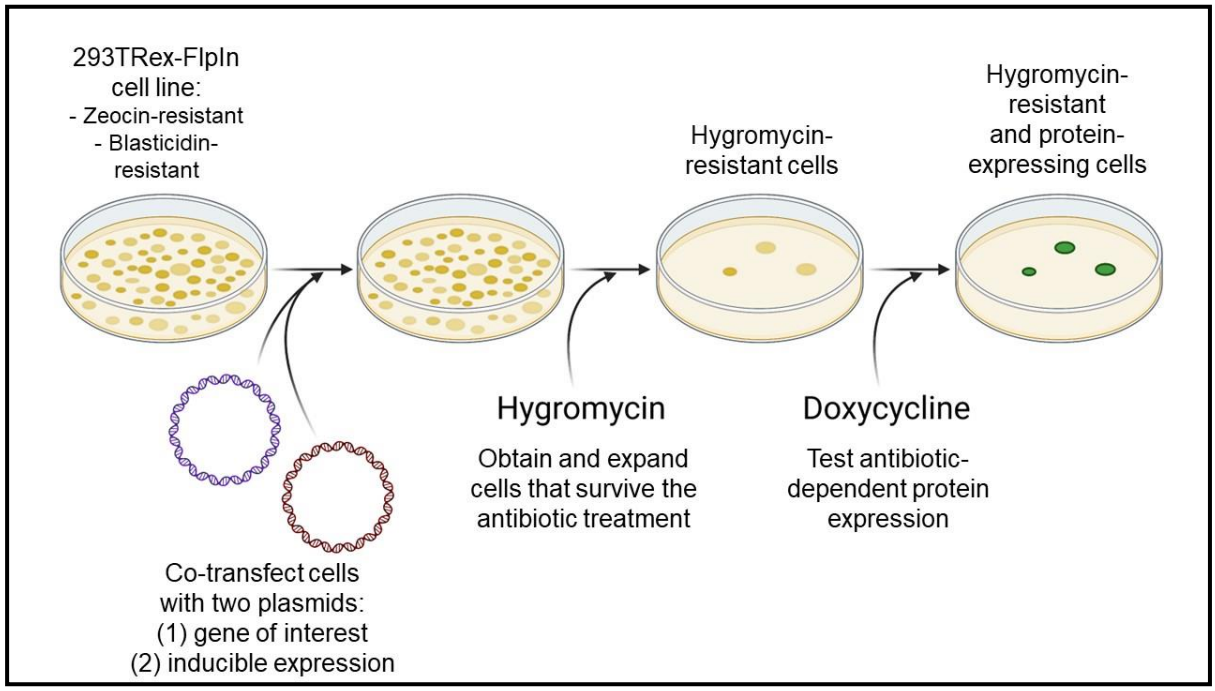

Fig.2.2.: Schematic of the 293TRex-FlpIn – based stable cell line development.

a high toxicity response after as short as 24h expression of Brd4-Nut. Next, I examined the fusion protein expression in Hek293 cells. While the protein was still toxic to these cells over a long period of time, the cells did survive a much longer time of protein expression than either U2OS or HeLa cells. Thus, I next sought out to find an inducible Hek239 – based expression system that would allow me to tune the expression levels to accurately mimic the expression of Brd4-Nut in Nut Carcinoma cells.

As reference, I used the HCC2429 patient derived Nut Carcinoma cell line, which was received as a kind gift from the Hamon Center for Therapeutic Oncology Research at UT Southwestern Medical Center. Then, I purchased 293TRex-FlpIn cells (Thermo Fisher Scientific), to use as a base cell line for stable cell line development. 293TRex-FlpIn cells contain two stably integrated features:

1. pFRT/*lac*Zeo element introduces a single FRT site into the genome and stably expresses *lac*Z-Zeocin gene, providing Zeocin resistance. The FRT site allows for a relatively easy, Flp recombinase – based insertion of a gene of interest. Because there is only one FRT site in this cell line, the gene of interest should be inserted in the same place in the genome of each cell, thus removing the necessity of a long process of clonal cell line development.
2. pcDNA6/TR element stably expresses the Tet repressor gene, under control of the CMV promoter. This allows for the gene of interest to be expressed in a Tetracycline – or Doxycycline – inducible manner.

To develop a stable cell line using this system, I co-transfected the 293TRex-FlpIn cells with two plasmids: pcDNA5/FRT/TO, containing my gene of interest and hygromycin resistance gene, and pOG44, allowing Flp recombinase expression. Next, I treated the cells with hygromycin, to select cells that have integrated the pcDNA5/FRT/TO plasmid (Fig.2.2). To account for the potential random integration of the gene of interest, the start codon of hygromycin resistance gene is already introduced to the genome of host cell line, 239TRex-FlpIn. In this way, only clones that have the gene of interest introduced at the exact correct locus, will be resistant to hygromycin. Finally, I tested the expression of the transgene by treating the hygromycin-resistant cells with doxycycline and monitoring the appearance of fluorescent signal from mNeonGreen protein, encoded within

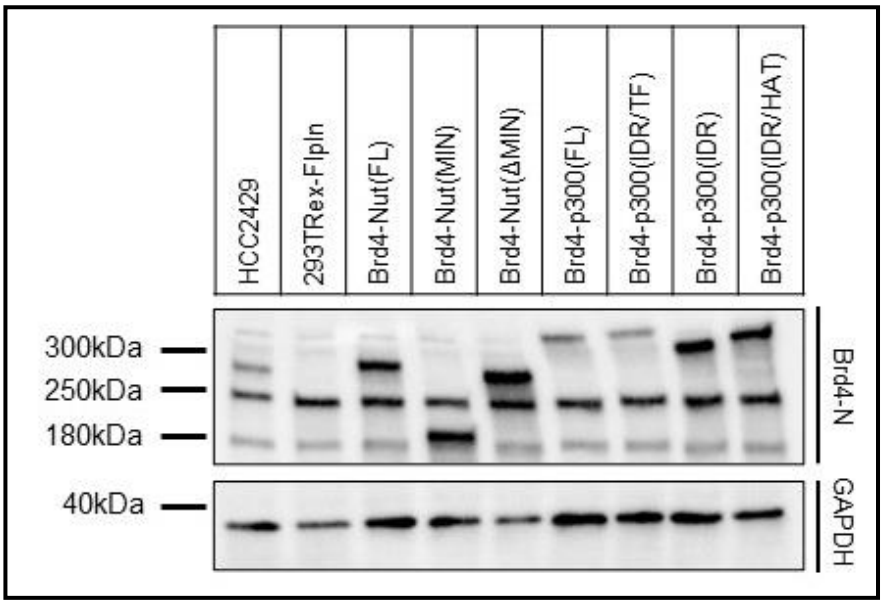

Fig.2.3.: Western blot showing the correct size of each fusion protein expressed in the different stable cell lines. *Antibodies used: Brd4-N, a kind gift from Cheng-Ming Chiang laboratory at UT Southwestern, to stain the fusion proteins, and GAPDH antibody for a loading control.*

the plasmid, via spinning disc confocal microscopy (Fig.2.2). The 293TRex-FlpIn cells are unique in that the newly introduced gene for inducible expression is incorporated in the same place in the genome of each cell, due to the single FRT site in the cells' genome. This feature removes potential expression variability due to differential placement of the transgene in the cell genome. In total, I developed seven stable cell lines, expressing mNeonGreen-tagged Brd4-Nut fusion protein and its mutants as well as Brd4-p300 fusion protein and its mutants, described in detail in chapter 3 of this dissertation (Fig.2.2).

Once developed, the Brd4-Nut(FL) stable cell line was used to find a correct treatment regime to induce protein expression to a level as similar to HCC2429 cells as possible. I first tested the canonical treatment with 1 µg/mL doxycycline for 24h and found that such mode of protein expression induction results in a very substantial overexpression of the protein (not shown). Thus, I lowered the concentration of the drug and tested several timepoints of treatment to assess protein expression levels via western blotting. Even as little as 50 ng/mL doxycycline for 4h caused a great overexpression of Brd4-Nut in the stable cell line, as compared to HCC2429 cells (Fig.2.4A). At the same time, I also observed that even if a whole cell population is treated with the same drug concentration for the same amount of time, the cells show a large level of variability of mNeonGreen signal intensity, when observed under a confocal microscope (Fig.2.4B). It became apparent that finding the right expression induction regime to ensure a low protein expression while allowing enough time for the condensates

to form was going to be challenging. I further lowered the concentration of doxycycline to 5 ng/mL and then treated the cells in a regime comprising two steps: a drug treatment step and a washout step. In this way, I was able to first induce the protein expression and then, without a further increase in expression, allow condensates to form. To further account for protein expression variability in the stable cell lines, I measured the protein expression levels via spinning disc confocal microscopy, in an immunostaining experiment, using a Nut antibody. Here, I treated cells for 1h with doxycycline, and followed the treatment with a 3h, 5h or 7h washout or I treated cells for 2h with doxycycline and followed the treatment with a 2h, 4h or 6h washout (Fig.2.5). I then measured the average pixel intensity across

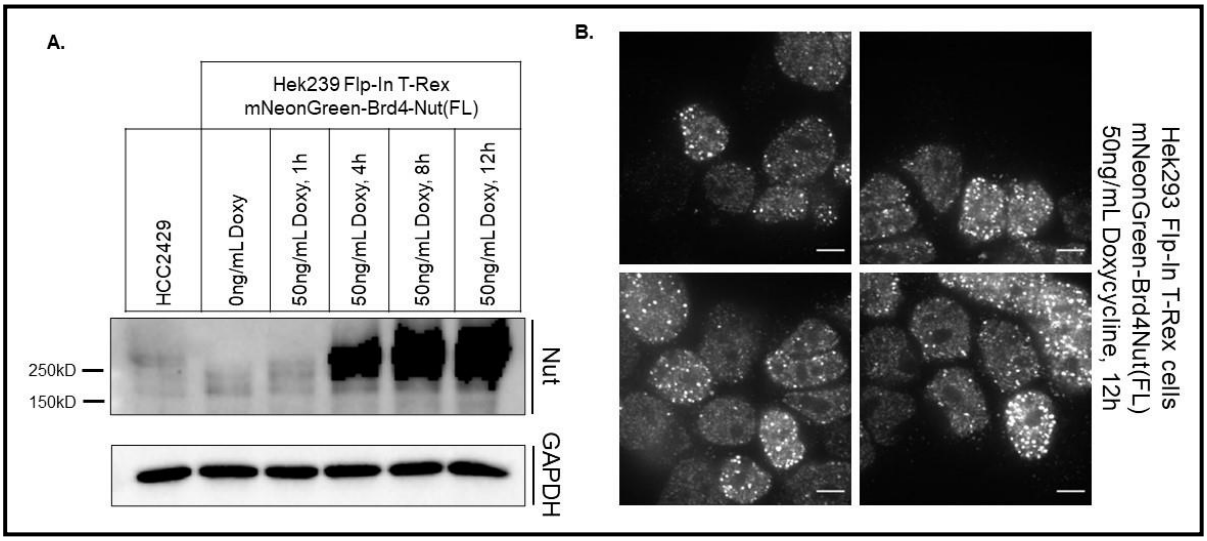

Fig.2.4.: Doxycycline – induced protein expression varies from cell to cell. ***A.*** *Western blot showing the difference in average expression level upon 1-12h treatment with 50 ng/mL Doxycycline.* ***B.*** *Example images of cells that have been treated with 100 ng/mL Doxycycline for 12 hours – expressing mNG-Brd4-Nut(FL) – notice the big expression differences within each field of view. Scale bar = 10 µm.*

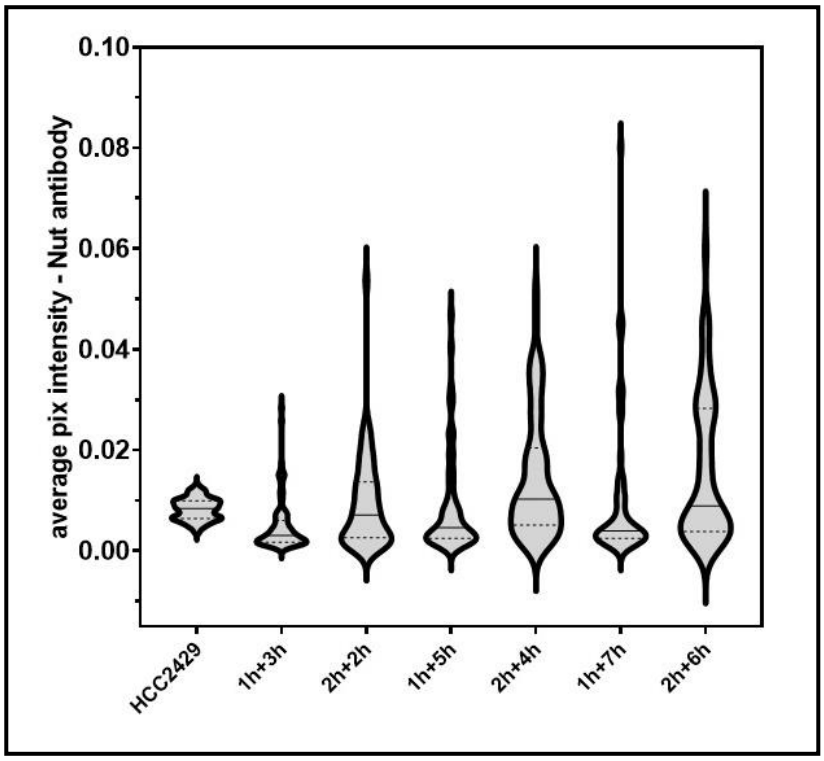

Fig.2.5.: Comparing doxycycline treatment regimes allows to find the protein expression pattern most similar to HCC2429 cells. *Protein expression level as measured by average pixel intensity in HCC2429 carcinoma cell line and Brd4-Nut(FL) – expressing stable cell line after 1 or 2h treatment with 5 ng/mL Doxycycline followed by 3/5/7h washout or 2/4/6h washout, e.g. 1h+3h = 1h treatment followed by 3h washout. Median indicated as a solid horizontal line and quartiles indicated as dotted lines. I chose the 2h treatment, followed by 4h washout for a good dynamic range and values similar to HCC2429.*

a cell nucleus, based on Nut antibody staining in single, representative Z-plane micrographs obtained via confocal fluorescence microscopy. There is a large dynamic range of expression levels upon each mode of treatment. However, the 2h treatment followed by 4h washout appeared to have majority of cells express the protein at a similar level to HCC2429 cells, within a reasonably large dynamic range. Thus, I decided to proceed with this mode of treatment.

Next, I treated all the cell lines in the same way and sought out to examine expression levels of each of the constructs. Since some of the constructs lack Nut (and thus could not be imaged using the α-Nut antibody), but all contain mNeonGreen, I first compared the immunofluorescence intensity based on the α-Nut and α-mNeonGreen antibodies. I decided to use an α-mNeonGreen antibody instead of relying on the autofluorescence of mNeonGreen tag, because the antibody staining resulted in a brighter signal. To compare the immunofluorescence signal intensity, I co-stained the stable cell line expressing Brd4-Nut(FL) protein with α-mNeonGreen and α-Nut primary antibodies, followed by Alexa Fluor 568nm or 647nm – conjugated secondary antibodies, respectively. Next,

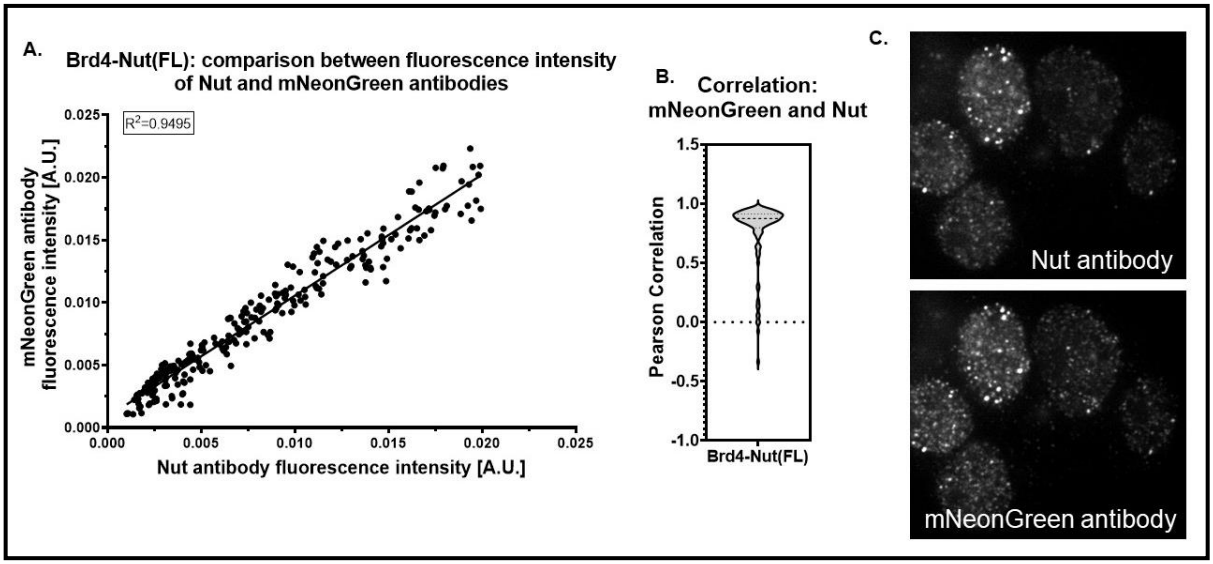

Fig.2.6.: Comparing mNeonGreen and Nut antibody staining and resulting fluorescence intensity. A. *Average fluorescence intensity from mNeonGreen antibody staining on y axis and from Nut antibody on x axis. Each datapoint is a single cell nucleus. The R^2=0.9495.* ***B.*** *Pearson correlation plot between the signal of mNeonGreen and Nut. Majority of the signal is close to 1, showing that there is a high level of correlation between the two channels.* ***C.*** *Example images of the cells co-stained with Nut antibody and mNeonGreen antibody to show the visible similarity between the two channels.*

I measured the resulting mean fluorescence intensity across the entire cell nucleus, using 568nm or 647nm channels to detect the fluorescent signal from either antibody. This measurement was performed in single Z-slice micrographs, acquired via confocal microscopy. I discovered that the two signals are highly correlated, as shown by R^2 (Fig.2.6A) and Pearson Correlation (Fig.2.6B). I also show the qualitative similarity in an example micrograph of the co-immunostained cells (Fig.2.6C). To control for the potential bleed through of fluorescent signal between the 568nm and 647nm channels, as well as possible cross-reactivity between antibodies, I additionally analyzed micrographs of cells that were co-stained with both α-mNeonGreen and α-Nut antibodies but lacked one or both antibody epitopes (Fig.2.7). The cell lines used here included:

1. The untransfected 293TRex-FlpIn cell line, not expressing any additional proteins, which should lack epitopes for either antibody,
2. HCC2429 Nut Carcinoma cell line, which constitutively expressed Brd4-Nut fusion protein and thus, should contain an epitope for the α-Nut antibody, but lack an epitope for the α-mNeonGreen antibody,
3. 293TRex-FlpIn – based stable cell line, inducibly expressing mNeonGreen-Brd4-p300(FL), which contains the α-mNeonGreen epitope, but lacks an epitope for the α-Nut antibody. should he Comparison of the average fluorescence intensity across the cell nuclei in a single Z-slice reveals that only cells expressing .

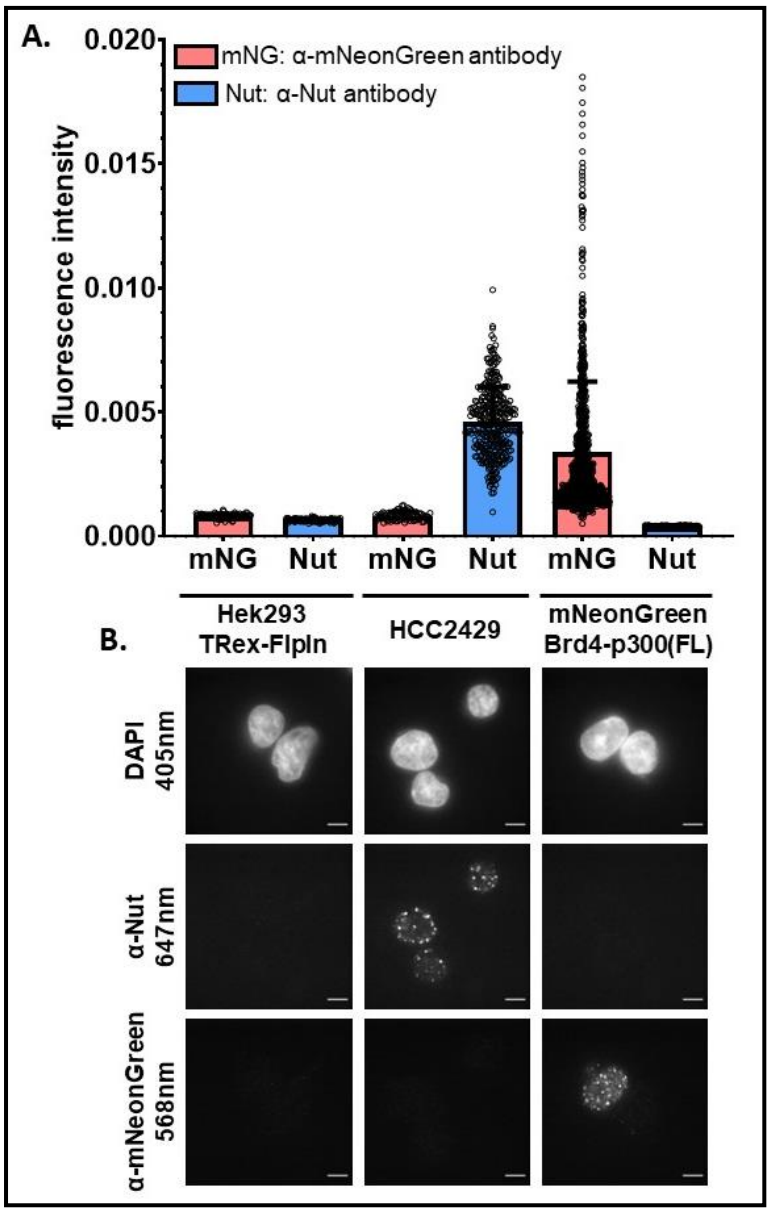

Fig.2.7.: Analysis of a potential bleed through between laser channels and cross-reactivity between the α-Nut and α-mNeonGreen antibodies. ***A.*** *Average fluorescence intensity from α-Nut and α-mNeonGreen antibodies staining across different cell lines. Each datapoint is a single cell nucleus.* ***B.*** *Representative micrographs of all three cell lines used in the analysis, as described. Scale bar = 10 µm.*

The obtained data show that there is a high intensity signal from either antibody only when an epitope for the antibody is expressed in cells (Fig.2.7A,B: blue bar in HCC2429 and pink bar in mNeonGreen-Brd4-p300(FL)). When a given epitope is absent, the signal from the corresponding antibody is negligible (Fig.2.7A,B: blue bar in mNeonGreen-Brd4-p300(FL) and pink bar in HCC2429). Additionally, I also observed presence of a low intensity background signal in untransfected 293TRex-FlpIn cells, with both α-Nut

and α-mNeonGreen antibodies (Fig.2.7A, 293TRex-FlpIn). The level of this signal is comparable to signal observed in epitope-negative cells stained with an antibody (i.e., cells stained with α-mNeonGreen antibody in absence of mNeonGreen expression or cells stained with α-Nut antibody in absence of Nut expression). This result indicates that the low fluorescence intensity observed in absence of an epitope for a given antibody is not an effect of bleed through between 568nm and 647nm channels but rather, low background antibody staining. All these results together suggested that cells stained with either α-Nut or α-mNeonGreen antibody can be accurately compared .

I next proceeded to treat all the stable cell lines in the same established regime of protein expression induction (2h treatment with 5 ng/mL doxycycline followed by 4h washout). I found that different cell lines express the fusion proteins to varying degrees, with median fluorescence intensity values varying by 4-fold, from ~0.005 to ~0.02 and fluorescence intensities per individual cell ranging from 0.002 to 0.16, depending on the cell line (Fig.2.8A). In comparison, the HCC2429 cell line expresses Brd4-Nut fusion protein at an average of 0.005 intensity signal, with fluorescence intensity values ranging from 0.002 to 0.015 (Fig.2.8A). This information allowed me to establish a top and bottom fluorescence intensity threshold for further image analysis purposes, to only consider cells that express the different fusion proteins within the same dynamic range as the HCC2429 cell line expresses Brd4-Nut (Fig.2.8B). I found that the fluorescence intensity range of 0.003 – 0.02 works well, as it is within the same

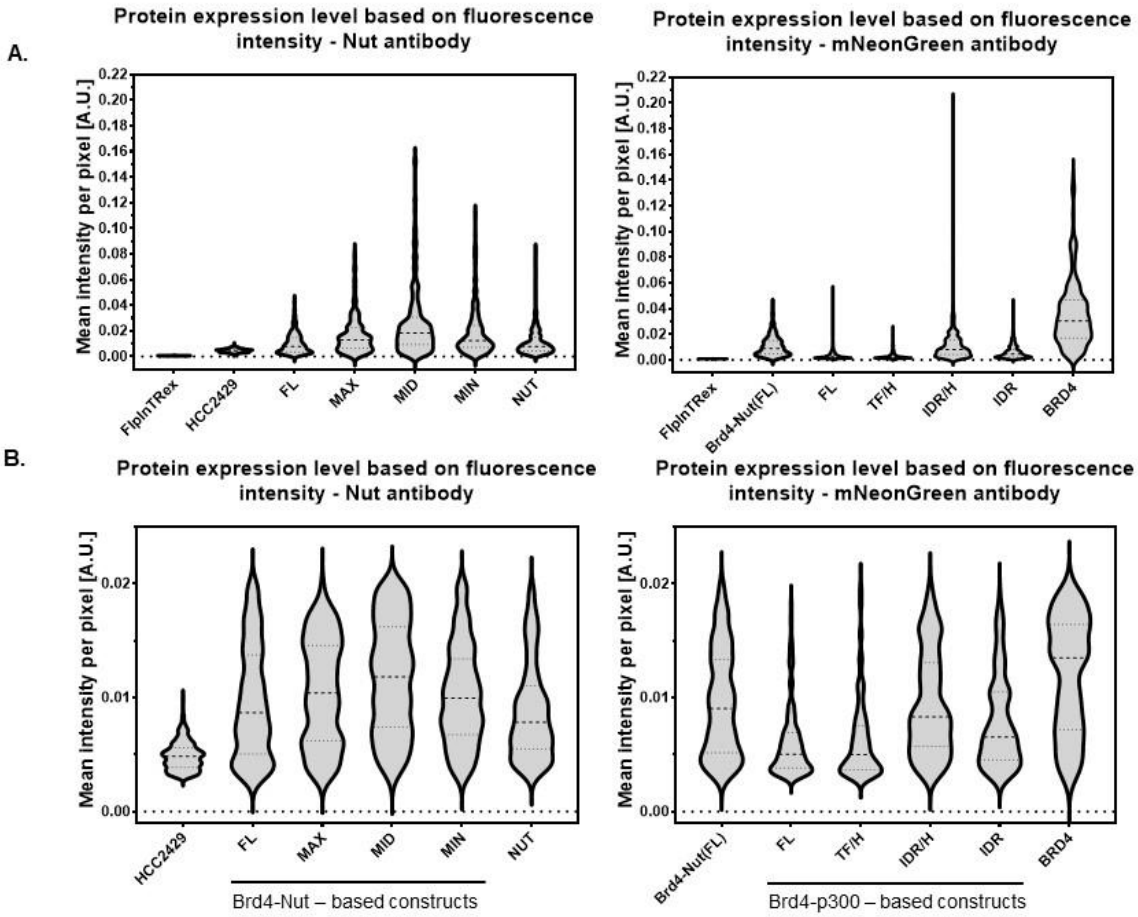

Fig.2.8.: Controlling doxycycline-induced protein expression levels allows to compare condensate formation between cells expressing different constructs. ***A.*** *Expression level of all imaged cells, using Nut antibody or mNeonGreen antibody; stable cell lines treated with 5ng/mL Doxycycline for 2h, followed by a 4h washout.* ***B.*** *Expression levels shown upon setting an expression cutoff at 0.003-0.02; stable cell lines treated with 5ng/mL Doxycycline for 2h, followed by a 4h washout. In all graphs, the thick dotted line represents the median, and the thin dotted lines represent the 25th and 75th percentiles of the data.*

range of expression as Brd4-Nut in HCC2429 cells, removing many of the outliers that express much more or much less protein than the Nut Carcinoma cells. At the same time, this range of fluorescence intensity also allows to keep the analyzed sample relatively large.

