

Molecular characterization of human histone H2BK120 Ubiquitin Ligase UBR7

**Thesis submitted for the partial fulfilment
of the requirements for the degree of
Doctor of Philosophy (Science)
of
Jadavpur University
2022**

**By
Anirban Dasgupta
Index No. 100/18/Life Sc./25**



**Structural Biology and Bioinformatics Division
CSIR – Indian Institute of Chemical Biology
Kolkata – 700032, India**



सी.एस.आई.आर-भारतीय रासायनिक जीवविज्ञान संस्थान

वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद की एक इकाई
विज्ञान एवं प्रौद्योगिकी मंत्रालय के अधीन, एक स्वायत्त निकाय, भारत सरकार
4, राजा एस. सी. मल्लिक रोड, यादवपुर, कोलकाता - 700 032





CSIR - INDIAN INSTITUTE OF CHEMICAL BIOLOGY

A Unit of Council of Scientific & Industrial Research
An Autonomous Body, under Ministry of Science & Technology, Government of India
4, Raja S. C. Mullick Road, Jadavpur, Kolkata-700 032

CERTIFICATE FROM THE SUPERVISOR(S)

This is to certify that the thesis entitled “**Molecular characterization of human histone H2BK120 Ubiquitin Ligase UBR7**” submitted by Sri. **Anirban Dasgupta** who got his name registered on **27.02.2018** for the award of Ph. D. (Science) degree of Jadavpur University, is absolutely based upon his own work under the supervision of **Dr. Siddhartha Roy** and that neither this thesis nor any part of it has been submitted for either any degree / diploma or any other academic award anywhere before.

Siddhartha Roy 03/06/2022

 **SIDDHARTHA ROY** 
Principal Scientist
Structural Biology & Bioinformatics
CSIR-Indian Institute of Chemical Biology
4, Raja S C Mullick Road
Jadavpur, Kolkata-700 032

(Signature of the Supervisor(s) date with official seal)

Preface

Histone monoubiquitination is a complex and dynamic post-translational modification associated with transcriptional regulation and DNA damage repair and is also implicated in various cancers. UBR7, a newly discovered E3 ligase enzyme which targets H2BK120 for monoubiquitination, was a very interesting candidate for mechanistic study as it distinguished itself from the UBR family members and demonstrated the previous unknown catalytic function of Plant homeodomain (PHD). The study and findings presented in this thesis entitled **“Molecular characterization of human histone H2BK120 Ubiquitin Ligase UBR7”** contribute towards understanding the mechanism of action of human H2BK120 monoubiquitin E3 ligases implicated in tumorigenesis.

I hereby declare this thesis contains literature survey and original research work done by the undersigned candidate as part of doctoral studies. This thesis research has been performed under the supervision of Dr. Siddhartha Roy, Principal Scientist, Structural Biology and Bioinformatics Division, CSIR-Indian Institute of Chemical Biology. The thesis has been divided into five major chapters. Chapter 1 contains the review of literature available on histone PTMs focusing mainly on histone ubiquitination and the enzyme involved in the process. Chapter 2 outlines the objectives of this study. Chapter 3 includes a detailed description of the materials and methods used to carry out this thesis work. Chapter 4 includes the results and data obtained from the present thesis work and finally, chapter 5 discusses the results and significance of the present study.

All information in this thesis has been obtained and presented in accordance with existing academic rules and ethical conduct. I also declare that, as required by these rules, I have fully cited and referred all materials and results that are not original to this work.

Anirban Dasgupta

Anirban Dasgupta

Structural Biology and Bioinformatics Division

CSIR-Indian Institute of Chemical Biology, Kolkata, India

Index No. 100/18/Life Sc./25, dated 27.02.2018

Dedicated to,

217

and everyone associated with it.

Acknowledgements

You either leave a hero, or live to write an extremely long acknowledgement.

Humour aside, I have an extremely long list of people who I am thankful for. First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Siddhartha Roy for entrusting me with this thesis project that was very close to his heart from the start. He has been a great mentor and had always been extremely patient with me over the years during the extended period of my struggles with this project. Beyond his suggestions on how to approach a scientific problem or how to critically interpret obtained data, he gave me a tremendous amount of freedom and support to pursue my research goals and curve my scientific niche and temperament. I am thankful for the countless life lessons he has casually passed on over the years, which I am beginning to appreciate more and more as time goes by.