Once I developed a reliable method to treat all the stable cell lines, I proceeded to perform the spinning disc confocal microscopy experiments. To this end, all cell samples were fixed with 4% paraformaldehyde and immunostained using α-mNeonGreen, α-Nut, α-p300, α-histone H3K27Ac or α-Brd4 antibodies and then counterstained using Hoechst 33342 nuclear dye (equivalent of DAPI). The cells were then imaged on a Nikon Eclipse Ti microscope base equipped with a Yokogawa CSUX1 spinning disk confocal scanner unit, using 100x / 1.49 NA oil objective and Andor EM-CCD camera. Images were acquired using MetaMorph software. A single in-focus Z-plane in the center of cell nucleus was acquired per image, with an average of 100-200 images acquired per sample.

Confocal Microscopy and Analyses

Following image acquisition, I next wanted to obtain detailed and quantitative image-based information for each cell line. The data I sought out to obtain included among others: numbers of nuclear condensates per cell, average sizes of the condensates, average brightness of condensates and surrounding nucleoplasm as well as overlap between condensates found using different immunostaining antibodies. Given the multiple cell lines and hundredths to thousands of images acquired per experiment, I deduced that the final amount of data collected would be very challenging and not objective to analyze manually. Thus, I utilized a series of image analysis tools and optimized them for use with my data. First, I performed nuclear segmentation based on the Hoechst 33342

nuclear dye staining for all images, using Cellpose (Stringer et al. 2021). Cellpose is a highly accurate, deep learning – based algorithm for image segmentation. The algorithm was pre-trained on a large dataset of highly variable cellular images, with over 70,000 segmented objects and thus, proved superior to other available image segmentation tools. Nuclear segmentation of multiple cell lines proved very challenging, as HCC2429 and 293TRex-FlpIn cells grow differently *in vitro,* with some cell lines being sparser and others growing more closely together, in clusters. Additionally, expression of different fusion proteins also affected the way cells grew and clustered to some extent. Thus, compared to other available tools for this kind of analysis, which often utilize diverse thresholding algorithms, Cellpose proved to be much more precise for my analyses. Here, I preset the nuclear diameter to 80 pixels and utilized a short code in Python to batch-analyze large sets of images. Once the images were segmented, I used CellProfiler (version 4.2.1 for Windows (Stirling et al. 2021)) to perform the remaining analysis. To find condensates with a high level of confidence, I needed to keep in mind the point spread function (PSF) of the microscope. While the theoretical resolution of a microscope provides valuable information on what size objects can be imaged and recognized

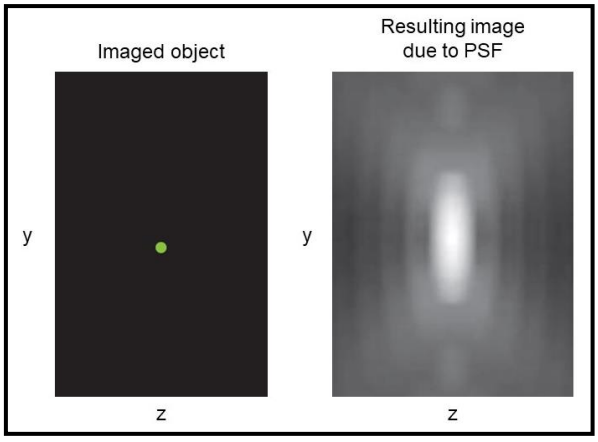

Fig.2.9.: Example of an imaged bead and the resulting microscopy image, blurred due to PSF. *Figure adapted from (Cole, Jinadasa, and Brown 2011).*

in micrographs, the obtained image is always a distorted representation of the actual light-emitting object (Fig.2.9). The PSF provides a mathematical description of this distortion (Cole, Jinadasa, and Brown 2011). It is the image produced by the microscope optical system of a point source of light and includes effects from the numerical aperture of the microscope objective, size of the imaged object and the wavelength emitted from the fluorescent object. Effects of the PSF influence both size and intensity measurements performed using microscopy-based imaging. Thus, in microscopy image processing the PSF must be empirically determined for each experimental condition and corrected for during image analysis. Generally, the PSF should be symmetrical can be mathematically modeled as 3D Gaussian.

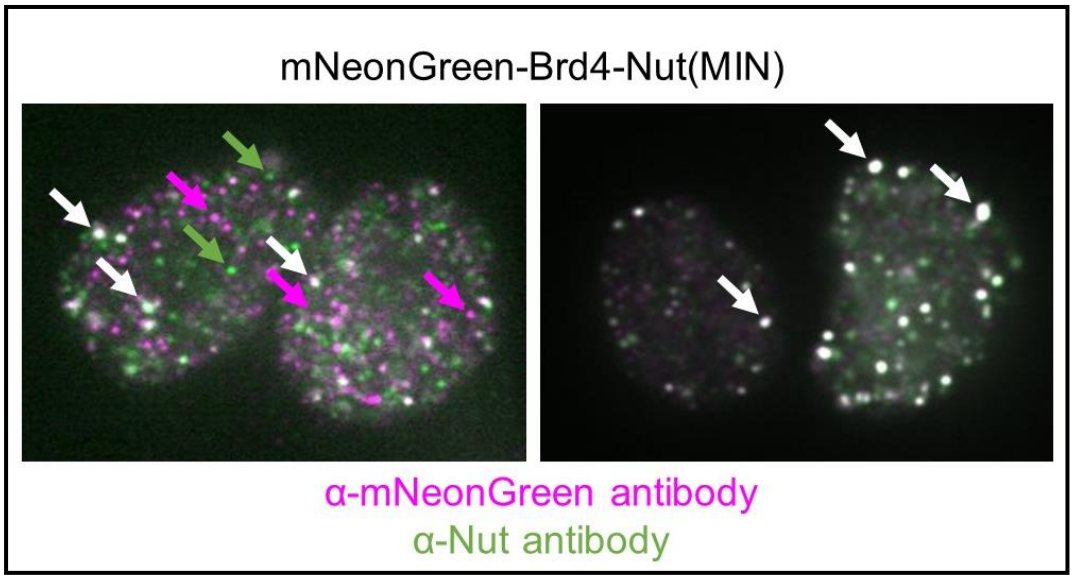

Fig.2.10.: Examples of micrographs of cells with large condensates and small condensates. *Note that the α-mNeonGreen and α-Nut antibody staining overlap well in large condensates (white arrows). Some of the small foci, where the antibody staining does not overlap well are indicated with magenta arrows (α-mNeonGreen antibody staining only) and green arrows (α-Nut antibody staining only).*

Not only is it useful to determine the actual resolution of an imaging system, but also as a tool used to deconvolute a blurred image. In my experimental setup, I determined the PSF of the microscope in the x-y dimensions to be on average about 600nm. This information allowed me to start working towards an image processing approach, which would enable me to quantitatively characterize differences in condensate formation observed upon expression of different fusion protein constructs. I observed that while fluorescent signal from large condensates overlapped between the α-mNeonGreen and α-Nut antibody staining, signal from small condensates did not coincide nearly as well (Fig.2.10). I believe that these small punctate structures do not represent fusion protein condensates, but rather, artifactual fluorescence intensities that might result from background antibody staining. Thus, throughout the follow-up quantitative image analyses, I decided to focus only on large condensates. While this approach allowed me to remove the small artifactual foci from my analysis pipeline, I acknowledge that it most likely also eliminated the smaller genuine condensates from my considerations. Through an empirical analysis, I determined that a size cutoff of 1.25 μm for the diameter of analyzed condensates, which is just slightly above 2xPSF, would provide a reasonable compromise between the competing goals of eliminating artifactual foci and yet, accounting for most actual condensates.

The following steps were next optimized using a set of representative micrographs and performed in the order they are listed:

1. <u>Background subtraction and flat-field correction:</u>

Illumination variations are often observed in microscopy images. Sometimes such variations can lead to artifactual fluorescence intensity fluctuations that could be as large as two-fold, due to optical system used or imperfections of the imaging slide used rather than actual differences in the biological system used. Therefore, correcting the illumination irregularities during image processing step is crucial for accurate intensity measurements.

- *Subtract background*: subtract an empty image from the empty image with fluorophore. To obtain the empty images, I used an empty well in microscopy plate with PBS, as imaging medium. To obtain the empty image with fluorophore, I made serial dilutions of Alexa Fluor – conjugated secondary antibodies used in immunostaining of cells and tested them all. Based on empirical analysis, I decided to use a 1:100 dilution of the antibodies as the fluorescent background (Fig.2.11A).
- *Calculate the illumination function*: determine a mathematical description of illumination pattern across a micrograph and create an image that is representative of the overall illumination pattern. This calculation is performed on the background-subtracted empty image, created using the module described above. The illumination function will later be used to correct for a potential uneven illumination in images (Fig.2.11B). In some cases, there is an additional smoothing step needed here, to prevent noise spikes influencing the maximum intensity. However, because in the case

of my studies the background noise is low, relative to the measured fluorescence intensities, I did not perform the smoothing step in my analysis.

- *Divide the illumination function image by its maximum*: this calculation normalized the illumination function, such that now the maximum intensity of the illumination function is 1 (Fig.2.11C).
- *Subtract background from the image with cells*: in the same way that it was done in the first background subtraction step described above (Fig.2.11D, compare with Fig.2.11A).
- *Divide by the illumination function*: in this step, I divided the background-subtracted image of cells by the illumination function image after corrections, determined in the steps described above. This produces a flat-field corrected image of cells (Fig.2.11E).

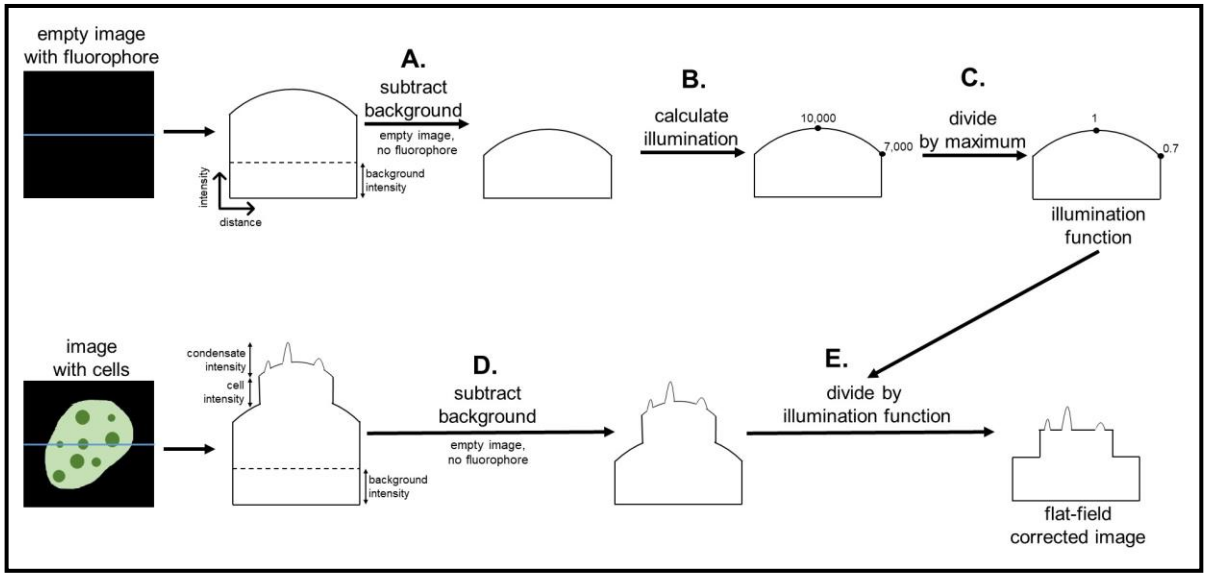

Fig.2.11.: Visual representation of the background subtraction and flat-field correction in image analysis pipeline.

2. Identifying condensates in cell nuclei: once the illumination corrections have been applied, I next sought out to determine the best way to segment the cell nuclei and then, identify large condensates inside the cell nuclei. Nuclear segmentation was performed using Cellpose, in a method described in detail at the beginning of the “Confocal Microscopy and Analyses” section of this chapter. The follow-up condensate segmentation was performed using CellProfiler.

- *Apply segmented images from Cellpose to CellProfiler:* I uploaded the DAPI stain-based segmented images produced through Cellpose into CellProfiler and masked them. I used a typical nucleus diameter of 80-300 pixels, to make sure all nuclei are included, and removed any nuclei that are touching the image borders. The best thresholding method for nuclei identification in the masks from Cellpose was empirically determined to be Otsu (Fig.2.12).

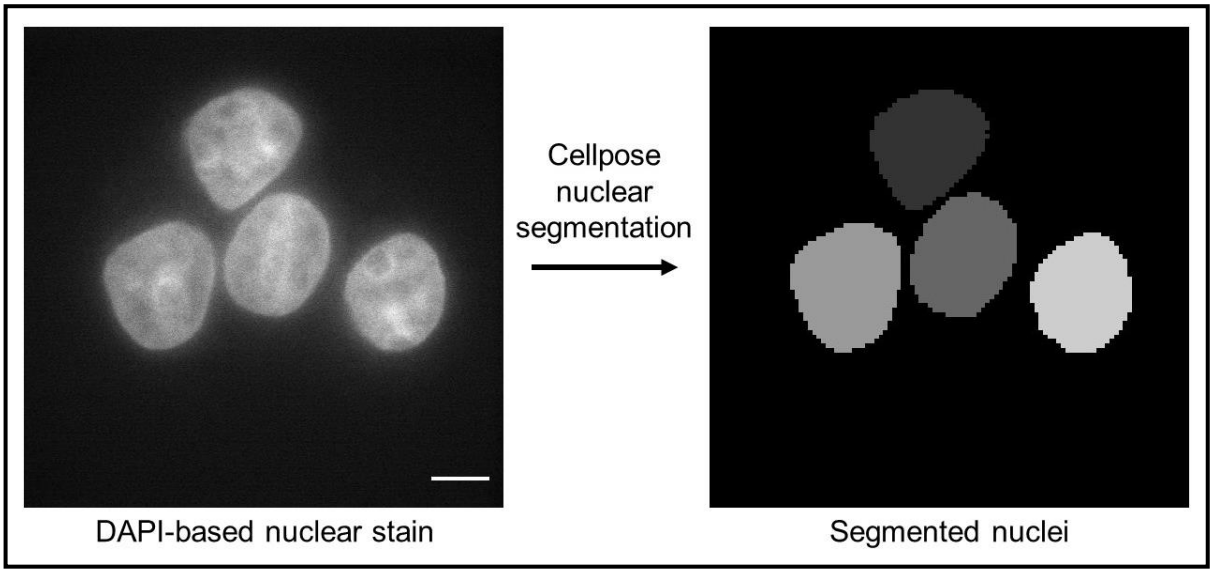

Fig.2.12.: Example of the Cellpose nuclear segmentation result

- *Enhance the fluorescence intensity of pixels within condensates* relative to the rest of the image, to improve subsequent identification of condensates.
- *Mask the nuclei in the image with cells* using the segmented images produced through Cellpose.
- *Identify the condensates – round I*: using a first, more lenient method of thresholding with a typical condensate diameter of 6-40 pixels, adaptive thresholding with the Robust Background thresholding method and size of adaptive window of 50 pixels, determined empirically (Fig.2.13A). Adaptive thresholding methods calculate a different threshold for each pixel, thus adapting to potential differences in fore- and background fluorescence intensities in the analyzed image. This step allows me to identify all large

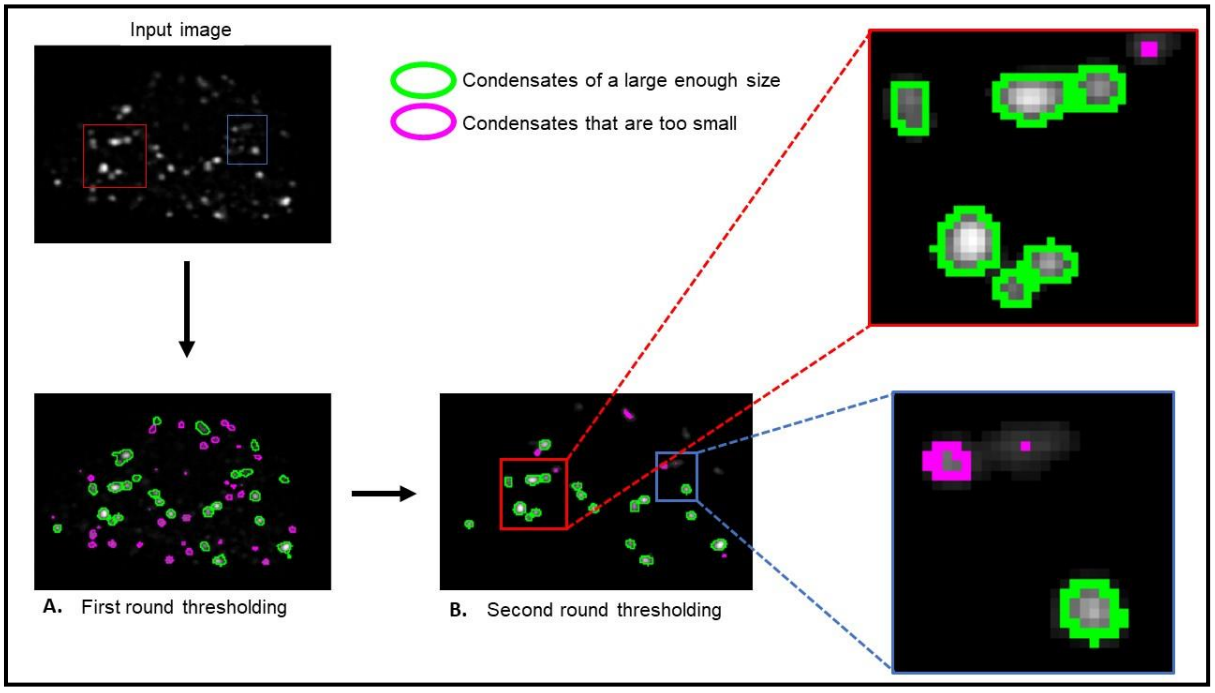

Fig.2.13.: Visual representation of the identification of condensates in image analysis pipeline.

condensates, but often does not account for their correct shapes, merging low intensities surrounding condensates or combining some of the large condensates with neighboring small condensates (Fig.2.13A).

- *Mask the condensates* identified during round I of thresholding.
- *Identify the condensates – round II*: using a second, more stringent method of thresholding: typical condensate diameter: 8-40 pixels, global thresholding performed with the global Otsu thresholding method with a threshold correction factor: 0.5 (Fig.2.13B). Global thresholding methods calculate a single threshold value for the input image. Then, they use this determined value to classify pixels that have intensities higher than the threshold as foreground and the ones with lower intensity as background. Since the condensates have already been masked, following round I of identification, I empirically determined that the global thresholding method works well to account for the large condensates' shapes and splits neighboring condensates that might otherwise have been merged in the first round of identification.

3. Quantifying the condensates: in this set of modules, I aimed to infer quantitative information based on the processed cellular images. The data I intended to obtain include the number of condensates per cell, condensates shape and size, as well as the fluorescence intensity of condensates and the surrounding nucleoplasm.

- *Measure the fluorescence intensity* across the nuclei as well as within the identified condensates. The quantitative fluorescence intensity information obtained here is reported as average within the nucleus, average among condensates within a given cell and average within a given condensate. Through this module, I can subsequently apply the fluorescence intensity cutoff in my analyses, as described in the “Stable Cell Lines and Protein Expression Tuning” section of this chapter.
- *Measure the condensate size and shape*: Through this module, I can analyze if there are differences in condensate size distribution upon expression of different constructs. Additionally, here I can determine the eccentricity of condensates, which is a numerical descriptor of roundness. This can be an important measure of morphology differences among different constructs.
- *Establish a relationship between nuclei and condensates*: here, the algorithm records which condensates are found in which nucleus. In this way I can analyze the numbers of condensates per nucleus.
- *Quantify the overlap between different channels*, e.g., in the case of co-immunostaining with two separate antibodies, such as α-Brd4 and α-p300 or α-Nut and α-histone H3K27Ac etc.
- *Quantify the overlap between different features found in all channels*, e.g., condensates found in a co-immunostaining experiment. One of the methods utilized here is Pearson Correlation.

Once I had all the quantitative information in-hand, additional filtering through the data was required. I needed to next remove outliers from the dataset and apply the expression cutoffs described above, in the sub-chapter “Stable Cell Lines and Protein Expression Tuning”. Only then was I able to confidently visualize the quantitative information collected via spinning disc confocal imaging.

Development of a series of stable cell lines, careful control for the different protein expression levels and development of a quantitative microscopy analysis pipeline for cellular imaging was non-trivial and required a long time of optimization at multiple steps. I hope that this rigorous way to analyze imaging data will soon become a standard among all biochemists and cell biologists.

CHAPTER THREE

OVERLAPPING, YET DISTINCT FEATURES OF P300 CONTRIBUTE TO BRD4-NUT CONDENSATE FORMATION AND TRANSCRIPTIONAL REGULATION

Brd4-Nut forms condensates and recruits p300

In my studies, I aimed to understand if and how Brd4-Nut condensate formation is correlated with gene expression changes in cells. To answer this question, I performed careful molecular dissections of Brd4-Nut fusion protein and p300 histone acetyltransferase. These two proteins have been shown to interact with one another and both colocalize inside Brd4-Nut condensates, in the HCC2429 patient-derived Nut Carcinoma cell line (Fig.2.1B). Because of the high level of acetylation observed in the condensates, it has also been proposed that both Brd4-Nut and p300 might play important roles in Nut Carcinoma-related transcriptional changes.

To start addressing the question of whether and how condensate formation might be correlated with gene expression changes, I developed a series of stable cell lines, using a commercially available, inducible protein expression system. Here, I chose 293TRex-FlpIn system, which provides a relatively fast, Flp-FRT-based method of developing stable cell lines with doxycycline-inducible

expression of the protein of interest (see methods for details on cell lines construction).

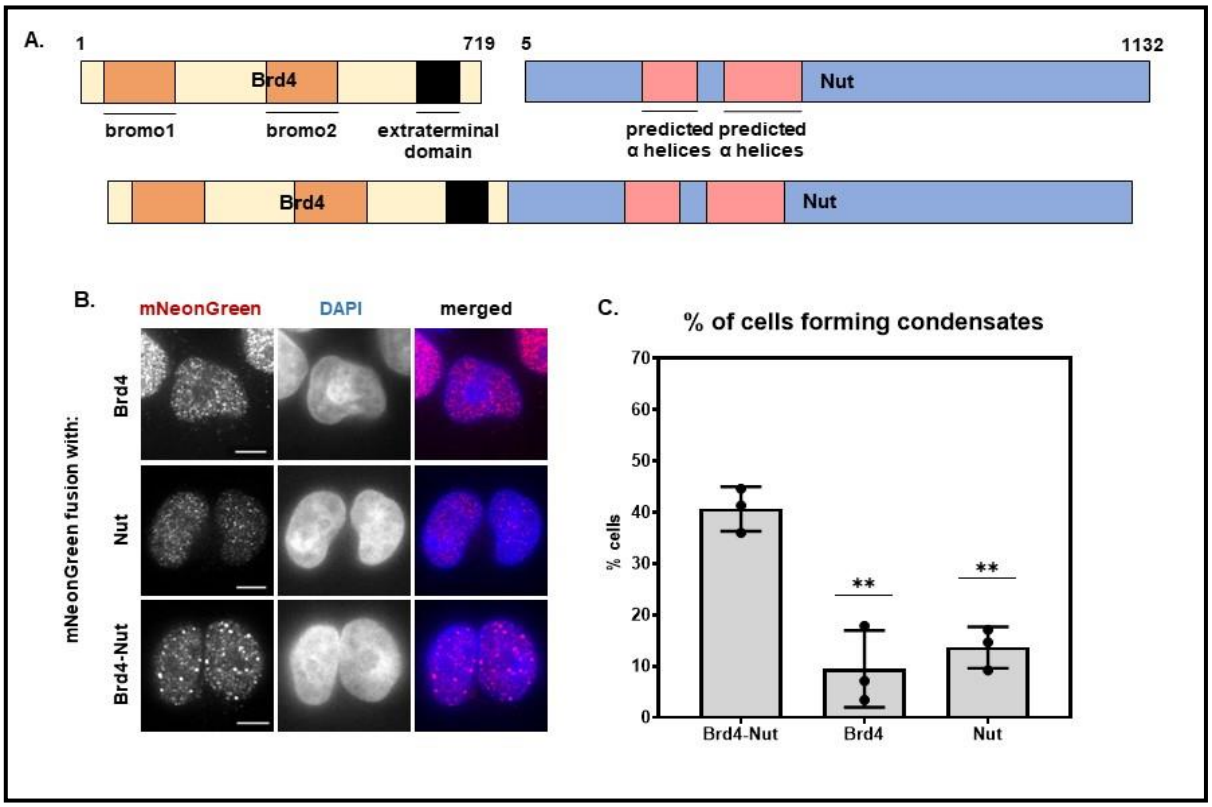

Fig.3.1.: Brd4-Nut fusion protein forms nuclear condensates in human cells. ***A.*** *Cartoon of Brd4, Nut and Brd4-Nut fusion proteins. Bromodomains 1 and 2 and extra terminal domain of Brd4 as well as predicted α-helices are indicated.* ***B.*** *Micrographs of cells expressing mNeonGreen-tagged Brd4, Nut and Brd4-Nut; scale bars = 10 μm.* ***C.*** *Quantification of percentage of cells forming condensates (at least 2 condensates larger than 1.25 μm in diameter).*

First, I wanted to test the similarities and differences in expression of the Brd4-Nut fusion protein, as compared to expression of either fusion component – Brd4 or Nut – alone. To this end, I analyzed the expression of Brd4, Nut and Brd4-Nut (Fig.3.1A), all tagged with mNeonGreen at their N-termini, in developed stable cell lines. I found that all three proteins can form condensates

(Fig.2.1A, Fig.3.1B,C). However, the condensates formed by Brd4-Nut fusion are larger than the ones formed by either Brd4 or Nut alone (Fig.2.1A). Furthermore, using the image analysis pipeline described in Chapter 2, I calculated the percentage of cells that form large condensates (>1.25 μm in diameter) and found that a higher fraction of cells expressing Brd4-Nut fusion forms condensates than cells expressing either Brd4 or Nut alone (Fig.3.1C). These data taken together indicate that Brd4-Nut expression leads to formation of larger condensates (Fig.2.1) and that a larger fraction of cells expressing Brd4-Nut fusion can form these condensates than upon expression of Brd4 or Nut alone (Fig.3.1). This further suggests that there is a cooperativity between Brd4 and Nut when the two proteins are directly fused to one another.

Since Nut has been previously shown to interact with p300 (Reynoird et al. 2010; Ibrahim et al. 2022; Yu et al. 2023), I next sought out to confirm the p300's recruitment to Brd4-Nut condensates (Fig.3.2A). Co-immunostaining of HCC2429 cells for p300 and Brd4 reveals that the signals from these two antibodies colocalize, suggesting that p300 is in fact recruited into Brd4-Nut condensates (Fig.3.2B). Additionally, immunoprecipitation of Brd4-Nut from the Brd4-Nut-expressing stable cell line also shows interaction with p300 (Fig.3.2C), consistent with previously published data (Reynoird et al. 2010; Ibrahim et al. 2022; Yu et al. 2023).

The data above provide evidence on the differences between condensates formed by Brd4-Nut fusion protein and either of the fusion protein components separately (Brd4 or Nut). I discovered that cells expressing Brd4-Nut have a higher

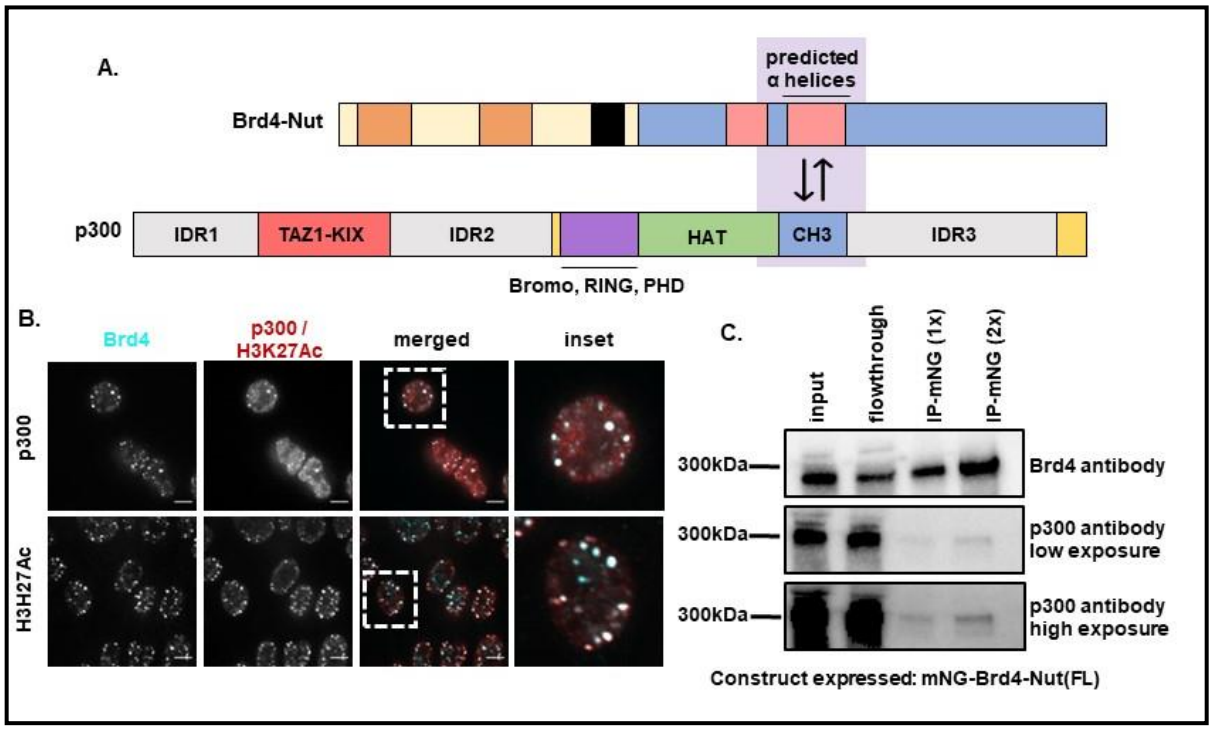

Fig.3.2.: Brd4-Nut condensates recruit p300 histone acetyltransferase and are heavily acetylated at H3K27. A. *Cartoon of Brd4-Nut and p300 interaction; known interaction motifs are indicated (Shiota, Barral, et al. 2018; Reynoird et al. 2010; Ibrahim et al. 2022)* ***B.*** *Micrographs of HCC2429 Nut Carcinoma cells co-stained with α-Brd4 and α-p300 or α-Brd4 and α-Histone H3K27Ac antibodies; scale bar = 10 µm.* ***C.*** *Immunoprecipitation against mNeonGreen; western blot with a p300 antibody showing that Brd4-Nut pulls down p300.*

propensity to form condensates than cells expressing either Brd4 or Nut alone, suggesting that there is cooperativity between these proteins, when in fusion. Furthermore, I confirmed that p300 is recruited into the large Brd4-Nut condensates in the patient-derived Nut Carcinoma cell line and that this histone acetyltransferase interacts with Brd4-Nut in a 293TRex-FlpIn-based stable cell line expressing the fusion protein .

Interaction with p300 is necessary and sufficient for condensate formation

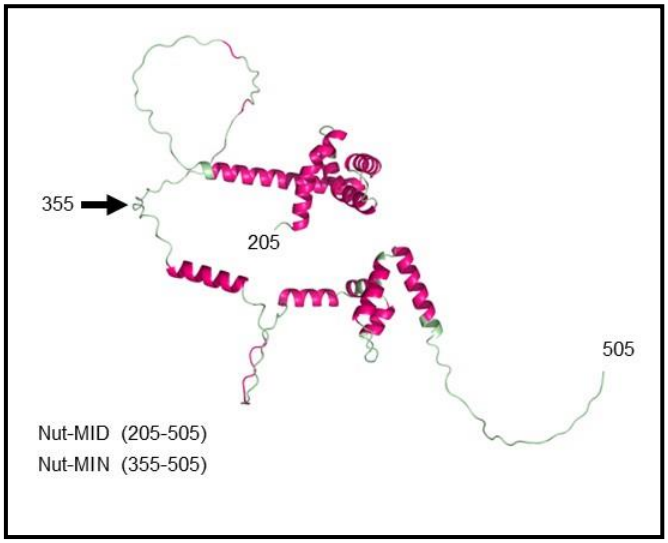

Fig.3.3.: AlphaFold2 secondary structure prediction of Nut(205-505), with residues belonging to Nut(MIN) and (MID) indicated.

I next sought out to determine which part of Nut is required for condensate formation. There are ~300 residues in the sequence of Nut, 208-476, that are predicted to form a series of α-helices (Fig.1.3, 3.3) and fold into a few small domains (Fig.3.3). A set of overlapping residues of Nut, 346-593, have been previously shown to bind to p300 (Reynoird et al. 2010). More recently, structures of two smaller regions within this folded part of Nut, spanning residues 403-418 and 419-470 were solved, with these parts of Nut directly bound too p300 (Yu et al. 2023). To preserve these potential structural elements, I designed a minimal fusion protein containing residues 355-505 of Nut, fused directly to Brd4, Brd4-Nut(MIN) (Fig.3.3, 3.4A). I also designed a complementary, Brd4-Nut(ΔMIN) fusion, lacking the MIN fragment of Nut (Fig.3.4A). I then developed stable cell lines expressing these constructs in an inducible manner (Fig.3.4B). Compared to Brd4-Nut(FL), a similar fraction of cells expressing Brd4-Nut(MIN) and a lower percentage of cells expressing Brd4-Nut(ΔMIN) form large condensates

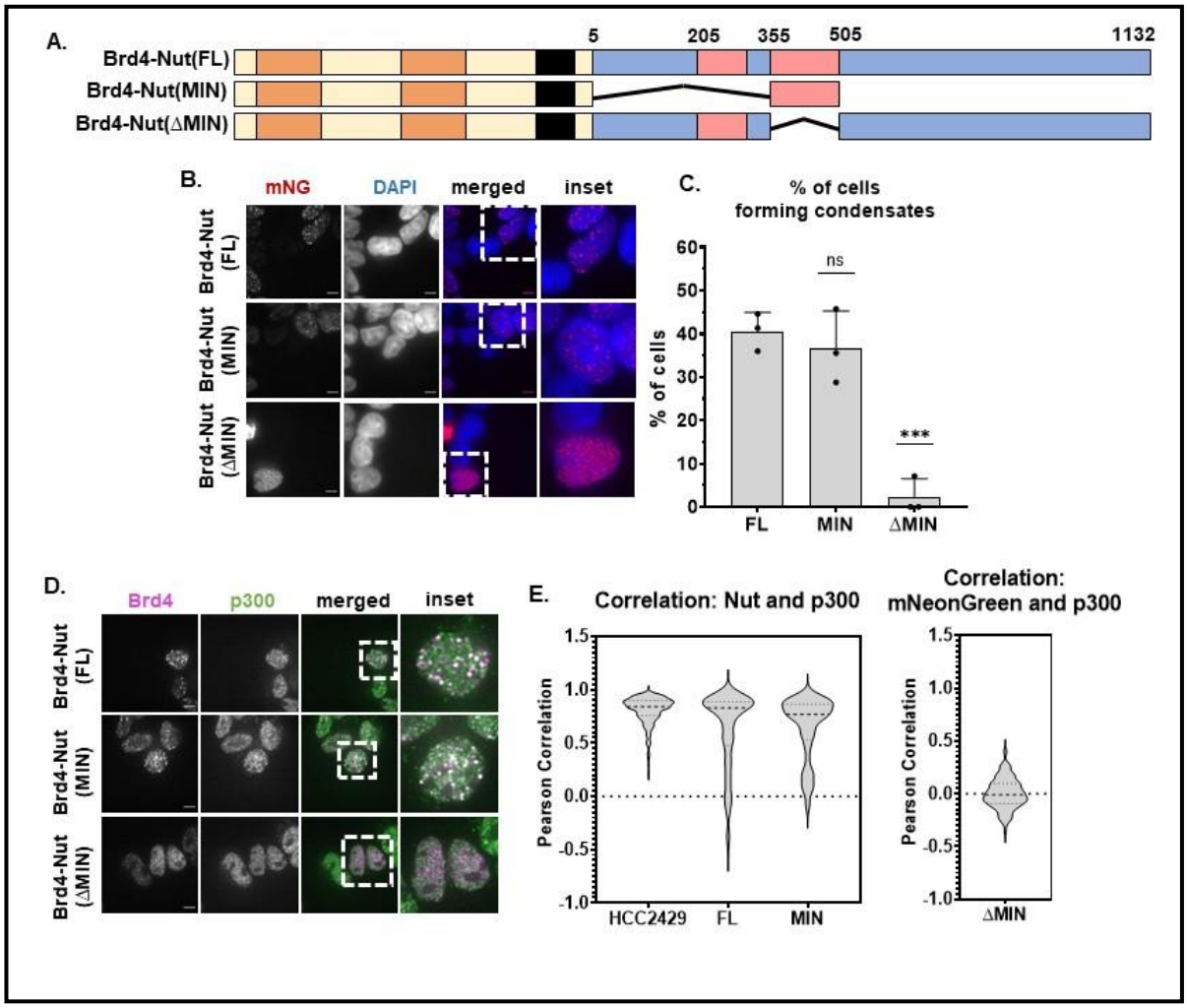

Fig.3.4.: Minimal p300-interaction fragment of Nut in Brd4-Nut fusion is necessary and sufficient for condensate formation. A. *Schematic of all Brd4-Nut – based constructs.* ***B.*** *Representative micrographs showing condensate formation in stable cell lines expressing different constructs. Staining against mNeonGreen shown in red, overlay with DAPI to indicate the nucleus. Scale bars = 10μm.* ***C.*** *Quantification of the micrographs represented in B.* ***D.*** *Representative micrographs comparing expression of constructs with or without p300-interaction motif in cells. Cells stained with anti-Brd4 antibody (magenta) and anti-p300 antibody (green). Scale bars = 10μm.* ***E.*** *Quantification of the overlap between condensates via co-staining shown as Pearson correlation. Cells co-stained with α-Nut and α-p300 antibodies (left-side graph) or α-mNeonGreen and α-p300 antibodies (right-side graph); Nut antibody epitope was removed in Brd4-Nut(ΔMIN) and thus, Nut antibody could not be used in this experiment. However, staining from α-mNeonGreen and α-Nut antibodies agrees very closely – see Fig.2.6 for more details.*

(Fig.3.4B,C). Knowing that a minimal p300-interacting region of Brd4-Nut is sufficient to form condensates, I then wanted to confirm that this region indeed recruits p300 into condensates. I performed a co-immunofluorescence experiment, staining cells for Brd4 and p300. I found that Brd4 and p300 colocalize in cells expressing Brd4-Nut(FL) or Brd4-Nut(MIN) (Fig.3.4D). I also quantified an overlap between Nut and p300, in an orthogonal co-staining experiment, using α-Nut and α-p300 antibodies. The quantification, shown as Pearson Correlation confirms a similar level of colocalization of Nut and p300 in HCC2429 cells as well as in stable cell lines expressing Brd4-Nut(FL) or Brd4-Nut(MIN) (Fig.3.4E). Additionally, in cells expressing Brd4-Nut(ΔMIN) mNeonGreen and p300 do not colocalize (Fig.3.4D,E). Here, I used the α-mNeonGreen antibody instead of Nut antibody to stain Brd4-Nut(ΔMIN) because the α-Nut antibody failed to interact with this construct (Fig.3.5). As detailed in Chapter 2, I confirmed a high degree of colocalization between α-mNeonGreen and α-Nut antibodies in cells expressing Brd4-Nut(FL), indicating that both

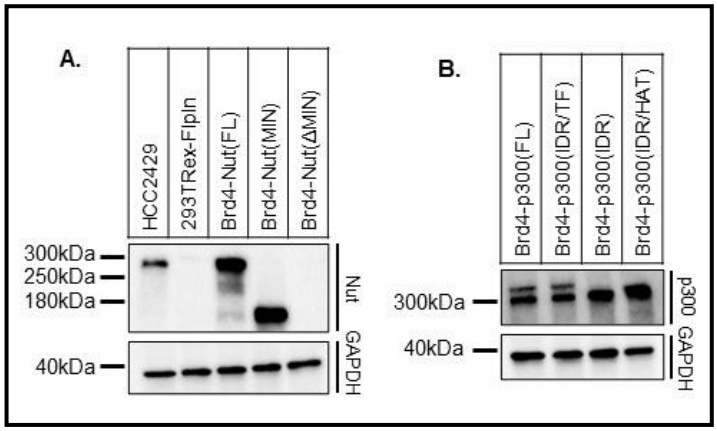

Fig.3.5.: : Western blots showing similarities and differences in the expression levels between cell lines expressing different constructs. *All cell lines treated with 5 ng/mL Doxycycline for 2h, followed by 4h washout. Antibodies used include: Nut and GAPDH* ***(A)*** *and p300 and GAPDH* ***(B).*** *Western blot including all stable cell lines showing levels of expression of all fusion proteins via Brd4-N antibody staining can be found in Fig.2.3.*

antibodies recognize mNeonGreen-labeled Brd4-Nut construct to a similar extent (Fig.2.6). These data confirm that Nut interacts with p300 through its MIN fragment and further suggest that this interaction contributes to condensate formation.

Having established that p300 is recruited to the Brd4-Nut condensates and that it is necessary for the condensate formation, I next asked whether p300 activity also contributes to condensate formation. I observed that Brd4-Nut condensates formed in the HCC2429 patient-derived Nut Carcinoma cells colocalize with fluorescent signal from an α-acetylated histone H3K27 (H3K27Ac) antibody (Fig.3.2B). Therefore, I questioned whether binding of Brd4 to acetylated histone tails and HAT activity of p300 are both necessary to form condensates. To address

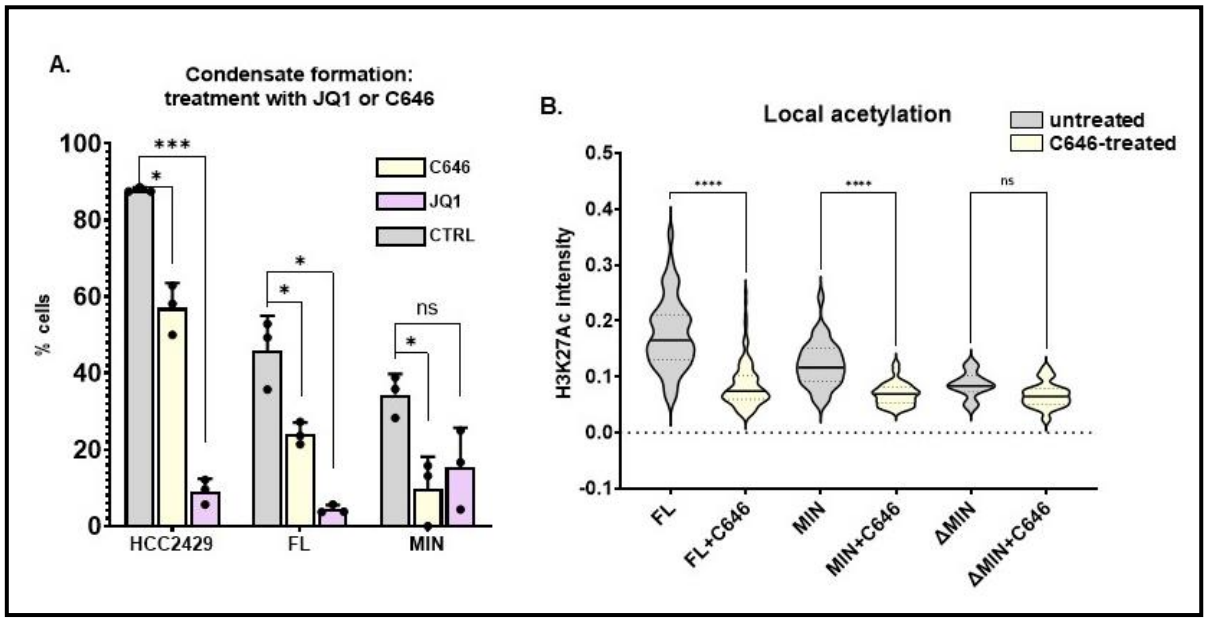

Fig.3.6.: Quantification of p300 activity influence on condensate formation. ***A.*** *Quantification of cells capacity to form condensates, comparing cells that are untreated and cells treated with C646 inhibitor or JQ1 inhibitor. Cells compared here include HCC2429 Nut Carcinoma cell line and stable cell lines expressing Brd4-Nut(FL) and Brd4-Nut(MIN).* ***B.*** *Local acetylation measured as average fluorescence intensity across a condensate. Data shown with or without C646 treatment, in cells expressing Brd4-Nut(FL), Brd4-Nut(MIN) and Brd4-Nut(ΔMIN).*

these questions, I used the established stable cell lines in co-immunostaining spinning disc confocal microscopy experiments, with addition of JQ1 inhibitor of bromodomains binding to acetylated histone tails (Filippakopoulos et al. 2010) and C646 inhibitor of p300 HAT (Zhao et al. 2015). A smaller fraction of cells expressing Brd4-Nut(FL) forms large condensates upon C646 treatment and this effect is even more pronounced upon treatment with JQ1 (Fig.3.6A). C646 and JQ1 also disrupt Brd4-Nut(MIN) condensates, although the effect of JQ1 is less noticeable than the effect of C646 (Fig.3.6A). Together, these data show that the predicted minimal folded fragment of Nut responsible for binding to p300 (MIN) in fusion with Brd4 is necessary and sufficient to form condensates and to recruit p300 into the condensates. Additionally, both acetylation of histone tails and Brd4 binding to acetylated histones contribute to Brd4-Nut condensate formation.

I then tested whether p300 is active in Brd4-Nut(MIN) condensates. Since H3K27Ac is a well-established marker of active transcription and it has been shown to be regulated by CBP/p300 and HDACs (Wang, Chen, and Zhang 2022), I decided to use it as a marker of p300 activity. I focused my analyses on the p300-triggered H3K27 acetylation inside the condensates. To this end, I measured the average α-H3K27Ac fluorescence intensity within the condensates via confocal microscopy and used it as a proxy for the acetylation level. I found that acetylation level in condensates formed by Brd4-Nut(FL) and Brd4-Nut(MIN) is high, and it significantly decreases upon C646 treatment (Fig.3.6). Since only a very small fraction of cells expressing Brd4-Nut(ΔMIN) form large condensates (Fig.3.4B,C),

the amount of local, in-condensate acetylation data to be collected for this construct was limited. Nonetheless, even based on the restricted-size sample, I was able to observe that the condensate-localized acetylation level was lower upon expression of Brd4-Nut(ΔMIN) than Brd4-Nut(FL) or Brd4-Nut(MIN) (Fig.3.6B). This observation was further corroborated by lack of additional acetylation decrease upon C646 treatment in cells expressing Brd4-Nut(ΔMIN) (Fig.3.6B). Based on these results, I conclude that interaction with p300 is important for local acetylation within Brd4-Nut condensates.

Taken together, the data above suggest that C646 HAT inhibitor causes a decrease in cells ability to for condensates. In agreement with this, acetylation is decreased in the remaining condensates that do form. Histone H3K27 acetylation in condensates is a result of p300 activity, as Brd4-Nut(ΔMIN), which lacks the ability to recruit p300, is unable to cause high levels of acetylation in the very few condensates that are formed by this construct.

Brd4-Nut(FL) and Brd4-Nut(MIN) have highly similar transcriptional profiles

Having established that the interaction of Brd4-Nut with p300 is important for condensate formation, I next sought to examine the potential functional consequences of this interaction and the resulting biomolecular condensates. Specifically, I studied the transcriptional changes that occur upon expression of Brd4-Nut and the role of p300 in driving these changes. I performed RNAseq

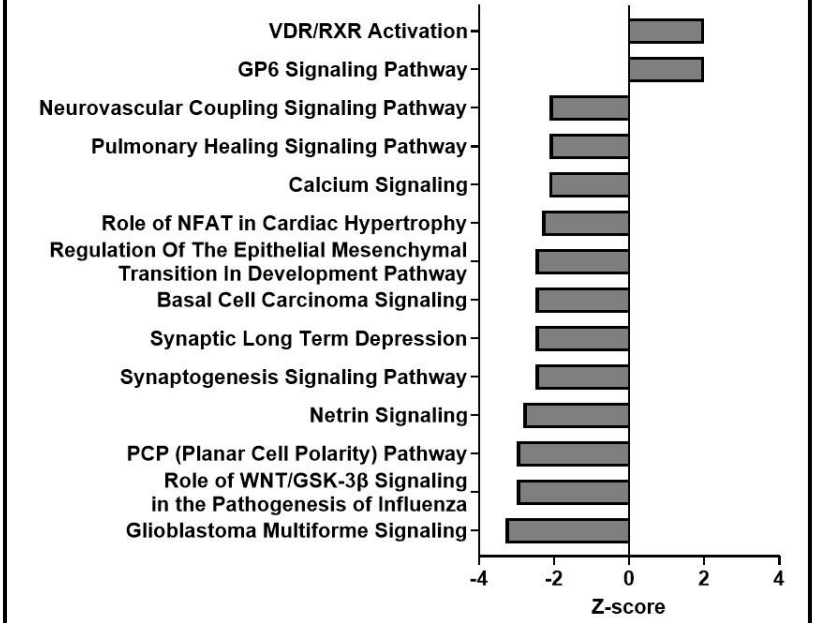

Fig.3.7.: Ingenuity Pathway Analysis: *most significantly up- and downregulated pathways, based on RNAseq of cells expressing Brd4-Nut(FL), as compared to 293TRex-FlpIn cells not expressing any fusion protein. Z-score is a measure of up- or downregulation. Only pathways with a p value of < 0.05 and Z-score ≥ 2 or ≤ (-2) are reported.*

and ChIPseq with an α-Nut antibody using the stable cell lines. I performed each of these experiments in two biological replicates and here I report results that are common between the replicates. Some of the genes that I identified to be differentially expressed in cell line expressing Brd4-Nut(FL) have previously been recognized as the signature genes of Nut Carcinoma, including *SOX2* and *TP63* (Eagen and French 2021; Alekseyenko et al. 2015). Additionally, RNAseq-based Ingenuity Pathway Analysis (IPA) revealed that expression of Brd4-Nut(FL) causes a significant up- and downregulation of multiple cellular pathways, many of which are related to cancer development and progression (Fig.3.7). More specifically, among the pathways:

- VDR and RXR receptors have been found to be overexpressed in many breast cancer patients (Heublein et al. 2017).