I would also like to thank our collaborator, Dr. Chandrima Das (Saha Institute of Nuclear Physics), for her wisdom, encouragement, and invaluable insights without which completion of this PhD thesis would not have been possible. I express sincere thanks to my doctoral advisory committee members Dr. Krishnananda Chattopadhyay (HOD, Structural Biology and Bioinformatics division, CSIR- Indian Institute of Chemical Biology) and Prof. Ratan Gachhui (Head of Life Science and Biotechnology Department, Jadavpur University) for their valuable time in evaluating my progress and for their warm encouragements.

I would like to thank the current director of CSIR-Indian Institute of Chemical Biology, Dr. Arun Bandyopadhyay and his predecessor Dr. Samit Chattopadhyay for being very supportive and allowing me to use the various departmental and well as other institutional facilities during my tenure. I would also like to thank Prof. Chitra Dutta, former HOD and Prof. Siddhartha Roy, former director for welcoming me into the department and the institute respectively, way back in 2014. In addition, I would like to acknowledge Dr. Amitava Sengupta, Dr. Ramalingam Natarajan, Dr. Jayati Sengupta and Dr. Hemanta K. Majumdar for their kind words of encouragement and inspiration. I must also thank Mr. Sandip Chakraborty, Mr. Jishu Mandal, Mr. K. Suresh Kumar, Dr. Ramdhan Majhi, Mr. T. Muruganandan and Mr. M. Vigneshwaram of the Central Instrumental Facility, IICB for their technical help regarding many of my experiments over the years.

I would like to acknowledge University Grants Commission (UGC) for my research fellowship.

I am indeed indebted to all the people of science I met along the way. Dr. Shelley Barton, Dr. Philippe Bouvet, Dr. Jerry Workman, Dr. Brenda Schulman, Dr. Cynthia Wolberger, Dr. Rachel Klevit, Dr. Chris Lima, Dr. Jonathan Pruneda and Dr. Bernhard Lechtenberg – thank you all for your valuable inputs, comments and constructive criticism – which ultimately, helped me better my work. I want to extend my deepest appreciation to Dr. Catherine Day, Dr. Gary S. Shaw, Dr. Cheryl Arrowsmith and most recently, Dr. Yannick Jacob and Dr. Andrew Bowman for sending me constructs. I want to thank my friends at the Fragile Nucleosome platform and the core journal club team for letting me be a part of an exciting new environment, albeit virtually, while we were stuck at our homes during Covid. It certainly helped restore some normalcy back into my life.

I have been extremely fortunate to be amongst a bunch of tremendously talented, witty and kind people in this institute. Anand, Nicky, Avijit – thank you for always lending me reagents, one after the other, and for never saying no. Arijit, Sujoy, Dheeren, Sushrita, Prathama – thank you for sharing double distilled water over the years. Sourav, Syamantak, Ishani and Shreya – thank you for letting me borrow ice whenever I wanted to. Sabyasachi Da, Isha Di, Sulagna Di, Payel, Vipin, Swagata – thank you for being a hoot in all those testing times. Ayan Da, Saroj, Sumit, Shiladitya, Chinmay Da and Shubhro - thank you for the access to the incubator whenever I asked for it! A big cheers to the old crowd – Suparno, Shankha Da, Mayukh, Debraj and Sayantani, Bartika Di, Neerajana Di, Dibyanti Di, Roopkatha Di, Aritra Da, Amrit Da, Gouranga, Oindrila, Satadeepa– for all those countless memories.

‘But we were all time travellers, watching the present turning into memories

And we were all brave, curious and together, wobbling into the future in little, tentative steps.

And so we were all time travellers, with nothing to do but soak in the moment -

Part of us bursting to talk about what we saw,

part of us longing to know what it means.’

Thank you all for being amazing friends along the way.

Forming, storming, norming and performing – four pillars for an age-old concept of teamwork. Cooped up together in a lab space for long hours does leave its mark on you. You end up sharing ideas, aspirations, music, films, food - and with it, some pretty remarkable memories along the way. What do you value most in those co-existing people?

Their continued existence.

Thank you - Dushyant, Sambit and Sinjini for your constant camaraderie. You shouldered with me when going got tough, and helped me get back up time and time again. I cherish you all more than I could ever express. Thank you, Santanu for the cookie-time stories and for guiding me through situations – every time I asked. A special mention to all who have been part of our lab, even if just for a bit – Debjani, Senjuti Di, Siddhartha, Sunil, Anushka- you all were great. Avradeep and Asmit, welcome to 217!

I will forever be grateful to my teachers, who were instrumental in me getting to where I am. Maini, Jem, Burokaku, Debabrata Sir, Subhen Sir, Jahar Da, Shishir Da (from Narendrapur), Rupa Ma'am, Sneha Ma'am, and then all the faculties of PESIT, especially Reshma Ma'am and V. Krishnamurthy Sir. A special shoutout to my brothers from Narendrapur. You have always been an inspiration. Avijit, Abhimanyu, Achira, Sandeep and Grishma, Charubak, Debarshi, Debolina, Sanchar, Soham, and Sritama, thank you for keeping me sane with your friendship. Thank you, AC Milan – for winning the scudetto!