- GP6 signaling pathway has been reported to be upregulated in metastasis but inhibited in tumorigenesis in endometrial cancer (Yadav et al. 2020).
- The epithelial-mesenchymal transition (EMT) is a pathway through which epithelial cells lose their cellular polarity and gain invasive properties, often observed in cancer progression.
- Netrin-1 signaling pathway has been shown to be generally downregulated in human cancers (Arakawa 2004). More recently, next-generation sequencing (NGS) studies identified somatic mutations in Netrin-1 receptors, named “deleted in colorectal cancer” (DCC), in Nut Carcinoma patients (Cavalieri et al. 2017).
- WNT signaling has been associated with multiple cancers, including colorectal, breast and lung cancer. Mutations in WNT have also been proposed to be involved in Nut Carcinoma progression (Zhang et al. 2021).
- Basal Cell Carcinoma and Glioblastoma Multiforme signaling pathways might be downregulated upon expression of Brd4-Nut(FL), as a cellular response to stress caused by oncogene expression.

All these changes in cellular pathways show that the expression of Brd4-Nut(FL) in cells results in relevant gene expression changes. To examine the importance of the MIN fragment of Nut in transcriptional changes, I wanted to perform a comparative analysis between cells expressing Brd4-Nut(FL) and (MIN).

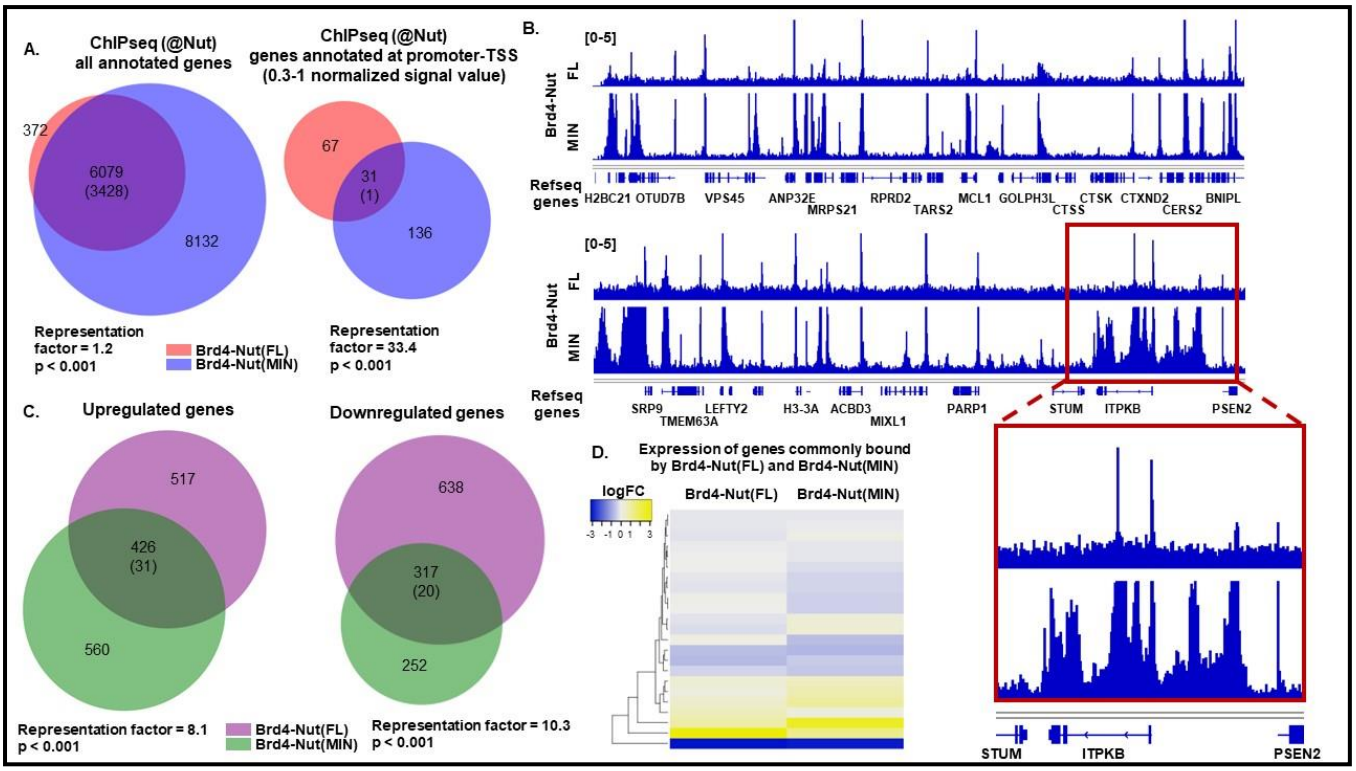

Fig.3.8.: Both Brd4-Nut(FL) and Brd4-Nut(MIN) – expressing cells have highly similar transcriptional profiles. A. *Venn diagrams of gene occupancy by Brd4-Nut(FL) and Brd4-Nut(MIN): left-hand side: all annotated genes occupied by either protein; right-hand side diagram: genes annotated at promoter-TSS regions of genes, with normalized signal value of 0.3-1. Values in the diagrams show the number of unique genes annotated; numbers in parentheses represent the mean overlap in 20 iterations with a randomly generated gene pool of the same size, from the human genome.* ***B.*** *Example ChIPseq tracks comparing gene occupancy by Brd4-Nut(FL) and Brd4-Nut(MIN). Data scale: 0-5. Inset shows more closely an example of the same genes being occupied by both proteins, but Brd4-Nut(MIN) occupying more loci.* ***C.*** *Venn diagrams of RNAseq genes that are significantly up- and down-regulated upon expression of Brd4-Nut(FL) or Brd4-Nut(MIN). Genes are only shown when fold change in expression was greater than two and p<0.05. Numbers in parentheses represent the mean overlap in 20 iterations with a randomly generated gene pool of the same size, from the human genome.* ***D****. RNAseq-ChIPseq data integration: heatmap showing up- and down-regulated genes found via RNAseq, that are bound by both Brd4-Nut(FL) and Brd4-Nut(MIN). Top 200 genes shown in the heatmap, based on p value.*

The ChIPseq data reveal that gene occupancy by Brd4-Nut(FL) and Brd4-Nut(MIN) is very similar, as almost all genes bound by the full-length fusion protein are also bound by Brd4-Nut(MIN) (Fig.3.8A, left-side diagram), consistent with the two

proteins recognizing similar elements of the genome. Furthermore, when I limit the analysis of gene occupancy to only the promoter-TSS (transcription start site) gene regions, 31 genes are found to be commonly bound by Brd4-Nut(FL) and Brd4-Nut(MIN), which is much higher than expected at random (1). Additionally, this overlap is statistically significant, as shown via p-value and the representation factor. The representation factor is the number of overlapping genes divided by the expected number of overlapping genes, when randomly drawn from two independent groups. If this value is larger than 1, it suggests that the overlap between groups is lower than expected at random. Here, the representation factor is 33.4, indicating that the gene occupancy by either Brd4-Nut(FL) or Brd4-Nut(MIN) is highly similar (Fig.3.8A, right-side diagram). It is important to note that, while the gene occupancy is highly similar between the two cell lines, there are many more genes that are occupied by Brd4-Nut(MIN) than by Brd4-Nut(FL) (Fig.3.8A, left-side diagram). This suggests that gene occupancy is more restricted for Brd4-Nut(FL) than for Brd4-Nut(MIN). This finding was corroborated upon closer examination of ChIPseq tracks, with majority of ChIPseq peaks coinciding at the same genomic loci for both cell lines, but some additional ones only appearing upon expression of Brd4-Nut(MIN) (Fig.3.8B). I hypothesize, based on the higher level of gene occupancy by Brd4-Nut(MIN), that residues 506-1132 of Nut, predicted to lack a defined secondary structure, may restrict access to certain loci in the full-length fusion protein (see Fig.1.3 for reference to the predicted secondary structure of Nut). Analogous to the ChIPseq results, the RNAseq data show that

a large portion of differentially expressed genes are common between Brd4-Nut(FL) and Brd4-Nut(MIN) – expressing cells (Fig.3.8C). The number of overlapping genes in RNAseq is more than 10 times higher than expected at random, for both up- and downregulated genes, indicating that there is a significant overlap in transcriptional profiles of these two cell lines. However, there are many genes, whose expression is altered exclusively upon expression of Brd4-Nut(FL) or Brd4-Nut(MIN) (Fig.3.8C). This again suggests that residues 506-1132 of Nut, which are mostly predicted to lack folded secondary structure, provide binding specificity and/or differentially recruit additional elements of the transcriptional machinery to particular genomic loci. A similar kind of observation has been described in the literature, where IDRs of transcription factors can direct specific binding to genomic loci (Brodsky et al. 2020). I also integrated the ChIPseq and RNAseq results for cells expressing either construct. To do so, I focused on the genes occupied by both proteins, as measured by ChIPseq annotations at promoter-TSS (Fig.3.8A, intersection in Venn diagram). I then analyzed the differential expression patterns of these genes via RNAseq and found that they are similarly up- or downregulated between both cell lines (Fig.3.8D). Several of the commonly upregulated genes here encode transcriptional regulators, e.g. CHD3 – a chromatin remodeler, FOXM1 – transcriptional activator, SETD2 and SETD5 – histone methyltransferases responsible for producing a histone methyl mark that is associated with active chromatin, MSL1 – predicted to enable chromatin binding activity and involved in histone acetylation. Some others encode proteins

involved in apoptotic signal transduction, including EIF5A, NGFRAP1 and FAIM. Finally, ZMYND11 encodes a zinc finger protein sharing a highly similar structural architecture to ZMYND8, which has been previously reported to be fused with NUT in some cases of Nut Carcinoma (Boyson et al. 2021). Together, these genetic analyses suggest that binding of both Brd4-Nut(FL) and Brd4-Nut(MIN) to specific promoter-TSS regions induces similar changes in gene expression. Therefore, I conclude that the Brd4-Nut(MIN) fusion is sufficient to mimic a large portion of the gene expression changes observed upon expression of Brd4-Nut(FL).

In summary, I discovered that just the minimal part of Nut, responsible for interaction with p300 (MIN), in fusion with Brd4 is sufficient to recapitulate majority of the transcriptional changes caused by Brd4-Nut(FL). Both gene occupancy and differential gene expression are highly similar for cells expressing either of the two fusion proteins. However, Brd4-Nut(MIN) binds to more genes, likely due to lack of the C-terminal IDR of Nut, which might be providing additional binding specificity.

Brd4-p300 recapitulates Brd4-Nut – mediated condensate formation and transcriptional changes

Since Brd4-Nut(MIN) mimics the condensate formation and transcriptional activity of Brd4-Nut(FL) to a large extent, I hypothesized that recruitment of p300 to Brd4-Nut condensates may be the main function of the Nut portion of the fusion protein. To study the roles of p300 in more detail, I designed an alternative

approach, in which Brd4 is fused directly to p300 (Fig.3.9A). This eliminates additional functions that Nut might be providing and focuses solely on p300. To start working towards this approach, I first established a new stable cell line, expressing mNeonGreen-tagged Brd4-p300 fusion protein in an inducible manner. I then established that the protein can be expressed to a similar level as Brd4-Nut(FL) (Fig.2.3). Next, I examined the ability of this new stable cell line to form condensates

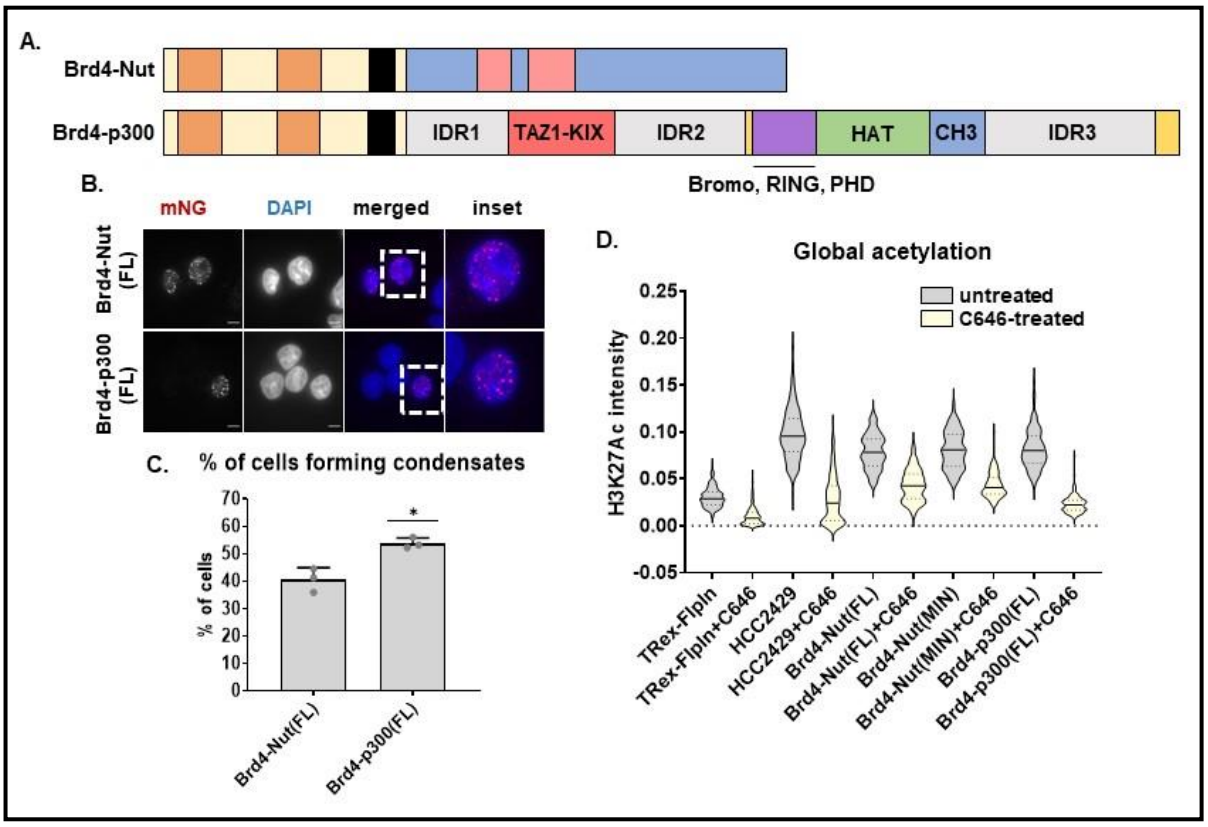

Fig.3.9.: Brd4-p300 fusion forms condensates and has similar acetylation pattern to Brd4-Nut. A. *Schematic of Brd4-Nut and Brd4-p300.* ***B.*** *Micrographs of Brd4-Nut and Brd4-p300 – expressing stable cell lines. Cells are stained with mNeonGreen antibody; scale bars = 10µm.* ***C.*** *Quantification of the micrographs represented in B.* ***D.*** *Quantification of histone H3K27Ac staining in untreated vs. C646-treated cells. Global acetylation shown as average fluorescence intensity across the entire cell nucleus.*

and I determined that cells expressing Brd4-p300 fusion can form condensates to a similar degree as Brd4-Nut, although the fraction of cells expressing Brd4-p300 fusion that can form condensates is slightly higher (Fig.3.9B,C). I further performed an immunofluorescence-based confocal microscopy experiment, where I stained the different cell lines with a α-H3K27Ac antibody and measured the average intensity of fluorescence across the cell nucleus. I then used the average fluorescence intensity as a proxy for average acetylation level in the nucleus. I found that – similarly to HCC2429 cell line – stable cell lines expressing Brd4-p300, Brd4-Nut(FL) and Brd4-Nut(MIN) each show elevated average acetylation in the cell nucleus, as compared to 293TRex-FlpIn cells (Fig.3.9D, gray violins). In all cases this acetylation level decreases upon treatment with C646 (Fig.3.9D, yellow violins). Through the experiments described above, I confirmed that the Brd4-p300 construct causes condensate formation to a similar degree as Brd4-Nut(FL) and Brd4-Nut(MIN). I also showed that all these constructs cause elevated acetylation in cell nuclei, as compared to 293TRex-FlpIn cells, which do not express any fusion transgene. This acetylation result is similar to what I observed in patient-derived Nut Carcinoma cell line, HCC2429.

Based on my previous data, I expected the Brd4-p300 construct to most closely resemble Brd4-Nut(MIN) in its activity. As opposed to Brd4-Nut-based constructs, I could not perform a direct genome occupancy study for Brd4-p300 via ChIPseq, as the antibody used for the Brd4-Nut constructs-expressing cells was an α-Nut antibody. Here, we could not use α-Nut antibody due to lack of the

appropriate epitope. Additionally, neither an α-Brd4 nor α-p300 antibody could be used, due to the wild-type Brd4 and p300 being expressed in the background in all stable cell lines. Finally, at the time of these experiments, there was no ChIP-grade α-mNeonGreen antibody available on the market. Thus instead, I examined the genome-wide acetylation profile, using ChIPseq against the H3K27Ac mark characteristic of active genes. I also performed RNAseq-based transcriptome analysis of cell lines expressing Brd4-p300-based constructs. I found that almost all the same genes are acetylated on histone H3K27 upon expression of either Brd4-Nut(MIN) or Brd4-p300(FL) (Fig.3.10A, left-side diagram). When I restrict the analysis to only gene annotations with the highest normalized signal value, overlap between the two cell lines is still significant, with the number of genes commonly annotated to have the H3K27Ac mark more than 7 times higher than predicted at random and the representation factor of the overlap of 6.0 (Fig.3.10A, right-side diagram). This suggests that Brd4-p300(FL) and Brd4-Nut(MIN) are likely activating transcription of similar genes. I also analyzed RNAseq results from the stable cell lines and compared them to the control 293TRex-FlpIn cell line, which does not express any fusion protein. I found that a large fraction of differentially expressed genes are commonly up- or downregulated in both Brd4-Nut(MIN) and Brd4-p300(FL) – expressing cells (Fig.3.10B). The fraction of genes up- or downregulated in both cell lines is more than 20-fold greater than expected by random overlap. Finally, I performed a ChIPseq-RNAseq integration, where I analyzed the expression patterns of the genes that were commonly acetylated

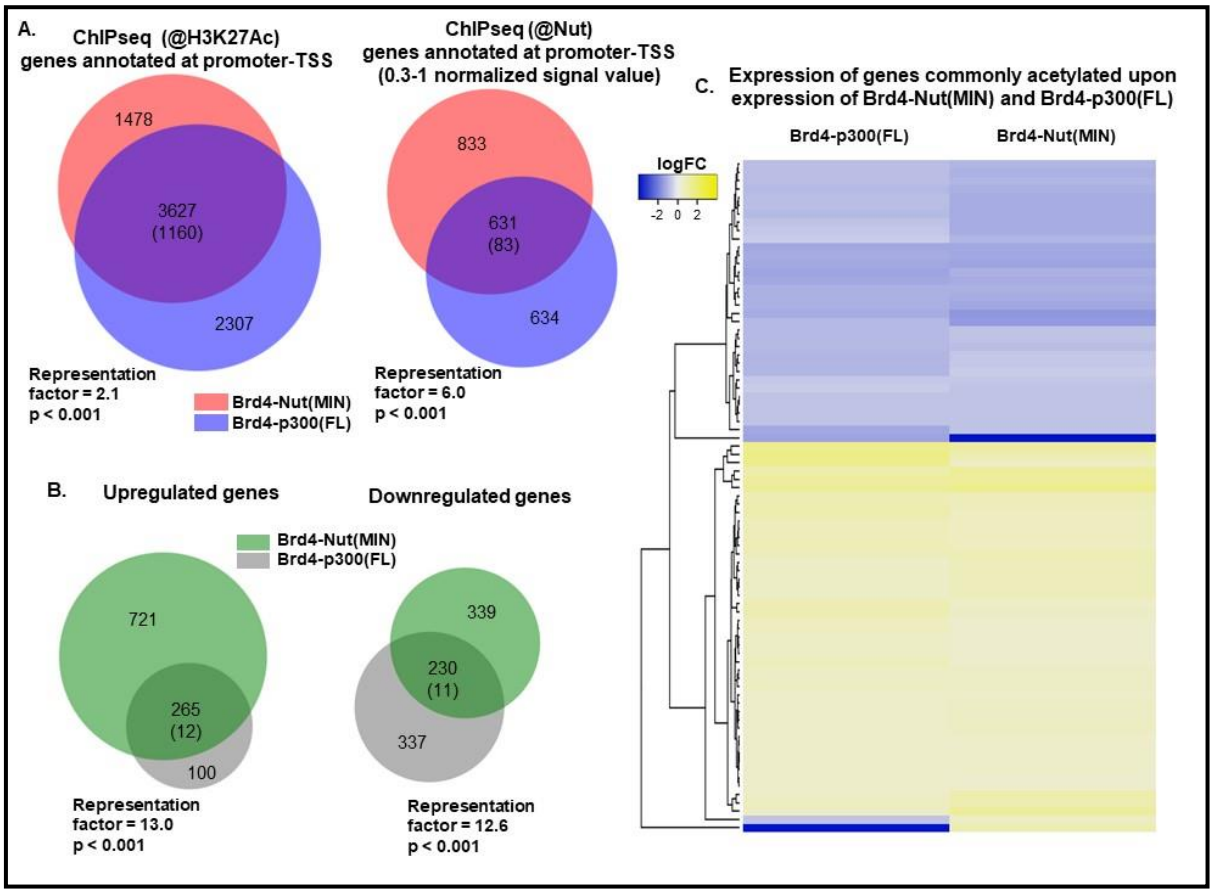

Fig.3.10.: Brd4-p300 fusion forms condensates and has similar acetylation pattern to Brd4-Nut. A. *Venn diagrams of genes acetylated at histone H3K27 site upon expression of Brd4-Nut(MIN) and Brd4-p300(FL) on promoter-TSS regions of genes. Left-side diagram represents all annotated genes with H3K27Ac mark at promoter-TSS region. Right-side diagram additionally includes a 0.3-1 normalized signal value cutoff. Numbers in parentheses represent expected overlap at random.* ***B.*** *Venn diagrams of RNAseq genes that are significantly up- and down-regulated upon expression of Brd4-Nut(MIN) or Brd4-p300(FL). Genes are only shown when fold change in expression was greater than two and p<0.05. Numbers in parentheses represent the mean overlap in 20 iterations with randomly generated genes from the human genome.* ***C.*** *RNAseq-ChIPseq data integration: heatmap showing up- and down-regulated genes found via RNAseq, that are bound by both Brd4-Nut(FL) and Brd4-Nut(MIN). Top up- and downregulated genes shown in the heatmap, based on log fold change, where logFC > 0.5 and < (-0.5).*

upon expression of either Brd4-Nut(MIN) or Brd4-p300(FL) (Fig.3.10A, overlap in the right-side Venn diagram). Out of the 631 common genes, I limited the number

of genes represented in the heatmap to 81, based on the most significant log fold change. This analysis showed that indeed both cell lines present highly similar differential gene expression patterns overall (Fig.3.10C). Among other genes found through this analysis, many encode zinc finger transcription factors, including: *ZBTB25*, *ZNF213*, *ZNF644*, *ZNF408*, *ZNF583* and *ZMYND8.* Interestingly, *ZMYND8* has been previously found as one of the fusion partners of *NUT* in some Nut Carcinoma patients (Boyson et al. 2021). Only one of these zinc finger transcription factors is downregulated, with the other five being upregulated. Furthermore, *EPC1* and *EPC2*, both found to be acetylated on H3K27 and both highly upregulated upon expression of either Brd4-Nut(MIN) or Brd4-p300(FL), encode chromatin binding proteins. Interestingly, there are many genes in this analysis that are downregulated, regardless of the activating acetylation marks on them. While I have not studied this phenomenon in depth, I hypothesize that the Brd4-Nut and Brd4-p300 condensates might play a repressing role on transcription of some genes, by potentially sequestering specific transcription factors and thus, preventing them from activating anti-tumorigenic genes and genes responsible for cell differentiation. Such effect has been proposed previously, in the context of EWS-Fli1, condensates, where it was proposed that increasing the condensate formation in this case results in sequestration of EWS-Fli1 and thus, inhibits EWS-Fli1-driven transcriptional program (Chong et al. 2022).

All these results support the idea that the MIN fragment of Nut is important in recruiting p300 to Brd4-Nut, and that acetyltransferase is responsible

for a significant portion of transcriptional changes observed in Brd4-Nut – expressing cells. Additionally, Brd4-Nut(MIN) and Brd4-p300(FL) – expressing cells behave similarly, both in terms of condensate formation and transcription. Having confirmed that recruitment of p300, either through binding in the case of Brd4-Nut(MIN) or through covalent attachment in the case of Brd4-p300(FL), is sufficient to recapitulate condensate formation and many of the transcriptional changes in cells expressing Brd4-Nut, we next examined what molecular features of p300 are responsible for these processes.

p300 IDRs, TF binding and enzymatic activity contribute to condensate formation

Having confirmed that recruitment of p300 is sufficient to recapitulate condensate formation and many of the transcriptional changes in cells expressing Brd4-Nut, I next sought out to examine what molecular features of p300 are responsible for these processes. The p300 histone acetyltransferase is a very complex molecule, with a plethora of different domains within its structure. Arguably, the most important domain is the histone acetyltransferase domain (HAT) – the catalytic center of the protein. In addition to HAT, p300 also contains multiple other folded domains, which are responsible for binding different molecular elements of transcriptional machinery. These domains include: bromodomain, PHD domain, ZZ, TAZ1, TAZ2, KIX and RING domains. Some of these domains can be seen in a structure reported earlier in this dissertation, in Fig.1.6 (Ortega et al. 2018).

Finally, there are also three long stretches of predicted intrinsically disordered regions within p300, which I named here IDR1, IDR2 and IDR3 (Fig.1.5B). All these different molecular parts can be generally categorized into three classes of features: (1) histone acetyltransferase (HAT), (2) folded transcription factor – binding domains (TF-binding domains) and (3) intrinsically disordered regions (IDRs) of p300. In my next steps, I designed a series of Brd4-p300 fusion mutations, with different molecular parts of p300 deleted or inactivated via a point mutation. The nomenclature for these mutants will collectively refer to the three classes of features, where “H” will stand for HAT domain, “T” will stand for all the TF-binding domains, and “I” will stand for all three IDRs of p300.

Protein IDRs have been often shown to be involved in biomolecular condensate formation (Pak et al. 2016; Lin, Currie, and Rosen 2017; Banjade et al. 2015; Nott et al. 2015; Burke et al. 2015). p300 possesses three long IDRs (IDR1, IDR2 and IDR3) (Erdos and Dosztanyi 2020; Meszaros, Erdos, and Dosztanyi 2018) (Fig.3.11A, 1.5B). Therefore, I first decided to investigate the role of p300 IDRs in condensate formation. To this end, I generated two new constructs: (1) just the IDRs of p300 directly fused to Brd4 [Brd4-p300(I)], and (2) all three IDRs deleted, with all structured domains of p300 directly fused to Brd4 [Brd4-p300(ΔI)] (Fig.3.11A). I attempted to develop two new stable cell lines but was only successful with the Brd4-p300(I) construct. Expression of the Brd4-p300(ΔI) in cells was highly toxic. Thus, all following experiments with Brd4-p300(ΔI) have been performed via transient transfections, rather than through use of a stable cell line. To use

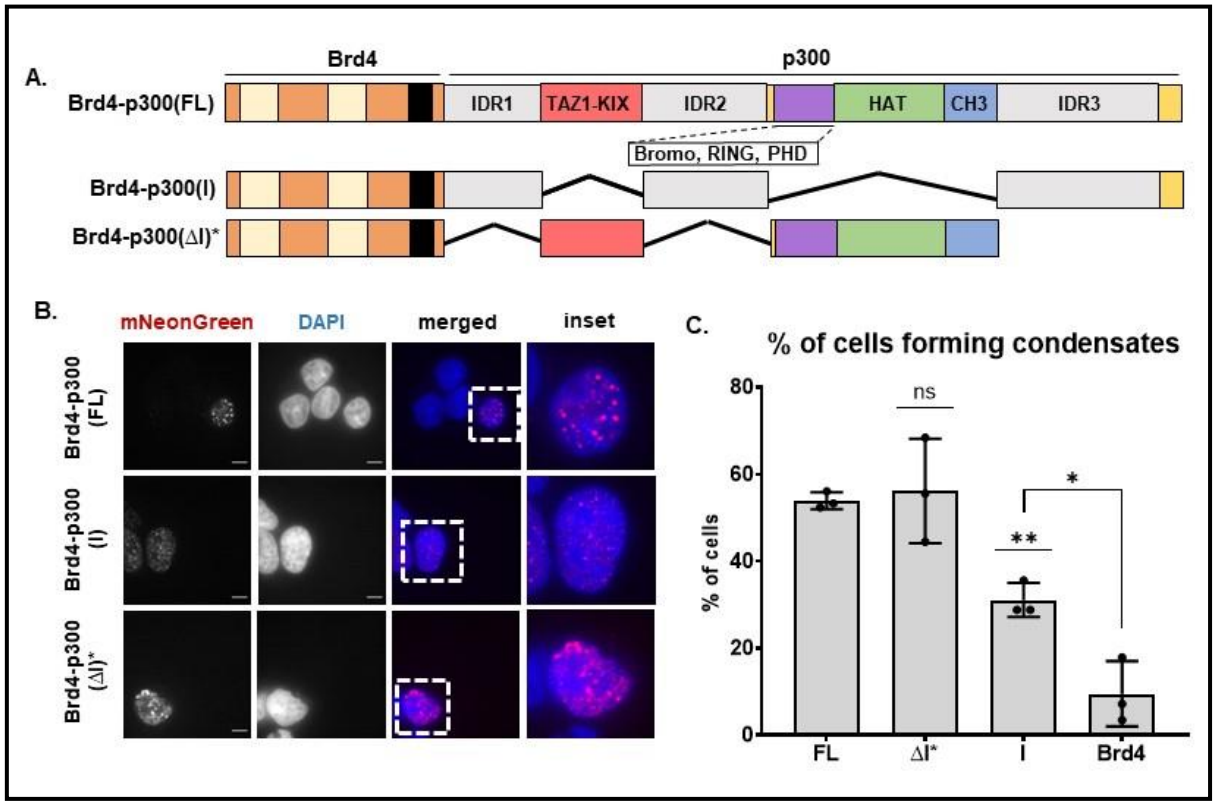

Fig.3.11.: p300 IDRs are dispensable for condensate formation. A. *Schematics of BRD4-p300(FL), (I) and (ΔI) constructs; different domains and disordered regions are indicated.* ***B.*** *Micrographs of all the Brd4-p300 constructs from A; scale bars = 10 µm. Data are a part of the same experiment as in Fig.3.8)* ***C.*** *Quantification of the micrographs represented in B.*

the Brd4-p300(ΔI) construct in the microscopy analyses and compare its ability to form condensates with stable cell lines expressing other fusion constructs, I applied stringent expression level cutoffs to the image analyses, like described in detail in Chapter 2. While I was able to apply such expression cutoff to the imaging data, I was not able to apply a similar data filtering strategy for the transcriptional analyses. Therefore, I was not able to perform RNAseq or ChIPseq with cells expressing Brd4-p300(ΔI). I analyzed the expression

of Brd4-p300(I) and Brd4-p300(ΔI) in cells and found that fewer cells expressing Brd4-p300(I) form condensates, than with Brd4-p300(FL) (Fig.3.11B,C). Conversely, cells expressing Brd4-p300(ΔI) at a comparable level form condensates to a similar extent as Brd4-p300(FL) (Fig.3.11B,C). However, the morphology of these condensates is quite different from the Brd4-p300(FL) ones: the condensates lacking p300 IDRs appear larger and less round (Fig.3.11B). Given the significant differences in shape compared to the other systems, I believe that just reporting the difference in fraction of cells able to form condensates as compared to the full-length construct may not accurately reflect the IDR functionality. In additional comparison, I discovered that there is a substantial difference in large condensates formation between cells expressing Brd4 alone and Brd4-p300(I) (Fig.3.11C). Thus, I conclude that while the IDRs of p300 are not a sole driver of Brd4-p300 condensation, they do promote condensate formation when in fusion with Brd4. Nevertheless, there are likely other elements of p300 that play additional, important roles in promoting formation of Brd4-p300 condensates.

I next tested whether the histone acetyltransferase (HAT) activity of p300 is important in condensate formation. I designed additional constructs, including: (1) Brd4-p300(H*IT) fusion with a HAT-inactivating point mutation (D1399Y (Ito et al. 2001)), which retained the IDRs and structured domains of p300 (folded domains collectively named “TF-binding domains”); and (2) Brd4-p300(HI), in which I kept the IDRs and the intact HAT domain fused to Brd4, removing all the TF-binding domains of p300 (Fig.3.12A). I found that inactivating the HAT domain via a point

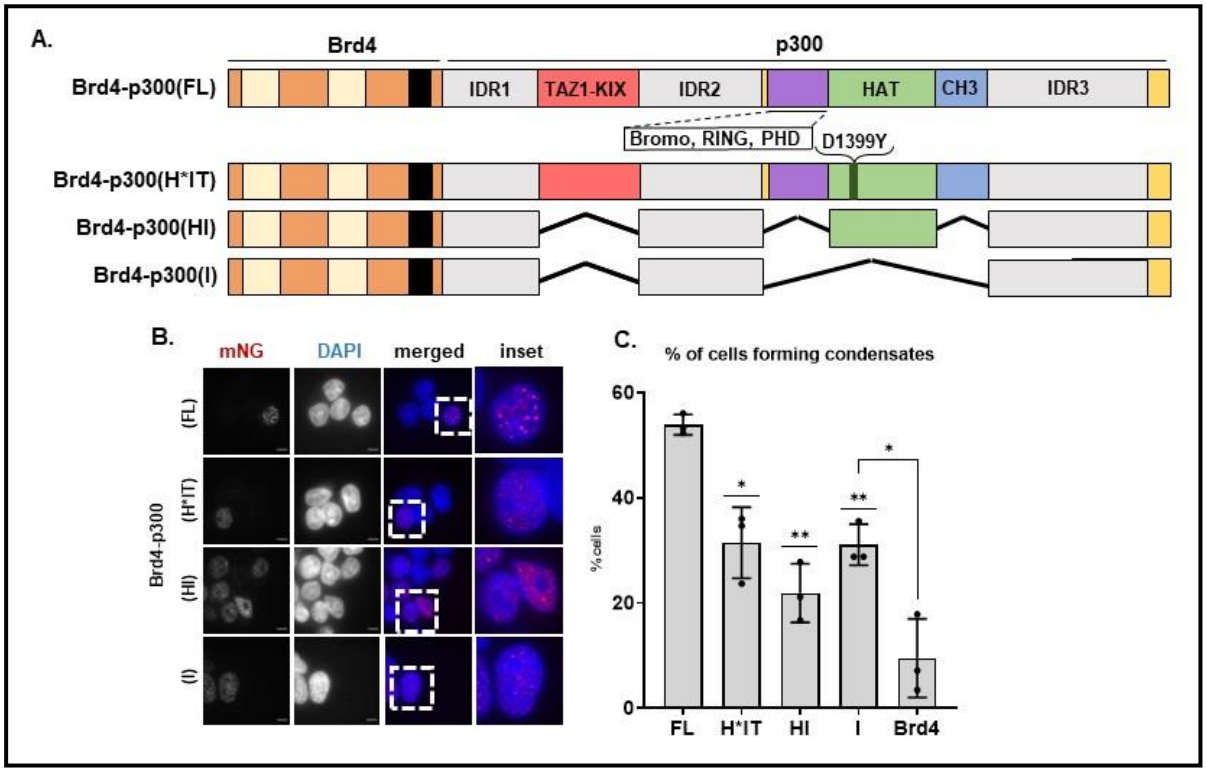

Fig.3.12.: IDRs, HAT and TF binding domains of p300 collectively contribute to condensate formation. A. *Schematic of Brd4-p300 – based constructs, including FL, H*IT, HI and I.* ***B.*** *Micrographs of all the Brd4-p300 constructs from a); scale bars = 10 µm. These data are a part of the same experiment as shown in figure 3.11 and 3.8.* ***C.*** *Quantification of the micrographs represented in B. Brd4 alone shown for reference.*

mutation in Brd4-p300(H*IT) construct decreases the percentage of cells forming condensates (Fig.3.12B,C). Expression of the Brd4-p300(HI) construct also results in fewer cells forming condensates, as compared to Brd4-p300(FL) (Fig.3.12B,C). Interestingly, while both TF-binding domains and HAT are important for condensate formation to some extent, when both are missing the effect is not additive. The fraction of cells forming condensates upon expression of a construct lacking HAT activity but retaining TF-binding [Brd4-p300(H*IT)] is similar to that of a construct lacking TF-binding domains but retaining HAT [Brd4-p300(HI)]

and one that lacks both TF-binding and HAT activity [Brd4-p300(I)]. This indicates that both TF-binding and HAT activity are required for condensate formation and neither one alone is sufficient to restore the condensate formation to the full extent of Brd4-p300(FL). Finally, retaining only IDRs of p300 in fusion with Brd4 is also sufficient for more cells to form condensates than upon expression of Brd4 only. These results together suggest that all three classes of molecular features of p300: IDRs, HAT and TF-binding are required for the formation of condensates in cells to the same extent as Brd4-p300(FL).

Since I have established that p300 acetyltransferase activity contributes to Brd4-Nut condensate formation, I next sought out to examine the acetylation patterns in cells expressing the different Brd4-p300 constructs. I found that there

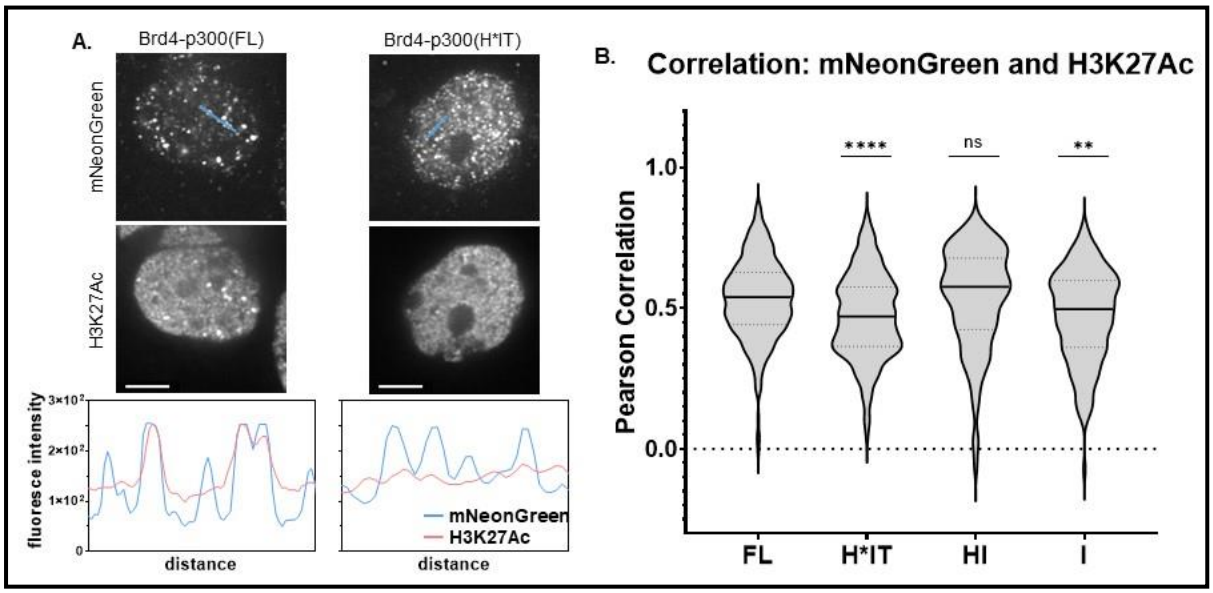

Fig.3.13.: Analyses of acetylation patterns among the Brd4-p300 mutant constructs. A. *Colocalization of Brd4-p300(FL) or Brd4-p300(H*IT), as visualized by α-mNeonGreen staining, with α-H3K27Ac staining, shown in micrographs and via line profiles.* ***B.*** *Quantification of correlation between α-mNeonGreen and α-H3K27Ac condensates, as shown via Pearson Correlation.*

is a high level of overlap between immunostaining with α-mNeonGreen antibody and α-H3K27Ac antibody in cells expressing Brd4-p300(FL) (Fig.3.13A, left side). Conversely, expression of a HAT-deficient mutant in cells [Brd4-p300(H*IT)] exhibits a much lower level of colocalization, as shown via line profiles (Fig.3.13A, right side). Moreover, when I measured the colocalization between α-mNeonGreen antibody and α-H3K27Ac antibody channels via Pearson Correlation in stable cell lines expressing all Brd4-p300 fusion mutants, I discovered that the Brd4-p300(FL) and Brd4-p300(HI) constructs both have higher levels of colocalization between channels than either of the HAT-deficient mutants [Brd4-p300(H*IT) and Brd4-p300(I)] (Fig.3.13B). These results together suggest that acetylation is concentrated in the condensates and that presence of an active HAT domain is necessary for this effect.

These data together indicate that IDRs, HAT and TF-binding all contribute to Brd4-p300 condensate formation. Furthermore, none of these elements alone is sufficient to cause condensate formation to the same extent as Brd4-p300(FL). It is possible that the three classes of molecular elements of p300 contribute to a combination of mechanisms. TF-binding domains along with HAT might be activating a positive feedback loop, causing hyperacetylation of chromatin and increasing the recruitment of additional Brd4-Nut molecules. At the same time, IDRs might act in an independent, self-association mechanism to enhance condensate formation.

p300 acetyltransferase activity is important for transcriptional changes in cells

Finally, I wanted to understand if and how changes in condensate formation and acetylation are reflected in transcriptional profiles. To address this question, I performed analyses of transcriptional changes upon expression of the Brd4-p300 mutant constructs. I first analyzed the trends in differential gene expression in stable cell lines inducibly expressing all the Brd4-p300 mutant fusion proteins, using RNAseq. I used the 293TRex-FlpIn cell line without any additional transgene as a point of reference. Through this analysis, I found that removal of HAT activity results in downregulation of most genes [Fig.3.14A, ~74% of all differentially expressed genes for Brd4-p300(H*IT) and ~60% for Brd4-p300(I)], whereas presence of active HAT [Brd4-p300(HI)] results in majority of genes being upregulated (Fig.3.14A, ~54%). This finding is consistent with the idea that higher acetylation level in the genome results in a more open chromatin and thus, activation of transcription. I also analyzed the expression of two of three Nut Carcinoma signature genes (*SOX2* and *TP63*) and found that they are both significantly upregulated in all cell lines expressing constructs with active HAT [Brd4-p300(FL) and Brd4-p300(HI)] (Fig.3.14B). On the contrary, in stable cell lines expressing Brd4-p300 mutations that lack HAT activity [Brd4-p300(H*IT) and Brd4-p300(I)], *SOX2* is downregulated and *TP63* is only slightly upregulated, but this upregulation effect is not statistically significant (Fig.3.14B). The third signature gene of NC – *MYC* – was not observed to be upregulated in our stable cell lines, but this

phenomenon has been described in non-NC cell lines (Alekseyenko et al. 2015; Eagen and French 2021). This result suggests that p300 HAT activity regulates transcriptional activation by the Brd4 fusions. Lastly, I performed RNAseq principal

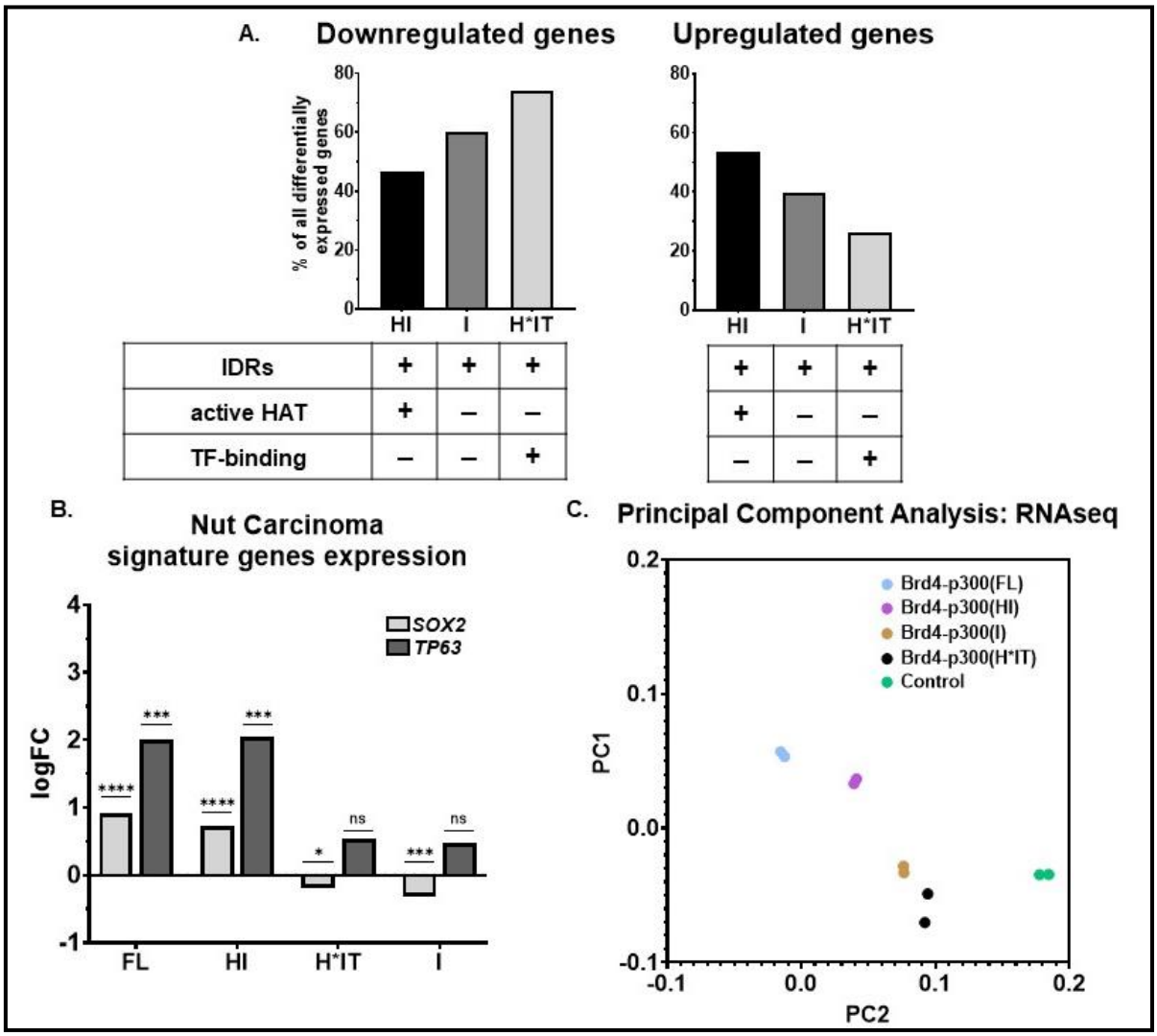

Fig.3.14.: Acetylation activity of p300 is necessary for transcriptional changes observed upon expression of Brd4-p300(FL). A. *Percent of all differentially expressed genes that are upregulated or downregulated, upon expression of different Brd4-p300 constructs.* ***B.*** *Table showing logFC for SOX2 and TP63 genes upon expression of different Brd4-p300 constructs.* ***C.*** *PCA plot summarizing RNAseq data for cells expressing all constructs. Cell lines excluded in this plot, present in full-size plot: Brd4-Nut(FL) and Brd4-Nut(MIN).*

component analysis (PCA) to compare transcription changes upon expression of all the Brd4-p300 constructs (Fig.3.14C). Through this analysis, I found that the Brd4-p300(H*IT)) and Brd4-p300(I) – expressing cells cluster closely to control (239TRex-FlpIn), especially when we consider the relatively higher contribution of PC1 to variability of the data (Fig.3.14C, y axis). This result indicates that lack of active HAT domain removes most of the transcriptional activity caused by Brd4-p300, making the gene expression more similar to cells that do not express any fusion protein. On the other hand, cells expressing Brd4-p300(FL) or Brd4-p300(HI) are clustered more closely together (Fig.3.14C), suggesting that just presence of an active HAT domain is sufficient to make the transcriptional profiles resemble the cells expressing the full-length Brd4-p300. In contrast, the presence or absence of TF-binding domains does not have a major effect on the transcription, as signified by the close proximity of data from Brd4-p300(H*IT) and Brd4-p300(I) – expressing cells (Fig.3.14C)

All these results taken together suggest that all three: IDRs, HAT activity and TF-binding collectively contribute to condensate formation, whereas only HAT activity seems most important for transcriptional changes in cells. The weight of each of these molecular elements is thus different when it comes to either condensate formation or modulation of gene expression.

Discussion

In my dissertation, I studied the connection between Brd4-Nut condensates formation and transcriptional changes in cells. I showed that p300 recruitment by Brd4-Nut is necessary and sufficient for condensate formation and different molecular features and functions of p300 collectively contribute to this process (Fig.3.15). However, only the acetylation activity of p300 HAT domain appears required to alter transcription (Fig.3.15). Thus, Brd4-Nut and Brd4-p300 fusion proteins integrate multiple functions and the regions responsible for condensate formation and transcription are not the same. I conclude that condensate formation and transcriptional changes upon expression of Brd4-Nut or Brd4-p300 might be decoupled to some extent.

Readout	Order of constructs	Most important factors
Condensate formation	FL > HAT+IDR (ΔTF) = TF+IDR (ΔHAT) = IDR (ΔTF+HAT)	TF HAT IDR
Acetylation localized to condensates	FL = HAT+IDR (ΔTF) > TF+IDR (ΔHAT) = IDR (ΔTF+HAT)	HAT
RNAseq	FL > HAT+IDR (ΔTF) >> TF+IDR (ΔHAT) = IDR (ΔTF+HAT)	HAT

Fig.3.15.: A summary of molecular parts of p300 required for condensate formation, condensate-localized acetylation and transcriptional changes in cells.

My studies were divided into two main parts: 1) condensate formation analysis and 2) gene expression analysis. To study both these aspects, I developed a series of stable cell lines, which all express different fusion proteins in an inducible manner. I controlled the expression levels very carefully, to make sure that my engineered system mimics the expression of Brd4-Nut in Nut Carcinoma cell line, HCC2429, as closely as possible. Next, I focused my microscopy-based experiments on large condensates that I can confidently differentiate from the background antibody staining. Based on empirical analyses (Fig.2.2, 2.3 and 2.4), I determined that size cutoff of 1.25 μm diameter per condensate is a secure choice. This size cutoff is equal to ~2xPSF of the microscope. However, keeping in mind that transcriptional biomolecular condensates are generally very small (Sabari et al. 2018; Forman-Kay et al. 2022), I acknowledge that a super-resolution microscopy approach might be advantageous, to understand this system in more detail. Indeed, many of the constructs I designed seem to form small, granular structures across the entire cell nucleus. Some of them form only a few large condensates that I focused the analyses on. In the future, it would be beneficial to analyze the differences in capacity of cells to form condensates as small as ~100 nm.

While I could set up thorough cutoffs for protein expression level based on fluorescent signal in microscopy on a cell-to-cell basis, such an approach was not feasible for batch experiments, such as RNAseq and ChIPseq. Here, I did my best to make the protein expression induction regime well fixed, so that there is little

variability from cell line to cell line. Nonetheless, I realize that there are some differences in the level of protein expression between cell lines (Fig.3.5A). These differences must be taken into account when drawing conclusions based on the transcriptomic data. In the future, it would be valuable to perform single cell RNAseq, combined with rapid fluorescence measurement to control protein expression in each cell. At this point, technology does not yet allow for such fast protein expression control and automated RNA extraction for sequencing.