And finally, the people back home. I am here because of their presence in my life. Ma-Baba, thank you for the warm hugs on the thunderous rainy days, for all the outpour of love, for constant sacrifices along the way and the moral principles – past, present and future. You make me feel homesick even before I leave. Kaka, and Nana, thank you for the incomprehensible love and support all these years. You all assert my reason to stay back a little while longer. Amma – I hope I can make you proud every single day. And Purbita – we have been through quite a lot already, haven't we? Sushi plates, airport goodbyes, surviving Covid scares, and most recently, writing this thesis! The fact that I get to cross this line is because of you, because of us. I look forward to all the life-long adventures with you.

Anirban Dasgupta

CONTENTS

List of Figures	1
List of Tables	4
List of Abbreviations	5
Abstract	7
Chapter 1: Review of Literature	8-59
1.1 Introduction to Chromatin	9
1.2 Early Research Leading to Determination of Nucleosome Structure	11
1.3 Nucleosome Structure	12
1.4 Histone Variants	15
1.5 Histone Post-Translational Modifications	18
1.5.1 Histone Methylation	20
1.5.2 Histone Acetylation	22
1.5.3 Histone Phosphorylation	23
1.5.4 ADP Ribosylation	24
1.5.5 Histone Ubiquitination	24
1.5.5.1 H2A Ubiquitination	26
1.5.5.2 H2B Ubiquitination	31
1.6 Ubiquitin Pathway	36
1.6.1 Types of Ubiquitination	37
1.6.2 Types of Enzymes Involved in Ubiquitin Pathway	38
1.6.2.1 Ubiquitin Activating Enzyme (E1)	39
1.6.2.2 Ubiquitin Conjugating Enzyme (E2)	40
1.6.2.3 Ubiquitin Ligase Enzyme (E3)	42
1.7 Chromatin Modifying Proteins	46
1.7.1 Chromatin Remodelers	47
1.7.2 Histone Modifying Enzymes	48
1.7.2.1 RING domain as writer of histone PTMs	49
1.7.2.2 PHD finger as reader of histone PTMs	50
1.8 Role of H2B Monoubiquitination in Cancer	54
1.9 UBR Ubiquitin Ligase Family	55

Chapter 2: Objectives	60-63
Chapter 3: Materials and Methods	64-78
3.1 Chemically Competent Cell Preparation	65
3.2 Cloning of Expression Constructs and Site-Directed Mutagenesis (SDM)	65
3.3 Recombinant Protein Expression and Purification	67
3.4 Purification and Reconstitution of H2A-H2B Dimers	69
3.5 E2 Charging Assay and Purification of UbcH6~Ub Thioester	70
3.6 <i>In vitro</i> Ubiquitination Assay	70
3.7 Thioester Hydrolysis Assay	71
3.8 GST Pull-down Assay	72
3.9 Hexa-histidine (6xHis) Pull-down Assay	72
3.10 Peptide Pull-down Assay	72
3.11 Fluorescence Spectroscopy	73
3.12 SEC-MALS	74
3.13 DFDNB-mediated Cross-linking Assay	74
3.14 Circular Dichroism Spectroscopy	75
3.15 Cell Culture	75
3.16 Gene Overexpression	75
3.17 Co-immunoprecipitation and Western Blot Analysis	76
3.18 Protein Structural Modelling	77
3.19 MALDI-TOF Mass Spectrometry Analysis	77
Chapter 4: Results	79-98
4.1 Preparation of E2-Ub Thioester	80
4.2 Identification and Characterization of UBR7-UbcH6 Interacting Region	83
4.3 Characterization of H2B Ubiquitination by UBR7-UbcH6 Complex	87
4.4 Determination of Oligomeric Status of UBR7 in Solution and its Role in E3 Ligase Function	91
4.5 Comparison of the Mechanism of UBR7 Mediated Ubiquitin Transfer with other H2BK120 Monoubiquitin Ligases	96

Chapter 5: Discussion	99-103
Bibliography	104-131
List of Publications	132
List of Conferences Attended	133
Reprints	134