Through the experiments, I found that the p300-interaction motif of Nut (MIN) in fusion with Brd4 is necessary and sufficient to form condensates. Indeed Brd4-Nut(MIN) forms condensates to a similar degree as full-length Brd4-Nut. While the remaining IDR parts of Nut are not critical for condensate formation, they have additional roles in transcription. Binding to genome is more restricted for Brd4-Nut(FL) than for Brd4-Nut(MIN) (Fig.3.8A) and it is possible that by designing the Brd4-Nut(MIN) construct, lacking the IDRs of Nut, I might have removed the binding specificity from Brd4-Nut(FL). This would allow the smaller fusion protein to bind to more spots in the genome, in a less specific manner, even though the expression level of the two fusion proteins is comparable (Fig.2.3). Such an effect would not be unprecedented. A similar phenomenon has been described where IDRs of transcription factors can assist in directing genomic binding specificity (Brodsky et al. 2020).

p300 is recruited into Brd4-Nut condensates even in the context of the minimal Brd4-Nut(MIN) construct. The transcriptional profile of cells

expressing either Brd4-Nut(MIN) or Brd4-p300(FL) is also similar. Thus, I focused on the roles of p300 in condensate formation and gene expression changes. I found that there are multiple different molecular features collectively contributing to biomolecular condensation and transcription: histone acetyltransferase activity (HAT), structured TF-binding domains and IDRs of p300. However, I determined that there is little overlap between the features that contribute to condensate formation and to transcriptional changes, as only HAT activity of p300 strongly influences transcriptional output. Importantly, these results demonstrate that it is not condensate formation, per se, that drives transcriptional activity, as p300 mutants lacking active HAT domain [Brd4-p300(I) and Brd4-p300(H*IT)] form condensates to some extent, but lack transcriptional changes. In contrast, p300 mutant with an active HAT domain [Brd4-p300(HI)] forms condensates to a similar extent as HAT-less mutants and regulates transcription similarly to Brd4-p300(FL). Thus, my careful molecular dissection suggests that the features that contribute to condensate formation and to transcriptional regulation are overlapping, yet distinct. Nonetheless, the biologically relevant molecules (Brd4-Nut fusion protein and full-length p300) combine these features to stimulate condensate formation and transcriptional regulation.

CHAPTER FOUR

CONCLUSIONS AND FUTURE DIRECTIONS

Summary of the studies

In this dissertation, I studied Brd4-Nut – a fusion oncoprotein involved in the pathogenesis of Nut Carcinoma. The Brd4-Nut fusion has been shown to form nuclear condensates in cells. I sought to understand if and how Brd4-Nut condensate formation might be related to changes in transcriptional profiles of cells. I also explored the roles that a Brd4-Nut interactor, p300 acetyltransferase, and its different molecular parts might play in this system.

To perform these molecular studies, I designed a series of mutant constructs of Brd4-Nut and p300 and developed stable cell lines, expressing each of these constructs in an inducible manner. I carefully controlled the expression levels of proteins in my stable cell lines, to mimic the expression of Brd4-Nut in the patient-derived Nut Carcinoma cell line, HCC2429. Then, I examined the ability of different fusion protein mutants to form condensates and to induce gene expression changes in cells. I found that only a small fragment of Nut (MIN) is required in fusion with Brd4 for condensate formation and that the same fragment is also responsible for recruitment of p300 into the condensates (Fig.3.4). Concomitantly, I established that the transcriptional profiles of cells expressing either full-length or minimal Brd4-Nut fusion are similar (Fig.3.8). Having determined the role of p300-interaction motif of

Nut (MIN) in condensate formation and transcription, I next moved on to study the roles of p300 itself in both these processes. First, I made a Brd4-p300 fusion construct and I found that cells expressing this fusion form condensates to a similar extent as cells expressing Brd4-Nut(FL) or Brd4-Nut(MIN) (Fig.3.9). Then I also showed that the Brd4-p300-induced gene expression changes in cells resemble the transcriptional profiles of cells expressing Brd4-Nut(MIN) (Fig.3.10). Subsequently, I focused on the different molecular parts of p300, including its HAT domain, IDRs and TF-binding domains, and their roles in condensate formation and transcription. Here, I found that all three: HAT, IDRs and TF-binding domains are needed for condensate formation in cells, (Fig.3.11and 3.12). At the same time, only HAT appears to have a significant role when it comes to causing high levels of acetylation in condensates (Fig.3.13). Accordingly, HAT also seems to be the most important molecular part of p300 that is responsible for changes in transcription upon expression of the fusion proteins (Fig.3.14 and 3.15). All in all, in my dissertation I discovered that the molecular features of Brd4-Nut and p300 that are responsible for condensate formation and transcription in cells are distinguishable. While there is a correlation between condensate formation and transcriptional changes occurring in a patient-derived Nut Carcinoma cell line, condensate formation per se is likely not the cause for these transcriptional changes to occur.

Brd4-Nut condensates and transcription

It has been previously proposed that biomolecular condensates might play important roles in concentrating transcriptional machinery and activating multiple genes at the same time in mammalian cells. Several groups have shown that important elements of transcriptional machinery, including RNA polymerase II (RNA Pol II), the over 1MDa-sized Mediator complex and several other transcriptional activators all form condensates in cell nuclei, and transcription factors also enter such transcriptional condensates, at super enhancer sites. These condensates have been shown to likely be involved in formation of transcriptional condensates and participates in regulation of all essential steps in transcription. (Sabari et al. 2018; Shrinivas et al. 2019; Lyons et al. 2023; Guo et al. 2019; Boija et al. 2018; Cho et al. 2018). Among others, Lyons et al. showed that increased local concentration of the MED1 IDR allows for positive transcriptional regulators to partition into a transcriptional condensate and thus, to activate gene expression (Lyons et al. 2023). At the same time, a negative transcriptional regulator, showed an opposite effect (Lyons et al. 2023). Cho et al. discovered that Mediator and RNA Pol II colocalize in condensates only when transcription is active (Cho et al. 2018). What is more, the same group used super resolution microscopy methods to show that Mediator condensates dynamically “kiss” actively transcribing genes (Cho et al. 2018). The connection between condensate formation and gene expression has also been extensively studied in the context of nuclear fusion oncoproteins, which often form

condensates and drive aberrant transcription (Chandra et al. 2022; Shirnekhi, Chandra, and Kriwacki 2023; Boulay et al. 2017; Owen et al. 2021).

However, the validity of some findings on transcriptional condensates has been questioned, due to a very dynamic nature of these transcriptional condensates, as well as their small size (McSwiggen et al. 2019). Doubt has also been cast on some of the methods canonically used in studies of condensates, as these in some cases might lead to qualitative and phenomenological conclusions, rather than robust, quantitative measurements (McSwiggen et al. 2019; Musacchio 2022). Indeed, some simplified, engineered condensate model systems have led to conclusions that there should be a single threshold concentration for condensate formation (Nott et al. 2015; Shin et al. 2017). However, in the case of small, multi-component condensates this model does not hold true (Riback et al. 2020). Rather, heterotypic multi-component interactions, which are more abundant in endogenous cellular condensates, are much more complex and do not have a single, fixed threshold concentration for the formation of condensates. Moreover, McSwiggen et al. claim that the concentration of components within condensates and in bulk should be stable as a single condensate component is titrated (McSwiggen et al. 2019). This again is true only for single-component condensates in *in vitro* experiments and not for multi-component condensates in the cellular environments, especially the small, transcriptional condensates, where stochastic fluctuations in concentration are expected (Riback et al. 2020).

Findings that I report in this dissertation indicate that in the case of Brd4-Nut and p300, different molecular parts of the proteins are responsible for condensate formation and for gene expression changes. While the two processes seem to be decoupled to an extent, acetylation appears to be involved in both condensate formation and transcription. It would be an interesting next step to study the exact molecular mechanisms of Brd4-Nut condensate formation. Additional mutant constructs of both Brd4-p300 and Brd4-Nut would be needed to further dissect the molecular details of the involvement of specific TF-binding domains, HAT and IDRs in condensate formation. It would also be worthwhile to decipher what additional elements of transcriptional machinery might be involved in the formation of Brd4-Nut condensates and what additional roles the presence or absence of these different factors might have on transcription.

Technical considerations in studies of condensates

In my dissertation, I did my best to utilize quantitative methods in studies of biomolecular condensates and to make clear conclusions, without extrapolating the results. First, I developed a series of stable cell lines and spent a great effort to ensure that all the proteins expressed in these cell lines are expressed at low levels, approximating the expression level of Brd4-Nut in HCC2429 patient-derived Nut Carcinoma cells (Fig.2.3, 2.5 and 2.8). Next, I limited my microscopy analyses to only large, >1.25 μm in diameter condensates that I could confidently quantify via confocal microscopy. I recognize that by setting the size cutoff this high,

I have inevitably missed small transcriptional condensates in my analyses. Indeed, some of the fusion constructs that I designed form many small, granular condensates in cells, and it would be interesting to study these as well. However, the size cutoff was established in empirical analysis, where I found that it allows me to account for majority of condensates formed by all constructs and at the same time, filter out fluorescent foci that are not condensates, but rather result from non-specific antibody staining or out-of-focus condensates in the imaged field of view (Fig.2.10). I found that it is absolutely crucial during fluorescent image analysis to account for the background signal, to perform flat-field corrections, to measure the potential bleed through effect between fluorescence channels, and to establish reliable methods for cellular segmentation. Through a rigorous design of an image analysis pipeline, described in detail in Chapter 2, I hope to set a standard for careful and meticulous microscopy image processing in the field of biomolecular condensates.

Having performed a careful analyses of condensate formation, I next wanted to understand a link between condensate formation and transcription. It quickly became apparent that connecting my findings on condensate formation to transcriptional output would not be a trivial task either. Some laboratories have been able to draw conclusions on correlation between condensate formation and transcription, by performing assays in vitro, on a cellular level and by translational research approach, involving the same constructs that do or do not form condensates (Chandra et al. 2022; Ahmed et al. 2021). Nevertheless, to the best of

my knowledge, there is still no experimental assay that could directly confirm or contradict the connection between condensate formation and transcription. To ensure correctness of my conclusions for Brd4-Nut, whenever possible, I performed my studies in a few orthogonal approaches, such as immunostaining of the cells with a few different antibodies to visualize the condensates or performing an integrated ChIPseq and RNAseq analysis to report transcriptional profiles of cells expressing the different proteins. One confounding issue in examining correlations is that unlike in the case of condensate formation studies, which are performed for collections of individual cells, for bulk biochemical assays including for example ChIPseq or RNAseq, I could not apply expression level cutoffs to only analyze cells that express proteins at a desired level. This should be considered when analyzing the Brd4-Nut condensate and transcriptional data. Such analyses could in theory be performed on a single-cell basis. However, in this dissertation, the time-sensitivity of expression level tuning in the developed stable cell lines, made it impossible to combine the gene expression analyses with tests of protein expression levels.

Potential mechanisms of Brd4-Nut condensate formation

Upon discovering that there is a specific element of Brd4-Nut that is responsible for condensate formation, and that the same part of the protein recruits p300, I wanted to study p300 and its roles in the same way. I performed molecular dissections of p300 and developed a series of stable cell lines, but here I found that the involvement of different molecular parts of p300 is not as

easily discernible as it was in the case of Nut. All three classes of molecular elements of p300 (HAT, IDRs and TF-binding domains) are involved in condensate formation. This result – while more complex than I anticipated – was thought-provoking.

First, although removal of the IDRs from the Brd4-p300 fusion [Brd4-p300(ΔI)] does not result in decreased condensate formation compared to Brd4-p300(FL) (Fig.3.11), fusing these IDRs directly to Brd4 [Brd4-p300(I)] results in increased condensate formation compared to cells expressing Brd4 only (Fig.3.11, 3.12). Additionally, the condensates formed upon expression of Brd4-p300(ΔI) have a very different morphology from the ones formed by Brd4-p300(FL). These observations suggest that IDRs of p300 likely have an important function in of condensate formation. I hypothesize that the IDRs might provide a self-association modality, potentially contributing to a phase separation-based mechanism of condensate formation. In this way, some cells expressing a construct that only has IDRs of p300 fused to Brd4 can form condensates, but this effect is decreased as compared to Brd4-p300(FL)-expressing cells, due to the lack of HAT and TF-binding. At the same time, the condensates formed by Brd4-p300(FL) are likely more round in shape and potentially larger, thanks to the presence of these IDRs.

Removal of either HAT activity [Brd4-p300(H*IT)] or TF-binding domains [Brd4-p300(HI)] results in a decreased fraction of cells forming condensates. However, this effect is not additive, as a similar percentage of cells forms

condensates upon expression of Brd4-p300(I), which lacks both the HAT and TF-binding domains. This suggests that both HAT and TF-binding domains are necessary for a specific mechanism involved in condensate formation. Keeping in mind that a positive feedback loop mechanism has previously been proposed, I hypothesize that this might be a second element of p300-mediated condensate formation. Here, TF-binding domains of p300 could enable direct interaction with other elements of transcriptional machinery. Thanks to this interaction, two molecules of p300 could be brought into close proximity and trans-autoacetylated within their HAT domains to become active. Next, activated HAT domain could cause increased acetylation of nearby histone tails. Hyperacetylated histone tails will recruit more Brd4-Nut molecules, and through the interaction with Nut, more p300 molecules to these highly acetylated domains, finally resulting in a positive feedback loop. In summary, I expect that p300 plays a role in Brd4-Nut condensates formation through a dual mechanism:

1. Self-association-based mechanism of phase separation, through the IDRs of p300.
2. Positive feedback loop mechanism, through the activity of HAT and TF-binding domains of p300.

All these molecular elements of p300 are necessary to form large condensates and neither of these elements on its own can mimic the condensate formation to the same extent as the full-length fusion protein.

Previous studies of other fusion oncoproteins that can form condensates have also proposed a positive feedback loop as a potential mechanism of condensate formation. An example here is NONO-TFE3 (Wang, Gan, et al. 2021). The fusion oncoprotein forms condensates and the mechanism of this process involves one of the IDRs of NONO and the NRF1 transcription factor. However, in some cases, IDR-based multivalent interactions have been also reported to be involved in the condensate formation, including for example NUP98-HOXA9 fusion (Terlecki-Zaniewicz et al. 2021; Chandra et al. 2022; Shirnekhi, Chandra, and Kriwacki 2023). Here, the multivalent interactions involve a combination of homotypic and heterotypic interactions, but in all cases, the FG-rich IDR of NUP98 appears to play an important role. NUP98-HOXA9 condensates are formed both in living cells and *in vitro*, using a purified protein. When formed in cells, condensates associate with chromatin. Given these few examples already described in the literature, I hypothesize that in the case of Brd4-Nut, the condensate formation involves both these types of mechanisms: the positive feedback loop and IDR-mediated self-association, potentially leading to phase separation. To test these mechanisms, it would be advantageous to perform *in vitro* condensate formation assays and in this way, study the involvement of different molecular parts of p300.

Future directions

Research in recent years in the fields of biomolecular condensates, transcription and fusion oncoprotein-based cancer biology has vastly improved

our understanding of the potential mechanisms that underlie condensate formation and their transcriptional functions. Nevertheless, many important questions remain unanswered. Through my project, I have shown that addressing these questions will depend on a specific biomolecular condensate system. Here I would like to suggest the future directions of the Brd4-Nut system.

Firstly, we still do not understand many details of how Brd4-Nut condensates form. To describe the role of p300 in condensate formation better, additional mutant constructs will be needed, including a Brd4-p300 fusion with only the TF-binding domains and a Brd4-p300 fusion with only the HAT domain. While I have theorized on the potential roles that TF-binding and HAT activity in condensate formation, I do not have direct evidence of the contributions of either of these elements alone. Furthermore, it would be valuable to further dissect the different folded TF-binding domains and understand the importance of each of them in the context of condensate formation. For example, the bromodomain of p300 could contribute to acetylated histone tail binding and in this way increase the Brd4-Nut and Brd4-p300 binding to acetylated chromatin. Next, knowing that it is the TAZ2 domain of p300 that interacts directly with Nut, it would be beneficial to determine which residues exactly contribute to this interaction and if single point mutations in these residues could abrogate the Brd4-Nut condensate formation.

Moreover, what are the remaining roles of Nut in condensate formation and transcription? While a Brd4-Nut(ΔMIN) mutant is incapable of forming

condensates, the long, C-terminal disordered tail of Nut contains some intriguing sequence elements that could be worth studying in more detail. These include for example stretches of charged residues that are highly conserved between species. Could these highly charged residues provide a direct DNA-binding ability? Can the Nut IDR enable self-assembly? Additionally, what are the roles of the N-terminal elements of Nut that are predicted to be folded, but do not bind to p300? Do they recruit additional factors of transcriptional machinery? How does their presence or absence contribute to transcriptional outputs?

Next, what do different epigenetic marks contribute in the contexts of condensate formation and transcription? In my work, I analyzed the histone H3K27 acetyl mark, but there are multiple other acetylated histone marks as well as methylated marks in human chromatin. Some of the differentially expressed genes that I found in my transcriptional analyses encode epigenetic methyl- and acetyl- writers. It would be fascinating to learn more details about how interplay between methylation and acetylation marks contribute to both condensate formation and transcriptional changes in cells. What particular residues are modified on the histone tails upon expression of Brd4-Nut and condensate formation? Furthermore, how do different small molecule perturbations influence condensate formation and gene expression?

Additionally, tackling all these problems would be done best with the use of higher resolution techniques. These would include super-resolution

microscopy to study the formation of smaller size condensates as well as single cell RNAseq, to study cell-to-cell variability in transcriptional output, depending on the level of fusion protein expression. This would be a non-trivial undertaking to tackle, as it would require further technology development. As of now, it is possible to visualize active transcription through a couple of methods, including tagging native mRNA with uridine analog that can be visualized via click chemistry or my labeling specific components of transcriptional machinery. Some groups have also used different reporter systems to monitor transcription of a specific gene in single cells, such as LacI-LacO arrays. Nevertheless, it is not possible to tag nuclear proteins with antibodies in living cells and thus, it is also impossible to measure protein expression level and extract and sequence its mRNA within a span of minutes and thus, it is impossible at this point to analyze whole-cell transcriptome and combine it with analyses of protein expression. If technology like this is ever developed, this would be a very valuable, high-resolution approach that would allow to make direct connections between protein expression levels and transcriptional profiles of cells.

Finally, and perhaps most importantly, we still lack a translational assay connecting all the fusion constructs and condensate formation to tumorigenesis. It would be of great value to use all the stable cell lines I developed and monitor their effects in an animal model. This could allow us to connect condensate formation and transcription with disease development. While this might turn out to be a very challenging task, as the cells often do not survive prolonged fusion

protein expression, it would be worthwhile to determine what molecular parts of Brd4-Nut and p300 could contribute specifically to cell transformation.

Answering all these questions would increase our understanding of the molecular mechanisms of condensate formation and its potential detailed relationship with transcriptional changes. Additionally, it would be very important to determine how these mechanisms contribute to tumorigenesis. Nevertheless, tackling all these problems will require new technological developments and contributions from translational sciences, in addition to cell biology and biochemistry. Answering these questions promises to strengthen our understanding of the potential connections between pathological condensate formation and related transcription- and cancer-related functions.

APPENDIX A

MATERIALS AND METHODS

Cell Culture

HCC2429 cells were a kind gift from the Hamon Center for Therapeutic Oncology Research at UT Southwestern Medical Center. The cells were grown in RPMI 1640 medium (Thermo Fisher Scientific, #11875119) with addition of 10% heat-inactivated Fetal Bovine Serum (Thermo Fisher Scientific, #10-438-026) and 1% penicillin – streptomycin (Thermo Fisher Scientific, #15140-122). The cell line was validated via immunofluorescence microscopy and western blotting for BRD4-NUT expression. HCC2429 cells were expanded during an early passage after receiving the cell line and multiple aliquots were frozen for future use. The cell line was regularly tested negative for mycoplasma, using the MycoAlert™ Detection Kit (Lonza, #LT07-418) and MycoAlert™ Control Set (Lonza, #LT07-518).

293TRex-FlpIn cells were purchased from Thermo Fisher Scientific (#R78007). The cells were grown in DMEM medium (Thermo Fisher Scientific, #11965-118) with addition of 10% heat-inactivated Fetal Bovine Serum (Thermo Fisher Scientific, #10-438-026), 1% penicillin – streptomycin (Thermo Fisher Scientific, #15140-122), 15 µg/mL blasticidin (Thermo Fisher Scientific, #A1113903) and 100 µg/mL zeocin (Thermo Fisher Scientific, #R250-01). The cell line was regularly tested negative for mycoplasma, as described above.

293TRex-FlpIn cell lines inducibly expressing different constructs were developed based on the original 293TRex-FlpIn cell line, according to the manufacturer's instructions. The list of constructs successfully introduced into the cell line are: mNeonGreen-Brd4(short), "Brd4", mNeonGreen-Nut, "Nut", mNeonGreen-Brd4-Nut(FL), "Brd4-Nut(FL)", mNeonGreen-Brd4-Nut(355-505), "Brd4-Nut(MIN)", mNeonGreen-Brd4-Nut(Δ355-505), "Brd4-Nut(ΔMIN)", mNeonGreen-Brd4-p300(FL), "Brd4-p300(FL)", mNeonGreen-Brd4-p300(D1399Y), "Brd4-p300(IDR/TF)", mNeonGreen-Brd4-p300(ΔTF-binding), "Brd4-p300(IDR/H)", and mNeonGreen-Brd4-p300(IDR-only), "Brd4-p300(IDR)". I was not able to develop a stable cell line inducibly expressing mNeonGreen-Brd4-p300(ΔIDR) construct, I think due to high toxicity of the construct. Experiments that present data with use of mNeonGreen-Brd4-p300(ΔIDR) were performed using transient transfections. All cell lines were maintained in DMEM medium (Thermo Fisher Scientific, #11965-118) with addition of 10% heat-inactivated Fetal Bovine Serum (Thermo Fisher Scientific, #10-438-026), 1% penicillin – streptomycin (Thermo Fisher Scientific, #15140-122), 15 µg/mL blasticidin (Thermo Fisher Scientific, #A1113903) and 100 µg/mL hygromycin B (Thermo Fisher Scientific, #10687010). All the cell lines were expanded in an early passage after developing and frozen in multiple aliquots for future use. Each cell line was tested for protein expression after induction, via immunofluorescence microscopy and western blotting. All cell lines regularly tested negative for mycoplasma, as described above.

Immunofluorescence

Immunofluorescence was performed with HCC2429 cells, 293TRex-FlpIn cells and all newly developed cell lines described above, with Hoechst 33342 nuclear counterstaining. The primary antibodies used were: Nut monoclonal rabbit antibody (Cell Signaling Technology, #3625S), mNeonGreen monoclonal mouse antibody (Chromotek, #32f6-100), p300 monoclonal mouse antibody (Santa Cruz Biotechnology, #sc-56455), Brd4 monoclonal rabbit antibody (Abcam, #ab128874) and H3K27Ac monoclonal rabbit antibody (Abcam, #ab4729). Secondary antibodies were: Alexa Fluor 568 – conjugated goat anti-mouse secondary antibody (Thermo Fisher Scientific, #A-11004) and Alexa Fluor 647 – conjugated goat anti-rabbit secondary antibody (Thermo Fisher Scientific, #A21245).

24-well glass bottom microscopy plates (Cellvis, #P24-1.5H-N) were treated with 5 µg/mL poly-D-lysine (Sigma-Aldrich, #P7405-5MG) for 1h at room temperature, washed with 1xPBS (Thermo Fisher Scientific, #10010049) and dried for 2h at room temperature in the tissue culture hood.

Cells were seeded 1 day before induction, assuring optimal density for imaging. After induction with doxycycline for 2h, followed by a 4h washout (Sigma-Aldrich, #D9891-5G), cells were washed with 1 mL PBS per well and fixed with 0.5 mL 4% paraformaldehyde (Thermo Fisher Scientific, #RT15710) for 20 minutes at room temperature. Next, the cells were washed 3 times with 1 mL PBS per well for 3 minutes. Cells were then permeabilized with 1 mL per well of 0.5% Triton-X-100 (Thermo Fisher Scientific, #J66624AP) in PBS for 10 minutes at room temperature

and then washed 3 times with 1 mL PBS per well for 3 minutes. Next, the cells were blocked in the blocking buffer – 5% BSA (Thermo Fisher Scientific, #BP9704100) in PBST (PBS with 0.1% Tween-20 (Sigma-Aldrich, #P7949-500ML)) for 1 hour at room temperature. Primary antibodies were diluted in the blocking buffer as follows: Nut monoclonal rabbit antibody – 1:500, mNeonGreen monoclonal mouse antibody – 1:750, p300 monoclonal mouse antibody – 1:500, Brd4 monoclonal rabbit antibody – 1:500, H3K27Ac monoclonal rabbit antibody – 1:2000. The cells were incubated with 300µL of the primary antibody dilution overnight at 4^{o}C. The cells were then washed 3 times, for 5 minutes each time, with 1 mL of PBST per well. Appropriate secondary antibodies were diluted 1:1000 in the blocking buffer. The cells were incubated with 300µL of the secondary antibody dilution for 2 hours at room temperature, during which the plate was covered with aluminum foil to protect the samples from light. Maintaining the protection from light, the cells were next washed 3 times for 5 minutes with 1 mL of PBST per well. Next, the cells were incubated with 500µL of 1:2000 dilution of Hoechst 33342 (Thermo Fisher Scientific, #H3570) in PBS for 10 minutes at room temperature. Finally, the cells were washed 3 times for 3 minutes with 1 mL of PBS and then, 1 mL of fresh PBS was used as a mounting medium for each well. Finally, a 1:100 dilution of the secondary antibodies in PBS were prepared and 500µL of such dilutions were placed in separate empty wells, for flat-field correction imaging. One well was always prepared with PBS only, for background imaging. Plates prepared in this way were kept at 4^{o}C and covered with aluminum foil until performing microscopy.

Confocal Microscopy

All samples were imaged on a Nikon Eclipse Ti microscope base equipped with a Yokogawa CSUX1 spinning disk confocal scanner unit, using 100x / 1.49 NA oil objective and Andor EM-CCD camera. Images were acquired using MetaMorph software. A single Z-slice in the center of cell nucleus was acquired per image, with an average of 100-200 images acquired per sample. The lasers used were: 405nm, 488nm, 561nm and 637nm.

Statistical analyses

Welch's t-test without equal SDs assumption was used for pairwise comparisons in figures: 1d, 3c and d, 5c, 6d, 7c and d and S4b. Kruskall-Wallis nonparametric test without assuming equal distribution of residuals with Dunn's multiple comparisons of ranks between preselected pairs in figures: 4b and S6a. The nonparametric Kolmogorov-Smirnov test was used to compare the cumulative distributions in figure S4d.

Cell cross-linking for ChIP

Cells were grown in 15cm dishes to confluency and 2-3 plates were used for each experiment. After doxycycline treatments, cells were washed once with 20 mL of PBS. Next, the cells were treated with 1% methanol-free formaldehyde (Thermo Scientific, #28908) diluted in PBS – 10 mL per plate for 8 minutes, with gentle agitation. Formaldehyde was then quenched with 500µL of 2.5M glycine (to a final

concentration of 125mM) for another 8 minutes, with gentle agitation. After quenching, cells were scraped off the plates, transferred to 50 mL conical tubes and centrifuged at 500xg at 4ºC for 5 minutes. The supernatant was removed, and the cell pellet was washed with 5 mL of cold PBS per plate. Following the wash, cells were again centrifuged at 500xg at 4ºC for 5 minutes and then, the supernatant was removed. Pellets were finally snap-frozen in liquid nitrogen and stored at –80ºC.

Chromatin extraction and shearing

Cell pellets were thawed and resuspended in LB1 buffer (50mM Hepes-KOH pH 7.9, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP40, 1% TritonX-100 and 1x Complete, EDTA-free protease inhibitor cocktail, Sigma Aldrich, #11873580001) to obtain ~1x10^7 cells/mL suspension and they were then incubated on a rotator at 4ºC for 20 minutes, for lysis. Lysis efficiency of at least 60% was determined using trypan blue. Next, the suspension was centrifuged at 1,350xg at 4ºC for 5 minutes. The supernatant was removed, and the pellet was resuspended in LB2 buffer (10mM Tris pH 8.0, 200mM NaCl, 1mM EDTA, 0.5mM EGTA and 1x Complete, EDTA-free protease inhibitor cocktail) to again obtain ~1x10^7 cells/mL suspension and they were then incubated on a rotator at 4ºC for 5 minutes. Next, the suspension was centrifuged at 1,350xg at 4ºC for 5 minutes. The supernatant was removed, and the pellet was resuspended in LB3 buffer (10mM Tris pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% SDS, 1% TritonX-100 and 1x Complete, EDTA-free protease inhibitor cocktail) to obtain

~1x10^7 cells/mL suspension, which was then passed through a 27G needle 3 times, to completely homogenize the pellet. The suspension was transferred to a Covaris millitube with AFA fiber (Fisher Scientific, # NC0597431) and sonicated using Covaris M220 sonicator (average incident power 7.5 Watts, peak incident power 75 Watts, duty factor 10%, cycles: 200 count, duration: 18 minutes, minimum temperature: 5^{o}C, temperature setpoint: 7^{o}C, maximum temperature: 9^{o}C). After sonication, samples were centrifuged at 15,000xg at 4^{o}C for 10 minutes, soluble supernatant was transferred to a new tube, snap-frozen in liquid nitrogen and stored at –80^{o}C. 10µL of each sample were incubated overnight at 65^{o}C and treated with Proteinase K and RNAse A and DNA was purified to test the sonication efficiency (details described below, in “Chromatin Immunoprecipitation” part).

Chromatin immunoprecipitation

Protein G-conjugated Dynabeads (Thermo Fisher Scientific, #10004D) (75uL of suspension per IP) were washed with 1 mL of blocking buffer (0.5% BSA in PBS) 3 times and collected on a magnet stand each time. Beads were resuspended in 500µL blocking buffer and mixed with 7.5µg antibody (Nut monoclonal rabbit antibody (Cell Signaling Technology, #3625S), or H3K27Ac monoclonal rabbit antibody (Abcam, #ab4729)) and incubated on a rotator at 4^{o}C overnight. Next day, the beads were washed 3 times with the same blocking buffer to remove unbound antibody and then they were resuspended in 75µL of the blocking buffer. The resuspended beads bound to antibody were mixed with chromatin extract and mixed

on a rotator at 4ºC overnight. After that, the beads were washed 3 times in 1 mL washing buffer 1 (50mM Hepes pH 7.0, 100mM LiCl, 1mM EDTA, 1% NP40, 0.7% sodium deoxycholate and 1x Complete, EDTA-free protease inhibitor cocktail). For the third wash, beads were incubated on a rotator at 4ºC for 10 minutes. The washes were repeated with washing buffer 2 (20mM Tris pH 8.0, 350mM NaCl, 1% TritonX-100, 0.1% SDS, 2mM EDTA). Next, the beads were washed with 1 mL TE buffer with 50mM NaCl and centrifuged shortly at 300xg and then placed on the magnet stand, to remove residual buffer. Then, 200µL of the elution buffer was added to the beads (50mM Tris pH 8.0, 10mM EDTA and 1% SDS) and the beads were incubated at 65ºC for 30 min. with gentle agitation. The beads were next centrifuged at 300xg and placed on the magnet stand. The 200µL elution was then transferred to a new tube. The elution was incubated at 65ºC overnight to reverse crosslinks. Next, samples were treated with RNase A (Thermo Scientific, #EN0531) for 1h at 37ºC and then with Proteinase K (Fisher Scientific, #25-530-049) for 2h at 55ºC. DNA was then purified from these samples using the Qiagen PCR purification kit (Qiagen, #28106).

ChIP data processing and representation

Libraries were sequenced on Illumina NextSeq 500 at Next Generation Sequencing Core, Eugene McDermott Center for Human Growth and Development). For alignment, single-end reads with a read length of 100 bp were generated for each library. FASTQ files were checked for quality using fastqc (v0.11.5)[1] and

fastq_screen (v0.11.4)[2] .Low-quality reads and adapter were removed using fastq-mcf (v 1.05) http://expressionanalysis.github.io/ea-utils/). The reads from FASTQ files were aligned to the human genome (hg19) using bowtie2[3] (v2.3.3.1). Picard-tool's (v 2.10.10 https://broadinstitute.github.io/picard/) MarkDuplicates module was then used to remove duplicate alignments

For peak calling, annotation and motif analysis, the duplicate removed alignment files were used to call peaks using MACS2 (v2.1.0)[4], with a q-value threshold of 0.05 and using DNA input as background controls. The fragment size of each library was used to extend reads at their 3′ ends to a fixed length with "–extsize" parameter in MACS2. The peak files from the peak calls were annotated using annotatePeaks module in HOMER[5].

ChIPseq tracks shown were visualized using Integrative Genomics Viewer (Robinson et al. 2011; Thorvaldsdottir, Robinson, and Mesirov 2013). Venn diagrams for ChIPseq data were generated using BioVenn (Hulsen, de Vlieg, and Alkema 2008). Top ~1000 genes are shown in ChIPseq Venn diagrams per sample, based on p value (Wingett and Andrews 2018; Langmead and Salzberg 2012; Zhang et al. 2008; Heinz et al. 2010).

RNA sequencing sample preparation

Cells were grown in 10cm plates to confluency. After treatment with doxycycline, cells were lysed and RNA was purified using the Qiagen RNeasy kit (Qiagen, # NC9677589). Purified RNA was stored at –20^{o}C until use.

RNA sequencing data processing and representation

Samples were sequenced on the Illumina NextSeq 500 with read configuration as 75 bp, single end reads. (This part should be include in the RNA library preparation section towards the end).

The Fastq files were subjected to quality check using fastqc (version 0.11.5, http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and fastq_screen (version 0.11.4, http://www.bioinformatics.babraham.ac.uk/projects/ fastq_screen). The reads from FASTQ files were aligned to the human genome (hg19) using STAR[1] (v2.5.3a), a splice-aware aligner for RNA-seq data.

For differential gene expression of RNA inputs using edgeR, read counts mapping to genomic feature for each samples were generated using featureCounts[2] from the Rsubread package (v1.4.6). Next, read count matrix was used to run the differential expression analysis using edgeR[3]. edgeR uses TMM normalization and estimates the dispersion of the negative binomial distribution from replicates in each group. Furthermore, edgeR applies the Benjamini-Hochberg method on the p-values to control for FDR. Statistical cutoffs of *p-value* < 0.05 and $\log_2$FC > 2 were used to identify statistically significant differentially expressed genes (Dobin et al. 2013; Liao, Smyth, and Shi 2014; Robinson, McCarthy, and Smyth 2010).

Venn diagrams for RNAseq data were generated using BioVenn (Hulsen, de Vlieg, and Alkema 2008). Heatmaps for RNAseq-ChIPseq integration were generated using Heatmapper (Babicki et al. 2016).

Western blotting

For western blotting, cells were grown to confluency in 6-well plates or 10cm dishes induced with doxycycline. Cells were then washed once with PBS and lysed with RIPA buffer, supplemented with 400mM NaCl for 30 minutes. Clarified cell lysate was tested via BCA assay (Thermo Fisher Scientific, #PI23227) to quantify the protein concentration. After adjusting the concentrations of samples, 15µL of samples were mixed with 15µL of 2x SDS buffer and heated on a 100^{o}C heating block for 5-10 minutes. Next, the samples were loaded on a 10% SDS-PAGE gel made in-lab or on a 4-15% pre-cast TGX gel (Bio-Rad, #4568086), along with the molecular marker. Gels were run for 40 minutes - 1h at 200V. Transfer onto a PVDF membrane (Sigma Aldrich, #: IPVH00010) was performed wet, using a transfer buffer with addition of 0.1% SDS. After transfer, membrane was blocked for 1h at room temperature using blocking buffer (5% BSA in TBST). Next the membrane was incubated with the appropriate primary antibody in blocking buffer overnight at 4^{o}C. The following primary antibodies were used for western blotting: rabbit anti-Brd4(N) antibody (generous gift from the Chiang lab at UT Southwestern), rabbit anti-Nut (C52B1) antibody (Cell Signaling Technology, #3625S), mouse anti-p300 antibody (Millipore Sigma, #05-257), mouse anti-GAPDH (Thermo Fisher Scientific, #MA515738). Next day the membrane was washed 3 times, for 5 minutes each with TBST. Then the membrane was incubated with the appropriate secondary antibody in the blocking buffer at room temperature for 2h. The following secondary antibodies were used for western blotting: mouse anti-rabbit HRP antibody (Santa

Cruz Biotechnology, #sc-2357) and goat anti-mouse HRP antibody (Thermo Fisher Scientific, #31430). Next, the membrane was washed 2 times 10 minutes. in TBST and once for 10 minutes. in TBS. For signal development, the antibody was removed and membrane was incubated for about 3 minutes. in a solution with chemiluminescent substrate (Fisher Scientific, #WBKL S0 100).

mNeonGreen pulldown

To perform the mNeonGreen pulldowns, the cells were grown on a 10cm dish and treated with doxycycline for induction. Next, the cells were lysed with RIPA buffer supplemented with 400mM NaCl. For the pulldown, I purchased the mNeonGreen-Trap Agarose Kit (Proteinteck, #ntak) and followed the manufacturer's instructions. After the pulldown, the samples were ran on a gradient, 4-15% gel (Bio-Rad, #4568086). Transfer and western blotting was performed as described above, in the "western blotting" section.

Molecular biology and cloning

The following constructs were cloned into the pcDNA5/FRT/TO vector, purchased from Thermo Fisher Scientific (#V652020): mNeonGreen-BRD4(short), named "Brd4", mNeonGreen-NUT, named "Nut", mNeonGreen-Brd4-Nut(FL), named "Brd4-Nut(FL)", mNeonGreen-Brd4-Nut(355-505), named "Brd4-Nut(MIN)", mNeonGreen-Brd4-Nut(Δ355-505), named "Brd4-Nut(ΔMIN)", mNeonGreen-Brd4-p300(FL), named "Brd4-p300(FL)", mNeonGreen-Brd4-p300(D1399Y), named "Brd4-p300(IDR/TF)", mNeonGreen-Brd4-p300(ΔTF-binding), named "Brd4-

p300(IDR/H)", and mNeonGreen-Brd4-p300(IDR-only), named "Brd4-p300(IDR)" and mNeonGreen-Brd4-p300(ΔIDR). Some of the constructs were initially introduced into pInducer20 vector, and then moved into the pcDNA5/FRT/TO vector. The inserts were amplified using the following primers:

Insert name	Forward primer	Reverse primer	Template used
mNeonGreen-BRD4(short)	name: F-EcoRV-NLSmNG-pcDNA5, GCAGATATCCAGCACAGTGGCGGCCGATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-XhoI-BRD4short, CTAGACTCGAGCCTCACTATGATCAGGTCGAgccTGTTTCGGAGTCTTCGCTG	pIND20-mNeonGreen-BRD4(short)
mNeonGreen-BRD4-NUT(FL)	name: F-EcoRV-NLSmNG-pcDNA5, GCAGATATCCAGCACAGTGGCGGCCGATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-XhoI-NUT(FL)-pcDNA5, CCTCTAGACTCGAGCCTCACTATGATCAGGTCGAgccctggctacgacgtcg	pIND20-mNeonGreen-BRD4-NUT(FL)
mNeonGreen-BRD4-NUT(350-500)	name: F-EcoRV-NLSmNG-pcDNA5, GCAGATATCCAGCACAGTGGCGGCCGATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-XhoI-NUT(1-500)-pcDNA5, CCTCTAGACTCGAGCCTCACTATGATCAGGTCGAgccATCCCCATCTTCATCC	pIND20-mNeonGreen-BRD4-NUT(350-500)
mNeonGreen-NUT	name: F-EcoRV-NLSmNG-pcDNA5, GCAGATATCCAGCACAGTGGCGGCCGATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-XhoI-NUT(FL)-pcDNA5, CCTCTAGACTCGAGCCTCACTATGATCAGGTCGAgccctggctacgacgtcg	pIND20-mNeonGreen-NUT(FL)
mNeonGreen-BRD4-p300(IDR)	name: F-HindIII-NLS, CGTTTAAACTTAAGCTTATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-XhoI-NotI-p300(FL), CTCGAGCGGCCGCTCATATGTCTAGTGTACTCTGTGAGAGGTTTGAATTCAAGTCTG	pIND20-mNeonGreen-BRD4-p300(N1+N2+PolyQ)
mNeonGreen-BRD4-p300(FL)	name: F-KpnI-Brd4, CGGTACCAAACACAACTCAAGCATCGACTCCTCCGC	name: R-XhoI-BRD4short, CTAGACTCGAGCCTCACTATGATCAGGTCGAgccTGTTTCGGAGTCTTC	pIND20-mNeonGreen-BRD4(short)

		GCTG	
mNeonGreen-BRD4-p300(D1399Y)	name: F-KpnI-Brd4, CGGTACCAAACACAACTCAAGCATC GACTCCTCCGC	name: R-XhoI-NUT(FL)-pcDNA5, CCTCTAGACTCGAGCCTCACTATGATCA GGTCGAgccctggctacgacgtcg	pIND20-mNeonGreen-BRD4-NUT(FL)
mNeonGreen-BRD4-p300(IDR/H)	name: F-KpnI-Brd4, CGGTACCAAACACAACTCAAGCATC GACTCCTCCGC	name: R-XhoI-NUT(1-500)-pcDNA5, CCTCTAGACTCGAGCCTCACTATGATCA GGTCGAgccATCCCCATCTTCATCC	pIND20-mNeonGreen-BRD4-NUT(1-500)

Amplification was performed using Pfu Turbo polymerase (Agilent, #600252), according to the manufacturer's instructions. After successful amplification, DNA was PCR-purified using Qiagen PCR purification kit (Qiagen, #28106). The amplified fragments and the pcDNA5/FRT/TO vector were then digested with the enzymes listed in the names of primers in the table above. Digestion was performed at 37ºC for 1-4h, using Cut Smart buffer from NEB, supplied along with the enzymes. After that time, Quick CIP phosphatase (NEB, # M0525S) was added to the vector and the mix was incubated at 37ºC for additional 10-15 minutes. The fragments were then ligated into the vector using T4ligase (Enzymatics, # L6030-LC-L) and transformed into Max Efficiency Stbl2 Competent Cells (Thermo Fisher Scientific, #10268019). The following constructs were obtained using exactly this method:

- mNeonGreen-BRD4(short)
- mNeonGreen-BRD4-NUT(FL)
- mNeonGreen-BRD4-NUT(350-500)
- mNeonGreen-NUT

- mNeonGreen-BRD4-p300(IDR)

For the remaining constructs, cloning was performed in two rounds, because the inserts to be ligated into the vector were larger than the vector itself. First, a 1366bp fragment was ligated into the vector, using the following primers:

Insert name	Forward primer	Reverse primer	Template used
HindIII - NLS - mNeonGreen - Brd4 - KpnI	name: F-HindIII-NLS, CGTTTAAACTTAAGCTTATGGGACCGGCTGCCAAGCGCGTGAAACTCG	name: R-KpnI-BRD4, GGTACCGTGGAAACGCCAGGTTTTGCTGTCCC	pIND20-mNeonGreen-BRD4(short)

The amplified fragment and the pcDNA5/FRT/TO vector were both digested using HindIII and KpnI enzymes, using the method described above. After obtaining and sequencing the new vector with the 1366bp insert, the remaining inserts were amplified, using the primers listed in the first table:

- mNeonGreen-BRD4-p300(FL)
- mNeonGreen-BRD4-p300(D1399Y)
- mNeonGreen-BRD4-p300(IDR/H)

The digest of the new vector (with 1366bp fragment insert) and digest of the new inserts was performed using the specified enzymes, in the method described above. The fragments were ligated with T4 ligase and DNA was transformed into the Stbl2 bacteria. After DNA purification all inserts were fully sequenced.

Finally, the mNeonGreen-Brd4-Nut(ΔMIN) construct was obtained using Gibson assembly protocol, with the following primers:

Insert name	Forward primer	Reverse primer	Template used
mNeonGreen-Brd4-	name: F-d350-500_Gibson341,	name: R-d350-500_Gibson341,	mNeonGreen-Brd4-Nut(FL)

Nut(ΔMIN)	GGCGTCAGCGTAAAGCCCAGAGACCTGGGCGGCTTCGGCCCTCACCTGGGCTTCAGGGG	CCCCTGAAGCCCAGGTGAGGGCCGAAGCCGCCCAGGTCTCTGGGCTTTACGCTGACGCC	
mNeonGreen-Brd4-Nut(ΔMIN)	Name: F-d350-500_Gibson556, GGAGAGGGACGATGTCTGTCTCAGCCCAGGAGTTTGGCTGAGCAGTGAGATGGATGC	R-d350-500_Gibson556, GCATCCATCTCACTGCTCAGCCAAACTCCTGGGCTGAGACAGACATCGTCCCTCTCC	mNeonGreen-Brd4-Nut(FL)

After DNA purification the construct was sequenced.

.

BIBLIOGRAPHY

Ahmed, N. S., L. M. Harrell, D. R. Wieland, M. A. Lay, V. F. Thompson, and J. C. Schwartz. 2021. 'Fusion protein EWS-FLI1 is incorporated into a protein granule in cells', *RNA*, 27: 920-32.

Ahn, J. H., E. S. Davis, T. A. Daugird, S. Zhao, I. Y. Quiroga, H. Uryu, J. Li, A. J. Storey, Y. H. Tsai, D. P. Keeley, S. G. Mackintosh, R. D. Edmondson, S. D. Byrum, L. Cai, A. J. Tackett, D. Zheng, W. R. Legant, D. H. Phanstiel, and G. G. Wang. 2021. 'Phase separation drives aberrant chromatin looping and cancer development', *Nature*, 595: 591-95.

Alberti, S., and D. Dormann. 2019. 'Liquid-Liquid Phase Separation in Disease', *Annu Rev Genet*, 53: 171-94.

Alberti, S., S. Saha, J. B. Woodruff, T. M. Franzmann, J. Wang, and A. A. Hyman. 2018. 'A User's Guide for Phase Separation Assays with Purified Proteins', *J Mol Biol*, 430: 4806-20.

Alekseyenko, A. A., E. M. Walsh, X. Wang, A. R. Grayson, P. T. Hsi, P. V. Kharchenko, M. I. Kuroda, and C. A. French. 2015. 'The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains', *Genes Dev*, 29: 1507-23.

Alekseyenko, A. A., E. M. Walsh, B. M. Zee, T. Pakozdi, P. Hsi, M. E. Lemieux, P. Dal Cin, T. A. Ince, P. V. Kharchenko, M. I. Kuroda, and C. A. French. 2017. 'Ectopic protein interactions within BRD4-chromatin complexes drive oncogenic megadomain formation in NUT midline carcinoma', *Proc Natl Acad Sci U S A*, 114: E4184-E92.

Anderson, P., and N. Kedersha. 2006. 'RNA granules', *J Cell Biol*, 172: 803-8.

Arakawa, H. 2004. 'Netrin-1 and its receptors in tumorigenesis', *Nat Rev Cancer*, 4: 978-87.

Babicki, S., D. Arndt, A. Marcu, Y. Liang, J. R. Grant, A. Maciejewski, and D. S. Wishart. 2016. 'Heatmapper: web-enabled heat mapping for all', *Nucleic Acids Res*, 44: W147-53.

Banani, S. F., H. O. Lee, A. A. Hyman, and M. K. Rosen. 2017. 'Biomolecular condensates: organizers of cellular biochemistry', *Nat Rev Mol Cell Biol*, 18: 285-98.

Banani, S. F., A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, and M. K. Rosen. 2016. 'Compositional Control of Phase-Separated Cellular Bodies', *Cell*, 166: 651-63.

Banjade, S., and M. K. Rosen. 2014. 'Phase transitions of multivalent proteins can promote clustering of membrane receptors', *Elife*, 3.

Banjade, S., Q. Wu, A. Mittal, W. B. Peeples, R. V. Pappu, and M. K. Rosen. 2015. 'Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck', *Proc Natl Acad Sci U S A*, 112: E6426-35.

Baumann, K. 2022. 'Phase separation in stress resistance', *Nat Rev Mol Cell Biol*, 23: 229.

Boija, A., I. A. Klein, B. R. Sabari, A. Dall'Agnese, E. L. Coffey, A. V. Zamudio, C. H. Li, K. Shrinivas, J. C. Manteiga, N. M. Hannett, B. J. Abraham, L. K. Afeyan, Y. E. Guo, J. K. Rimel, C. B. Fant, J. Schuijers, T. I. Lee, D. J. Taatjes, and R. A. Young. 2018. 'Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains', *Cell*, 175: 1842-55 e16.

Boulay, G., G. J. Sandoval, N. Riggi, S. Iyer, R. Buisson, B. Naigles, M. E. Awad, S. Rengarajan, A. Volorio, M. J. McBride, L. C. Broye, L. Zou, I. Stamenkovic, C. Kadoch, and M. N. Rivera. 2017. 'Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain', *Cell*, 171: 163-78 e19.

Boyson, S. P., C. Gao, K. Quinn, J. Boyd, H. Paculova, S. Frietze, and K. C. Glass. 2021. 'Functional Roles of Bromodomain Proteins in Cancer', *Cancers (Basel)*, 13.

Brangwynne, C. P., T. J. Mitchison, and A. A. Hyman. 2011. 'Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes', *Proc Natl Acad Sci U S A*, 108: 4334-9.

Brodsky, S., T. Jana, K. Mittelman, M. Chapal, D. K. Kumar, M. Carmi, and N. Barkai. 2020. 'Intrinsically Disordered Regions Direct Transcription Factor In Vivo Binding Specificity', *Mol Cell*, 79: 459-71 e4.

Burke, K. A., A. M. Janke, C. L. Rhine, and N. L. Fawzi. 2015. 'Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II', *Mol Cell*, 60: 231-41.

Cabral, S. E., J. P. Otis, and K. L. Mowry. 2022. 'Multivalent interactions with RNA drive recruitment and dynamics in biomolecular condensates in Xenopus oocytes', *iScience*, 25: 104811.

Carracedo, A., K. Ito, and P. P. Pandolfi. 2011. 'The nuclear bodies inside out: PML conquers the cytoplasm', *Curr Opin Cell Biol*, 23: 360-6.

Case, L. B., M. De Pasquale, L. Henry, and M. K. Rosen. 2022. 'Synergistic phase separation of two pathways promotes integrin clustering and nascent adhesion formation', *Elife*, 11.

Case, L. B., X. Zhang, J. A. Ditlev, and M. K. Rosen. 2019. 'Stoichiometry controls activity of phase-separated clusters of actin signaling proteins', *Science*, 363: 1093-97.

Cavalieri, S., A. Stathis, A. Fabbri, A. Sonzogni, F. Perrone, E. Tamborini, G. Pelosi, F. de Braud, and M. Platania. 2017. 'Uncommon somatic mutations in metastatic NUT midline carcinoma', *Tumori*, 103: e5-e8.

Chandra, B., N. L. Michmerhuizen, H. K. Shirnekhi, S. Tripathi, B. J. Pioso, D. W. Baggett, D. M. Mitrea, I. Iacobucci, M. R. White, J. Chen, C. G. Park, H. Wu, S. Pounds, A. Medyukhina, K. Khairy, Q. Gao, C. Qu, S.

Abdelhamed, S. D. Gorman, S. Bawa, C. Maslanka, S. Kinger, P. Dogra, M. C. Ferrolino, D. Di Giacomo, C. Mecucci, J. M. Klco, C. G. Mullighan, and R. W. Kriwacki. 2022. 'Phase Separation Mediates NUP98 Fusion Oncoprotein Leukemic Transformation', *Cancer Discov*, 12: 1152-69.

Cho, W. K., J. H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, and Cisse, II. 2018. 'Mediator and RNA polymerase II clusters associate in transcription-dependent condensates', *Science*, 361: 412-15.

Chong, S., T. G. W. Graham, C. Dugast-Darzacq, G. M. Dailey, X. Darzacq, and R. Tjian. 2022. 'Tuning levels of low-complexity domain interactions to modulate endogenous oncogenic transcription', *Mol Cell*, 82: 2084-97 e5.

Cole, R. W., T. Jinadasa, and C. M. Brown. 2011. 'Measuring and interpreting point spread functions to determine confocal microscope resolution and ensure quality control', *Nat Protoc*, 6: 1929-41.

Collier, N. C., and M. J. Schlesinger. 1986. 'The dynamic state of heat shock proteins in chicken embryo fibroblasts', *J Cell Biol*, 103: 1495-507.

Davis, R. B., T. Kaur, M. M. Moosa, and P. R. Banerjee. 2021. 'FUS oncofusion protein condensates recruit mSWI/SNF chromatin remodeler via heterotypic interactions between prion-like domains', *Protein Sci*, 30: 1454-66.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras. 2013. 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29: 15-21.