List of Figures

Fig. 1.1.1 Electron micrograph depicting chromatin spread in low ionic strength, resembling beads on a string	9
Fig. 1.1.2 Chromosomes are composed of hierarchical packaging of fibers of DNA tightly wrapped around histones	10
Fig. 1.3.1 Nucleosome core particle structure	13
Fig. 1.3.2 The nucleosomal acidic patch	15
Fig. 1.4 Protein domain architecture for the core canonical histones (H3, H4, H2A, and H2B), linker histone H1, and major variants of histones H3 and H2A	17
Fig. 1.5.1 Schematic diagram of nucleosome showing all the different histone tail modification sites	19
Fig. 1.5.2 Schematic diagram of nucleosome showing front and side views of histone globular domain modifications	20
Fig. 1.5.3 Structural insights into the PRC1-nucleosome core particle complex	27
Fig. 1.5.4 Structure of RNF168 RING in complex with nucleosome	28
Fig. 1.5.5 Structure of NCP-ubme–GST–53BP1 complex	29
Fig. 1.5.6 BRCA1-UbcH5c/BARD1/nucleosome complex	30
Fig. 1.5.7 Structural insights into the trans-histone crosstalk between H2BK120 monoubiquitination and Dot1L mediated H3K79 trimethylation	32
Fig. 1.5.8 Structural basis of the crosstalk between H2BK120 monoubiquitination and COMPASS mediated H3K4 trimethylation	33
Fig. 1.5.9 Overview of the SAGA DUB module containing the Ubp8 subunit bound to the monoubiquitinated nucleosome	35
Fig. 1.6.1 Crystal structure of ubiquitin (PDB ID: 1UBQ)	37
Fig. 1.6.2 Types of ubiquitination events (mono, multi-mono and poly-ub) and their physiological function	38
Fig. 1.6.3 Schematic representation of ubiquitin pathway	39
Fig. 1.6.4 Structural insights into Ube1-Ub complex	40
Fig. 1.6.5 Overview of E2 structure (Ube2D3 and Ube2G2 superimposed) with important structural features and binding surfaces highlighted	41

Fig. 1.6.6 Mechanism of ubiquitin transfer by HECT E3 Ligases	42
Fig. 1.6.7 Overview of the RING finger domain and its mechanism of ubiquitin transfer	44
Fig. 1.6.8 Structural insights into an E3-E2-Ub complex	45
Fig. 1.7.1 Domain architecture of the chromatin remodeler families, with their ATPase sub-unit highlighted	47
Fig. 1.7.2 Schematic diagram depicting the action of histone modifying enzymes (writers, erasers and readers)	48
Fig. 1.7.3 Structural insights into RNF20 RING domain	49
Fig. 1.7.4 Overview of H3K4me3 binding by BPTF PHD finger	53
Fig. 1.9.1 The UBR protein family	56
Fig. 1.9.2 Cryo-electron structure of the initiation complex comprising of UBR1, Ub Ubc2 and degron peptide	57
Fig. 1.9.3 UBR7 PHD finger is distinct from other canonical RING and PHD fingers	59
Fig. 4.1 Purification of untagged ubiquitin	80
Fig. 4.2 Purification of Ube1	81
Fig. 4.3 Purification of UbcH6 core domain (UbcH6c)	82
Fig. 4.4 Purification of UbcH6c~Ub thioester	83
Fig. 4.5 Identification of E2 interacting domain of UBR7 <i>in vitro</i>	83
Fig. 4.6 Mapping the E2 interacting domain of UBR7 <i>ex-vivo</i>	84
Fig. 4.7 UBR7 does not interact with ubiquitin <i>in-vitro</i> in absence of E2 and its PHD finger is sufficient to mediate ubiquitin transfer	85
Fig. 4.8 Crystal structure of UbcH6 (PDB ID: 5LBN)	86
Fig. 4.9 Exploring the role of UbcH6 loops in stabilizing the UBC fold and mediating ubiquitin charging	86
Fig. 4.10 UbcH6 loops 4 and 7 are involved in UBR7 binding	87
Fig. 4.11 UBR7/UbcH6 enzyme complex catalyzes ubiquitin transfer on to histone H2B C-terminus	88
Fig. 4.12 H2B C-terminal tail is sufficient to act as substrate in UBR7 mediated ubiquitin transfer	89
Fig. 4.13 The affinity of UBR7-PHD is specific towards H2B C-terminal tail	90

Fig. 4.14 Detection and determination of molecular mass of ubiquitinated H2B peptide using MALDI-TOF mass spectra	90
Fig. 4.15 UBR7-PHD finger exists as a dimer in solution	91
Fig. 4.16 UBR7-PHD finger remains conserved among higher eukaryotes	92
Fig. 4.17 DFDNB mediated cross-linking of UBR7-PHD WT and triple mutant (DDM)	93
Fig. 4.18 SEC-MALS analysis of untagged UBR7-PHD WT and DDM mutant	94
Fig. 4.19 Dimerization of UBR7-PHD is essential for its E3 ligase activity	95
Fig. 4.20 Dimerization of UBR7 regulates its E2 and substrate binding ability	96
Fig. 4.21 E2~Ub thioester targeting of UBR7-PHD and RNF20-RING in absence of substrate	97
Fig. 4.22 Substrate induced E2~Ub thioester hydrolysis of UBR7-PHD	97
Fig. 4.23 Substrate association induces conformational change in UBR7-PHD finger	98
Fig. 5.1 Working model of ubiquitin transfer mechanism of UBR7	102

List of Tables

Table 1.1	Summary of histone ubiquitination events and the enzymes involved	25
Table 1.2	Summary of PHD containing proteins and their target histone modifications	51
Table 3.1	Summary of cloned bacterial constructs used in this study	66
Table 3.2	Summary of mutants generated in this study	67
Table 3.3	Summary of antibodies used in this study	76

List of abbreviations

ATP: Adenosine Triphosphate

BARD1: BRCA1-associated RING domain protein

BPTF: bromodomain PHD finger transcription factor

BRCA1: breast cancer susceptibility gene 1

CD: Circular Dichroism

COMPASS: Complex Proteins Associated with Set1

DDM: Dimer Deficient Triple Mutant

DOT1L: Disruptor of telomeric silencing-1 -Like

DFDNB: 1,5-difluoro-2,4-dinitrobenzene

DSB: Double Stranded Break

EDTA: Ethylenediaminetetraacetic acid

EMT: Epithelial to Mesenchymal Transition

HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

H2BK120Ub – H2B Ubiquitinated at Lysine 120

kDa: Kilo Dalton

K_m: Michaelis Constant

NCP-ubme: Nucleosome with ubiquitinated and methylated histones

NP-40: Nonidet P-40

PAGE: Polyacrylamide Gel Electrophoresis

PDB: Protein Data Bank

PHD: Plant Homeodomain

PTM: Post Translational Modifications

RING: Really Interesting New Gene

RNF20: RING finger protein 20

RNF40: RING finger protein 40

SDS: Sodium dodecyl sulphate

SAGA: Spt–Ada–Gcn5 Acetyltransferase

TNBC: Triple-Negative Breast Cancer

Tris – HCl: Tris-Hydrochloride

UBE1 : Ubiquitin-activating Enzyme 1

UBR7: Ubiquitin Protein Ligase E3 Component N-Recognin 7

WT: Wild-Type.

All other abbreviations have been defined during their usage in the text.

Abstract

Ubiquitination of histones is involved in the maintenance of chromatin dynamics and genome stability through its ability to regulate gene expression and DNA damage repair. Nucleosomal histones H2A and H2B have been predominantly targeted for monoubiquitination. In mammals, Histone H2B gets mono-ubiquitinated in the conserved Lys-120 (K120), corresponding to Lys-123 (K123) in Budding Yeast. This initiates a degradation independent pathway which facilitates recruitment of other ‘writer’ enzymes involved in H3K4 and H3K79 methylation which subsequently helps establish an epigenetic crosstalk network with the H2BK120Ub mark. In humans, H2B K120 monoubiquitination is associated with disruption of chromatin compaction and increased transcription. Recently, our work had identified Ubiquitin Protein Ligase E3 Component N-Recognin 7 (UBR7) to be a novel H2BK120 monoubiquitin ligase which pairs with E2 conjugate UbcH6 for its E3 ligase function. In the present thesis work, to gain insight into the previously unknown mechanism of UBR7 mediated ubiquitin transfer, I have extensively mapped the E3-E2 binding interface between UBR7 and UbcH6. The results obtained suggest that atypical PHD finger of UBR7 is crucial for interaction with UbcH6 and concomitant substrate histone H2B monoubiquitination. The critical loop regions of UbcH6 involved in UBR7 interaction were identified and their role in UBR7 mediated H2B monoubiquitination was assessed. The histone H2B C-terminal tail (114-125) is necessary and was found sufficient by itself to undergo UBR7/UbcH6-mediated monoubiquitination. PHD finger was found to mediate dimerization of UBR7 and I used SEC-MALS to determine the molecular mass of UBR7-PHD in solution. Furthermore, the residues mapped to the dimerization interface were found implicated in cancer and were also critical in regulating E3 ubiquitin ligase function of UBR7. Upon dimer deficiency, the E2 and substrate binding of UBR7 was found to be compromised. Finally, I have also compared the mode of ubiquitin transfer of UBR7 to RNF20, a previously reported H2B K120 ubiquitin ligase, through single turnover ubiquitin discharge assays. Interestingly, unlike RNF20, the UbcH6~Ub hydrolysis mediated by UBR7 requires substrate histone H2B association. Substrate H2B binding to UBR7 brings about a change in conformation within the PHD finger which was found to be critical for efficient ubiquitin transfer. RNF20 was not subjected to any such conformational change. Thus, the mechanism of ubiquitin transfer by UBR7 was found significantly distinct from that of RNF20.