Donati, B., E. Lorenzini, and A. Ciarrocchi. 2018. 'BRD4 and Cancer: going beyond transcriptional regulation', *Mol Cancer*, 17: 164.

Duster, R., I. H. Kaltheuner, M. Schmitz, and M. Geyer. 2021. '1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities', *J Biol Chem*, 296: 100260.

Eagen, K. P., and C. A. French. 2021. 'Supercharging BRD4 with NUT in carcinoma', *Oncogene*, 40: 1396-408.

Erdos, G., and Z. Dosztanyi. 2020. 'Analyzing Protein Disorder with IUPred2A', *Curr Protoc Bioinformatics*, 70: e99.

Filippakopoulos, P., J. Qi, S. Picaud, Y. Shen, W. B. Smith, O. Fedorov, E. M. Morse, T. Keates, T. T. Hickman, I. Felletar, M. Philpott, S. Munro, M. R. McKeown, Y. Wang, A. L. Christie, N. West, M. J. Cameron, B. Schwartz, T. D. Heightman, N. La Thangue, C. A. French, O. Wiest, A. L. Kung, S. Knapp, and J. E. Bradner. 2010. 'Selective inhibition of BET bromodomains', *Nature*, 468: 1067-73.

Flory, P. J. 1941. 'Thermodynamics of High Polymer Solutions', *Journal of Chemical Physics*, 9: 660.

Forman-Kay, J. D., J. A. Ditlev, M. L. Nosella, and H. O. Lee. 2022. 'What are the distinguishing features and size requirements of biomolecular condensates and their implications for RNA-containing condensates?', *RNA*, 28: 36-47.

Fox, A. H., and A. I. Lamond. 2010. 'Paraspeckles', *Cold Spring Harb Perspect Biol*, 2: a000687.

French, C. A., I. Miyoshi, I. Kubonishi, H. E. Grier, A. R. Perez-Atayde, and J. A. Fletcher. 2003. 'BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma', *Cancer Res*, 63: 304-7.

French, C. A., S. Rahman, E. M. Walsh, S. Kuhnle, A. R. Grayson, M. E. Lemieux, N. Grunfeld, B. P. Rubin, C. R. Antonescu, S. Zhang, R. Venkatramani, P. Dal Cin, and P. M. Howley. 2014. 'NSD3-NUT fusion oncoprotein in NUT midline carcinoma: implications for a novel oncogenic mechanism', *Cancer Discov*, 4: 928-41.

Gibson, B. A., Blaukopf, C., Lou, T., Doolittle, L. K., Finkelstein, I., Narlikar, G. J., Gerlich, D. W., Rosen, M. K. . 2021. 'In Diverse Conditions Intrinsic Chromatin Condensates Have Liquid-like Material Properties'.

Gibson, B. A., L. K. Doolittle, M. W. G. Schneider, L. E. Jensen, N. Gamarra, L. Henry, D. W. Gerlich, S. Redding, and M. K. Rosen. 2019. 'Organization of Chromatin by Intrinsic and Regulated Phase Separation', *Cell*, 179: 470-84 e21.

Guo, Y. E., J. C. Manteiga, J. E. Henninger, B. R. Sabari, A. Dall'Agnese, N. M. Hannett, J. H. Spille, L. K. Afeyan, A. V. Zamudio, K. Shrinivas, B. J. Abraham, A. Boija, T. M. Decker, J. K. Rimel, C. B. Fant, T. I. Lee, Cisse, II, P. A. Sharp, D. J. Taatjes, and R. A. Young. 2019. 'Pol II phosphorylation regulates a switch between transcriptional and splicing condensates', *Nature*, 572: 543-48.

Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, and C. K. Glass. 2010. 'Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities', *Mol Cell*, 38: 576-89.

Henninger, J. E., O. Oksuz, K. Shrinivas, I. Sagi, G. LeRoy, M. M. Zheng, J. O. Andrews, A. V. Zamudio, C. Lazaris, N. M. Hannett, T. I. Lee, P. A. Sharp, Cisse, II, A. K. Chakraborty, and R. A. Young. 2021. 'RNA-Mediated Feedback Control of Transcriptional Condensates', *Cell*, 184: 207-25 e24.

Heublein, S., D. Mayr, A. Meindl, A. Kircher, U. Jeschke, and N. Ditsch. 2017. 'Vitamin D receptor, Retinoid X receptor and peroxisome proliferator-activated receptor gamma are overexpressed in BRCA1 mutated breast cancer and predict prognosis', *J Exp Clin Cancer Res*, 36: 57.

Hnisz, D., K. Shrinivas, R. A. Young, A. K. Chakraborty, and P. A. Sharp. 2017. 'A Phase Separation Model for Transcriptional Control', *Cell*, 169: 13-23.

Holehouse, A. S., R. K. Das, J. N. Ahad, M. O. Richardson, and R. V. Pappu. 2017. 'CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins', *Biophys J*, 112: 16-21.

Huang, Q. W., L. J. He, S. Zheng, T. Liu, and B. N. Peng. 2019. 'An Overview of Molecular Mechanism, Clinicopathological Factors, and Treatment in NUT Carcinoma', *Biomed Res Int*, 2019: 1018439.

Huggins, M. L. 1941. 'Solutions of Long Chain Compounds', *Journal of Chemical Physics*, 9: 440.

Hulsen, T., J. de Vlieg, and W. Alkema. 2008. 'BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams', *BMC Genomics*, 9: 488.

Hyman, A. A., C. A. Weber, and F. Julicher. 2014. 'Liquid-liquid phase separation in biology', *Annu Rev Cell Dev Biol*, 30: 39-58.

Ibrahim, Z., T. Wang, O. Destaing, N. Salvi, N. Hoghoughi, C. Chabert, A. Rusu, J. Gao, L. Feletto, N. Reynoird, T. Schalch, Y. Zhao, M. Blackledge, S. Khochbin, and D. Panne. 2022. 'Structural insights into p300 regulation and acetylation-dependent genome organisation', *Nat Commun*, 13: 7759.

Ito, A., C. H. Lai, X. Zhao, S. Saito, M. H. Hamilton, E. Appella, and T. P. Yao. 2001. 'p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2', *EMBO J*, 20: 1331-40.

Jain, A., and R. D. Vale. 2017. 'RNA phase transitions in repeat expansion disorders', *Nature*, 546: 243-47.

Jung, M., S. Kim, J. K. Lee, S. O. Yoon, H. S. Park, S. W. Hong, W. S. Park, J. E. Kim, J. Kim, B. Keam, H. J. Kim, H. J. Kang, D. W. Kim, K. C. Jung, Y. T. Kim, D. S. Heo, T. M. Kim, and Y. K. Jeon. 2019. 'Clinicopathological and Preclinical Findings of NUT Carcinoma: A Multicenter Study', *Oncologist*, 24: e740-e48.

Kato, M., T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko, and S. L. McKnight. 2012. 'Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels', *Cell*, 149: 753-67.

Lafarga, M., I. Casafont, R. Bengoechea, O. Tapia, and M. T. Berciano. 2009. 'Cajal's contribution to the knowledge of the neuronal cell nucleus', *Chromosoma*, 118: 437-43.

Langmead, B., and S. L. Salzberg. 2012. 'Fast gapped-read alignment with Bowtie 2', *Nat Methods*, 9: 357-9.

Lee, J. K., S. Louzada, Y. An, S. Y. Kim, S. Kim, J. Youk, S. Park, S. H. Koo, B. Keam, Y. K. Jeon, J. L. Ku, F. Yang, T. M. Kim, and Y. S. Ju. 2017. 'Complex chromosomal rearrangements by single catastrophic pathogenesis in NUT midline carcinoma', *Ann Oncol*, 28: 890-97.

Levy, D. E., and J. E. Darnell, Jr. 2002. 'Stats: transcriptional control and biological impact', *Nat Rev Mol Cell Biol*, 3: 651-62.

Li, P., S. Banjade, H. C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q. X. Jiang, B. T. Nixon, and M. K. Rosen. 2012. 'Phase transitions in the assembly of multivalent signalling proteins', *Nature*, 483: 336-40.

Liao, Y., G. K. Smyth, and W. Shi. 2014. 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30: 923-30.

Lin, Y., S. L. Currie, and M. K. Rosen. 2017. 'Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs', *J Biol Chem*, 292: 19110-20.

Liu, S., and G. Ferzli. 2018. 'NUT carcinoma: a rare and devastating neoplasm', *BMJ Case Rep*, 2018.

Lu, H., D. Yu, A. S. Hansen, S. Ganguly, R. Liu, A. Heckert, X. Darzacq, and Q. Zhou. 2018. 'Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II', *Nature*, 558: 318-23.

Lyons, H., R. T. Veettil, P. Pradhan, C. Fornero, N. De La Cruz, K. Ito, M. Eppert, R. G. Roeder, and B. R. Sabari. 2023. 'Functional partitioning of

transcriptional regulators by patterned charge blocks', *Cell*, 186: 327-45 e28.

Marazzi, I., B. D. Greenbaum, D. H. P. Low, and E. Guccione. 2018. 'Chromatin dependencies in cancer and inflammation', *Nat Rev Mol Cell Biol*, 19: 245-61.

McSwiggen, D. T., M. Mir, X. Darzacq, and R. Tjian. 2019. 'Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences', *Genes Dev*, 33: 1619-34.

Meszaros, B., G. Erdos, and Z. Dosztanyi. 2018. 'IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding', *Nucleic Acids Res*, 46: W329-W37.

Mondal, S., K. Narayan, S. Botterbusch, I. Powers, J. Zheng, H. P. James, R. Jin, and T. Baumgart. 2022. 'Multivalent interactions between molecular components involved in fast endophilin mediated endocytosis drive protein phase separation', *Nat Commun*, 13: 5017.

Moreno, V., K. Saluja, and S. Pina-Oviedo. 2022. 'NUT Carcinoma: Clinicopathologic Features, Molecular Genetics and Epigenetics', *Front Oncol*, 12: 860830.

Morrison-Smith, C. D., T. M. Knox, I. Filic, K. M. Soroko, B. K. Eschle, M. K. Wilkens, P. C. Gokhale, F. Giles, A. Griffin, B. Brown, G. I. Shapiro, B. E. Zucconi, P. A. Cole, M. E. Lemieux, and C. A. French. 2020. 'Combined Targeting of the BRD4-NUT-p300 Axis in NUT Midline Carcinoma by Dual Selective Bromodomain Inhibitor, NEO2734', *Mol Cancer Ther*, 19: 1406-14.

Musacchio, A. 2022. 'On the role of phase separation in the biogenesis of membraneless compartments', *EMBO J*, 41: e109952.

Nair, S. J., L. Yang, D. Meluzzi, S. Oh, F. Yang, M. J. Friedman, S. Wang, T. Suter, I. Alshareedah, A. Gamliel, Q. Ma, J. Zhang, Y. Hu, Y. Tan, K. A. Ohgi, R. S. Jayani, P. R. Banerjee, A. K. Aggarwal, and M. G. Rosenfeld.

2019. 'Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly', *Nat Struct Mol Biol*, 26: 193-203.

Nott, T. J., E. Petsalaki, P. Farber, D. Jervis, E. Fussner, A. Plochowietz, T. D. Craggs, D. P. Bazett-Jones, T. Pawson, J. D. Forman-Kay, and A. J. Baldwin. 2015. 'Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles', *Mol Cell*, 57: 936-47.

Ortega, E., S. Rengachari, Z. Ibrahim, N. Hoghoughi, J. Gaucher, A. S. Holehouse, S. Khochbin, and D. Panne. 2018. 'Transcription factor dimerization activates the p300 acetyltransferase', *Nature*, 562: 538-44.

Owen, I., D. Yee, H. Wyne, T. M. Perdikari, V. Johnson, J. Smyth, R. Kortum, N. L. Fawzi, and F. Shewmaker. 2021. 'The oncogenic transcription factor FUS-CHOP can undergo nuclear liquid-liquid phase separation', *J Cell Sci*, 134.

Pak, C. W., M. Kosno, A. S. Holehouse, S. B. Padrick, A. Mittal, R. Ali, A. A. Yunus, D. R. Liu, R. V. Pappu, and M. K. Rosen. 2016. 'Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein', *Mol Cell*, 63: 72-85.

Parker, R., and U. Sheth. 2007. 'P bodies and the control of mRNA translation and degradation', *Mol Cell*, 25: 635-46.

Pavlaki, I. 2021. 'Fusion proteins drive cancer through phase separation', *Nat Cancer*, 2: 1285.

Piha-Paul, S. A., C. L. Hann, C. A. French, S. Cousin, I. Brana, P. A. Cassier, V. Moreno, J. S. de Bono, S. D. Harward, G. Ferron-Brady, O. Barbash, A. Wyce, Y. Wu, T. Horner, M. Annan, N. J. Parr, R. K. Prinjha, C. L. Carpenter, J. Hilton, D. S. Hong, N. B. Haas, M. C. Markowski, A. Dhar, P. J. O'Dwyer, and G. I. Shapiro. 2020. 'Phase 1 Study of Molibresib (GSK525762), a Bromodomain and Extra-Terminal Domain Protein

Inhibitor, in NUT Carcinoma and Other Solid Tumors', *JNCI Cancer Spectr*, 4: pkz093.

Qin, Z., H. Sun, M. Yue, X. Pan, L. Chen, X. Feng, X. Yan, X. Zhu, and H. Ji. 2021. 'Phase separation of EML4-ALK in firing downstream signaling and promoting lung tumorigenesis', *Cell Discov*, 7: 33.

Quiroga, I. Y., J. H. Ahn, G. G. Wang, and D. Phanstiel. 2022. 'Oncogenic fusion proteins and their role in three-dimensional chromatin structure, phase separation, and cancer', *Curr Opin Genet Dev*, 74: 101901.

Rahman, S., M. E. Sowa, M. Ottinger, J. A. Smith, Y. Shi, J. W. Harper, and P. M. Howley. 2011. 'The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3', *Mol Cell Biol*, 31: 2641-52.

Reynoird, N., B. E. Schwartz, M. Delvecchio, K. Sadoul, D. Meyers, C. Mukherjee, C. Caron, H. Kimura, S. Rousseaux, P. A. Cole, D. Panne, C. A. French, and S. Khochbin. 2010. 'Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains', *EMBO J*, 29: 2943-52.

Riback, J. A., L. Zhu, M. C. Ferrolino, M. Tolbert, D. M. Mitrea, D. W. Sanders, M. T. Wei, R. W. Kriwacki, and C. P. Brangwynne. 2020. 'Composition-dependent thermodynamics of intracellular phase separation', *Nature*, 581: 209-14.

Ritter, R. 1890. 'Die Entwicklung der Geschlechtsorgane und des Darmes bei Chiromomus', *Zeit furr Wiss Zool Bd*, 50: 408-27.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. 'Integrative genomics viewer', *Nat Biotechnol*, 29: 24-6.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth. 2010. 'edgeR: a Bioconductor package for differential expression analysis of digital gene expression data', *Bioinformatics*, 26: 139-40.

Rosencrance, C. D., H. N. Ammouri, Q. Yu, T. Ge, E. J. Rendleman, S. A. Marshall, and K. P. Eagen. 2020. 'Chromatin Hyperacetylation Impacts Chromosome Folding by Forming a Nuclear Subcompartment', *Mol Cell*, 78: 112-26 e12.

Rousseaux, S., N. Reynoird, and S. Khochbin. 2022. 'NUT Is a Driver of p300-Mediated Histone Hyperacetylation: From Spermatogenesis to Cancer', *Cancers (Basel)*, 14.

Ruff, K. M., F. Dar, and R. V. Pappu. 2021. 'Ligand effects on phase separation of multivalent macromolecules', *Proc Natl Acad Sci U S A*, 118.

Sabari, B. R., A. Dall'Agnese, A. Boija, I. A. Klein, E. L. Coffey, K. Shrinivas, B. J. Abraham, N. M. Hannett, A. V. Zamudio, J. C. Manteiga, C. H. Li, Y. E. Guo, D. S. Day, J. Schuijers, E. Vasile, S. Malik, D. Hnisz, T. I. Lee, Cisse, II, R. G. Roeder, P. A. Sharp, A. K. Chakraborty, and R. A. Young. 2018. 'Coactivator condensation at super-enhancers links phase separation and gene control', *Science*, 361.

Sampson, J., M. W. Richards, J. Choi, A. M. Fry, and R. Bayliss. 2021. 'Phase-separated foci of EML4-ALK facilitate signalling and depend upon an active kinase conformation', *EMBO Rep*, 22: e53693.

Schwartz, B. E., M. D. Hofer, M. E. Lemieux, D. E. Bauer, M. J. Cameron, N. H. West, E. S. Agoston, N. Reynoird, S. Khochbin, T. A. Ince, A. Christie, K. A. Janeway, S. O. Vargas, A. R. Perez-Atayde, J. C. Aster, S. E. Sallan, A. L. Kung, J. E. Bradner, and C. A. French. 2011. 'Differentiation of NUT midline carcinoma by epigenomic reprogramming', *Cancer Res*, 71: 2686-96.

Shin, Y., J. Berry, N. Pannucci, M. P. Haataja, J. E. Toettcher, and C. P. Brangwynne. 2017. 'Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets', *Cell*, 168: 159-71 e14.

Shiota, H., S. Barral, T. Buchou, M. Tan, Y. Coute, G. Charbonnier, N. Reynoird, F. Boussouar, M. Gerard, M. Zhu, L. Bargier, D. Puthier, F. Chuffart, E.

Bourova-Flin, S. Picaud, P. Filippakopoulos, A. Goudarzi, Z. Ibrahim, D. Panne, S. Rousseaux, Y. Zhao, and S. Khochbin. 2018. 'Nut Directs p300-Dependent, Genome-Wide H4 Hyperacetylation in Male Germ Cells', *Cell Rep*, 24: 3477-87 e6.

Shiota, H., J. E. Elya, A. A. Alekseyenko, P. M. Chou, S. A. Gorman, O. Barbash, K. Becht, K. Danga, M. I. Kuroda, V. Nardi, and C. A. French. 2018. '"Z4" Complex Member Fusions in NUT Carcinoma: Implications for a Novel Oncogenic Mechanism', *Mol Cancer Res*, 16: 1826-33.

Shirnekhi, H. K., B. Chandra, and R. W. Kriwacki. 2023. 'The Role of Phase-Separated Condensates in Fusion Oncoprotein–Driven Cancers', *Annual Review of Cancer Biology*, 7.

Shrinivas, K., B. R. Sabari, E. L. Coffey, I. A. Klein, A. Boija, A. V. Zamudio, J. Schuijers, N. M. Hannett, P. A. Sharp, R. A. Young, and A. K. Chakraborty. 2019. 'Enhancer Features that Drive Formation of Transcriptional Condensates', *Mol Cell*, 75: 549-61 e7.

Stirling, D. R., M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, and A. Goodman. 2021. 'CellProfiler 4: improvements in speed, utility and usability', *BMC Bioinformatics*, 22: 433.

Stringer, C., T. Wang, M. Michaelos, and M. Pachitariu. 2021. 'Cellpose: a generalist algorithm for cellular segmentation', *Nat Methods*, 18: 100-06.

Taniue, K., and N. Akimitsu. 2022. 'Aberrant phase separation and cancer', *FEBS J*, 289: 17-39.

Terlecki-Zaniewicz, S., T. Humer, T. Eder, J. Schmoellerl, E. Heyes, G. Manhart, N. Kuchynka, K. Parapatics, F. G. Liberante, A. C. Muller, E. M. Tomazou, and F. Grebien. 2021. 'Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression', *Nat Struct Mol Biol*, 28: 190-201.

Thompson-Wicking, K., R. W. Francis, A. Stirnweiss, E. Ferrari, M. D. Welch, E. Baker, A. R. Murch, A. M. Gout, K. W. Carter, A. K. Charles, M. B. Phillips,

U. R. Kees, and A. H. Beesley. 2013. 'Novel BRD4-NUT fusion isoforms increase the pathogenic complexity in NUT midline carcinoma', *Oncogene*, 32: 4664-74.

Thorvaldsdottir, H., J. T. Robinson, and J. P. Mesirov. 2013. 'Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration', *Brief Bioinform*, 14: 178-92.

Tulpule, A., J. Guan, D. S. Neel, H. R. Allegakoen, Y. P. Lin, D. Brown, Y. T. Chou, A. Heslin, N. Chatterjee, S. Perati, S. Menon, T. A. Nguyen, J. Debnath, A. D. Ramirez, X. Shi, B. Yang, S. Feng, S. Makhija, B. Huang, and T. G. Bivona. 2021. 'Kinase-mediated RAS signaling via membraneless cytoplasmic protein granules', *Cell*, 184: 2649-64 e18.

Virarkar, M., M. Mallery, M. Saleh, N. S. Ramani, A. C. Morani, and P. Bhosale. 2021. 'Clinical, Radiographic, Pathologic Characterization and Survival Outcomes of Nuclear Protein of the Testis Carcinoma', *J Comput Assist Tomogr*, 45: 431-41.

Wang, B., W. Gan, X. Han, N. Liu, T. Ma, and D. Li. 2021. 'The positive regulation loop between NRF1 and NONO-TFE3 fusion promotes phase separation and aggregation of NONO-TFE3 in NONO-TFE3 tRCC', *Int J Biol Macromol*, 176: 437-47.

Wang, B., L. Zhang, T. Dai, Z. Qin, H. Lu, L. Zhang, and F. Zhou. 2021. 'Liquid-liquid phase separation in human health and diseases', *Signal Transduct Target Ther*, 6: 290.

Wang, M., Z. Chen, and Y. Zhang. 2022. 'CBP/p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos', *EMBO J*, 41: e112012.

Wang, R., Q. Li, C. M. Helfer, J. Jiao, and J. You. 2012. 'Bromodomain protein Brd4 associated with acetylated chromatin is important for maintenance of higher-order chromatin structure', *J Biol Chem*, 287: 10738-52.

Wang, R., W. Liu, C. M. Helfer, J. E. Bradner, J. L. Hornick, S. M. Janicki, C. A. French, and J. You. 2014. 'Activation of SOX2 expression by BRD4-NUT oncogenic fusion drives neoplastic transformation in NUT midline carcinoma', *Cancer Res*, 74: 3332-43.

Wang, R., and J. You. 2015. 'Mechanistic analysis of the role of bromodomain-containing protein 4 (BRD4) in BRD4-NUT oncoprotein-induced transcriptional activation', *J Biol Chem*, 290: 2744-58.

Wang, S., J. Li, W. Tong, H. Li, Q. Feng, and B. Teng. 2020. 'Advances in the pathogenesis and treatment of nut carcinoma: a narrative review', *Transl Cancer Res*, 9: 6505-15.

Wang, T., H. Gao, W. Li, and C. Liu. 2019. 'Essential Role of Histone Replacement and Modifications in Male Fertility', *Front Genet*, 10: 962.

Wei, M. T., Y. C. Chang, S. F. Shimobayashi, Y. Shin, A. R. Strom, and C. P. Brangwynne. 2020. 'Nucleated transcriptional condensates amplify gene expression', *Nat Cell Biol*, 22: 1187-96.

Weinert, B. T., T. Narita, S. Satpathy, B. Srinivasan, B. K. Hansen, C. Scholz, W. B. Hamilton, B. E. Zucconi, W. W. Wang, W. R. Liu, J. M. Brickman, E. A. Kesicki, A. Lai, K. D. Bromberg, P. A. Cole, and C. Choudhary. 2018. 'Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome', *Cell*, 174: 231-44 e12.

White, M. E., J. M. Fenger, and W. E. Carson, 3rd. 2019. 'Emerging roles of and therapeutic strategies targeting BRD4 in cancer', *Cell Immunol*, 337: 48-53.

Wingett, S. W., and S. Andrews. 2018. 'FastQ Screen: A tool for multi-genome mapping and quality control', *F1000Res*, 7: 1338.

Wojciak, J. M., M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. 2009. 'Structural basis for recruitment of CBP/p300 coactivators by STAT1 and STAT2 transactivation domains', *EMBO J*, 28: 948-58.

Yadav, V. K., T. Y. Lee, J. B. Hsu, H. D. Huang, W. V. Yang, and T. H. Chang. 2020. 'Computational analysis for identification of the extracellular matrix molecules involved in endometrial cancer progression', *PLoS One*, 15: e0231594.

Yan, J., J. Diaz, J. Jiao, R. Wang, and J. You. 2011. 'Perturbation of BRD4 protein function by BRD4-NUT protein abrogates cellular differentiation in NUT midline carcinoma', *J Biol Chem*, 286: 27663-75.

Yu, D., Y. Liang, C. Kim, A. Jaganathan, D. Ji, X. Han, X. Yang, Y. Jia, R. Gu, C. Wang, Q. Zhang, K. L. Cheung, M. M. Zhou, and L. Zeng. 2023. 'Structural mechanism of BRD4-NUT and p300 bipartite interaction in propagating aberrant gene transcription in chromatin in NUT carcinoma', *Nat Commun*, 14: 378.

Zbinden, A., M. Perez-Berlanga, P. De Rossi, and M. Polymenidou. 2020. 'Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force', *Dev Cell*, 55: 45-68.

Zee, B. M., A. B. Dibona, A. A. Alekseyenko, C. A. French, and M. I. Kuroda. 2016. 'The Oncoprotein BRD4-NUT Generates Aberrant Histone Modification Patterns', *PLoS One*, 11: e0163820.

Zhang, Y., K. Han, X. Dong, Q. Hou, T. Li, L. Li, G. Zhou, X. Liu, G. Zhao, and W. Li. 2021. 'Case Report and Literature Review: Primary Pulmonary NUT-Midline Carcinoma', *Front Oncol*, 11: 700781.

Zhang, Y., T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, and X. S. Liu. 2008. 'Model-based analysis of ChIP-Seq (MACS)', *Genome Biol*, 9: R137.

Zhao, D., S. Fukuyama, Y. Sakai-Tagawa, E. Takashita, J. E. Shoemaker, and Y. Kawaoka. 2015. 'C646, a Novel p300/CREB-Binding Protein-Specific Inhibitor of Histone Acetyltransferase, Attenuates Influenza A Virus Infection', *Antimicrob Agents Chemother*, 60: 1902-6.

Zumbro, E., and A. Alexander-Katz. 2021. 'Multivalent polymers can control phase boundary, dynamics, and organization of liquid-liquid phase separation', *PLoS One*, 16: e0245405.