Chapter 1

Review of literature

1.1 Introduction to Chromatin

The human genome in a haploid state contains approximately 3 billion base pairs of DNA packed within 23 chromosomes. Since the length of each DNA base pair measures 0.34 nm roughly, the diploid mammalian cells each contain around 2 meters of linear DNA which undergo elaborate packing to get accommodated inside the nucleus of about 10 μm diameter. Within the eukaryotic nucleus, chromatin exists in the form of a dynamic complex of DNA and proteins, both histone and nonhistone, that together collectively undertakes the implementation of proper structural and functional regulation of different DNA mediated processes. Chromatin is constituted of functional units called nucleosomes, which are linked together by linker histones and threaded in the form of arrays, resembling the classic 'beads on a string' appearance, as reported by Ada and Donald Olins in 1974 using electron microscopy (Fig. 1.1.1).

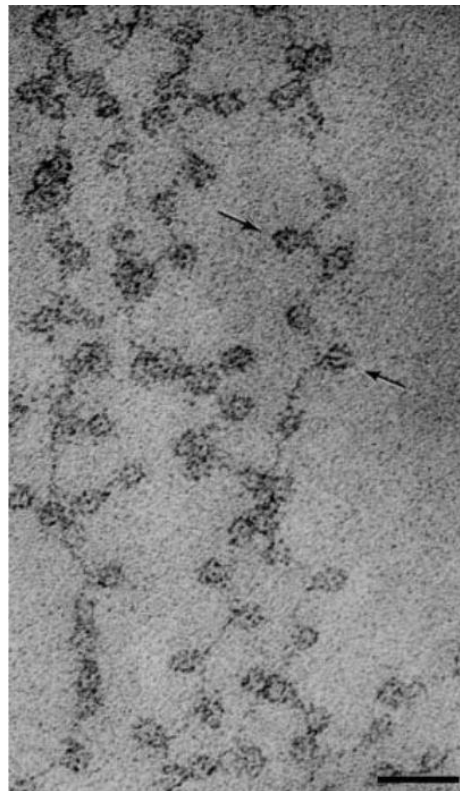


Fig. 1.1.1 Electron micrograph depicting chromatin spread in low ionic strength, resembling beads on a string. Size marker: 30 nm (Adapted from Olins *et al*, 2003)

The nucleosomes in the array further interact with each other to form a compact 30-nm fiber, referred to as the secondary structure of chromatin, which is further organized into higher-

order chromatin structures. This detailed chromatin architecture restricts DNA accessibility and thus it is required for nucleosomes to maintain dynamicity concerning their positioning and state of assembly to allow access to the base readout of DNA (Fig.1.1.2). For this purpose, chromatin is subjected to stringent regulation by multiple factors such as histone modifications or variants, chromatin remodelers, epigenetic modifiers, and histone chaperones.

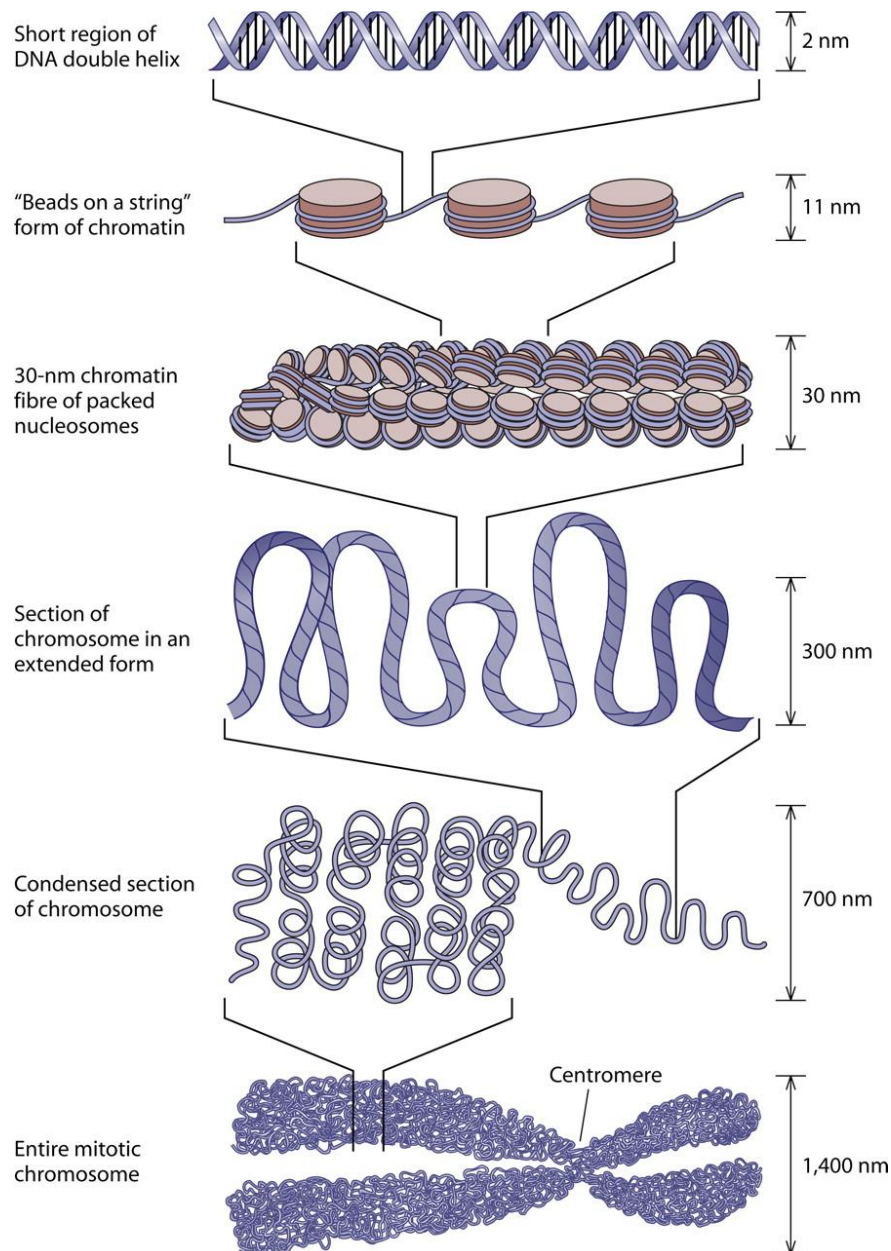


Fig. 1.1.2 Chromosomes are composed of hierarchical packaging of fibers of DNA tightly wrapped around histones. (Adapted from Jansen *et al*, 2011)

1.2 Early Research Leading to Determination of Nucleosome Structure

In the first half of the twentieth century, significant advances were made in the emerging field of genetics. However, not much study went into exploring the structure and function of chromatin. It was not until the discovery of the double-helical structure of DNA (Watson & Crick, 1953), and the subsequent demonstration that DNA constitutes the backbone of continuous units that forms the chromatid (Gall, 1963) that people in this field began to work towards the discovery of fundamental chromatin subunit structure. This was followed by a series of important studies that had a profound influence on the development of this field – preparation of soluble histones (Zubay & Doty, 1959), discovery of the link between acetylation and methylation of histones and chromatin transcription (Allfrey *et al*, 1964) and the fractionation of histones (Johns, 1969), which collectively allowed the scientists to use biophysical tools in their efforts of determining the chromatin structure. In 1971, two independent groups reported that approximately 50% of the DNA in isolated chromatin remained accessible to nuclease degradation (Clark & Felsenfeld, 1971; Itzhaki *et al*, 1971), which was followed by an elegant study reporting the electron micrographs depicting the beads on a string appearance of chromatid strands from chicken erythrocyte nuclei (Olins *et al*, 1974). In the same year, research from two more groups independently established the existence of histone-histone interactions within the chromatin subunit (D'Anna & Isenberg, 1974; Roark *et al*, 1974).

Later in 1974, Roger Kornberg came up with his model of chromatin structure which suggested that a DNA stretch of around 200 bp was in complex with four histone pairs (Kornberg, 1974) which were supported by nuclease digestion and histone crosslinking data generated in collaboration with J. Thomas (Kornberg and Thomas, 1974). In 1975, the structural subunit of chromatin was called “Nucleosome” due to their nuclear origin (Oudet *et al*, 1975) and resemblance to the chicken erythrocyte ‘nu’ bodies as described by Olins *et al*. The coined name has been unanimously accepted ever since. Subsequently, the role of nucleosomes as gene repressors was demonstrated by Kornberg group *in vitro* (Lorch *et al*, 1977) and by Han and Grunstein *in vivo* (Han & Grunstein, 1988). In the 1980s, groundbreaking work from Aaron Klug's group presented structural evidence that a histone protein octamer wraps DNA around itself in about 1.7 turns of a left-handed superhelix (Richmond *et al*, 1984). Tim Richmond's group 1997 solved the near-atomic resolution crystal structure of

the nucleosome for the first time (Luger *et al*, 1997) which helped in gaining insights into the molecular details of the histone constituents and their interacting surfaces. Since then, structures of over 25 different nucleosome core particles (NCP) have been solved, including NCPs containing histone variants and histones from different species. These structures have been solved mostly using X-ray crystallography and more recently, using cryo-electron microscopy.

1.3 Nucleosome Structure

As mentioned earlier, the fundamental repeating unit of the chromatin is called the nucleosome. Each nucleosome contains a core region which is wrapped around by 145 to 147 bp of DNA. The nucleosome core, in turn, is composed of an octameric complex of the histone proteins, two copies each of histones H2A, H2B, H3, and H4 (Arents *et al*, 1991). The adjacent nucleosome cores are interconnected with the help of a linker DNA, which is generally connected with the linker histone protein (H1 or H5). The nucleosome core together with the associated linker histone has been coined as chromatosome (Simpson, 1978). The chromatosome together with the linker DNA is collectively called Nucleosome. Within the nucleus, the nucleosomes organize approximately 200 bp of DNA and contribute toward genomic compaction. They act as a scaffold that hosts and displays a combinatorial array of post-translational modifications (PTMs), besides providing a platform for the binding of several chromatin-modifying enzymes during DNA-mediated processes. The vast variety of PTMs plays crucial roles in the recruitment of chromatin-modifying enzymes, thus regulating the chromatin signalling cascade. The nucleosomes can also undergo self-assembly to form higher-order chromatin structures, resulting in further genome compaction.

The crystal structure of the nucleosome core particle (NCP) solved by the Richmond group (Luger *et al*, 1997) at 2.8 Å resolution helped the researchers working in the field to understand the structural intricacies of the DNA and histone surfaces involved in the assembly of this genomic unit. From the structure, it is evident that 146 bp of alpha-satellite DNA sequence wraps 1.65 times around the core octamer of histone proteins in a left-handed superhelix (Fig 1.3A). A single base pair is positioned at the centre of the nucleosome dyad, which demarcates the 2-fold pseudo-symmetrical axis of the NCP. Super helical locations (SHL) are used to denote DNA locations ranging from SHL -7 to SHL 7, with the

nucleosome dyad at SHL 0. The core histones (two copies each of H2A, H2B, H3, and H4) come together in the form of four histone-fold heterodimers with the N-terminal flexible tails protruding out of the NCP.

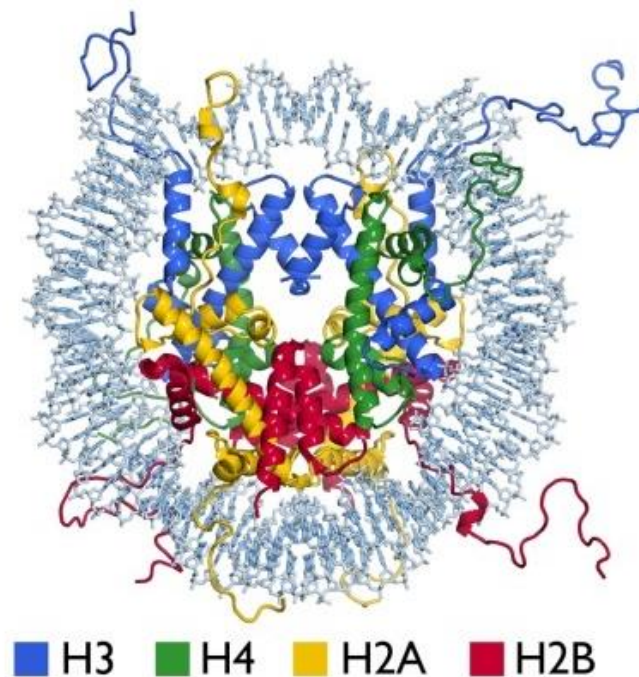


Fig. 1.3.1 Nucleosome core particle structure. Histones are shown as cartoon representations where DNA is highlighted as sticks representation. Histones are differentially coloured as indicated. (Adapted from McGinty & Tan, 2014; PDB ID 1KX5)

Each of the histones contains a central core region of α -helices that forms a histone-fold motif, with N- and C-terminal tails extended outward. The histone fold comprises three α helices connected by two intervening loops (Luger *et al*,1997). The two peripheral helices are positioned to pack against the longer helix located in the centre. Histone fold of H2A pairs with the complementary H2B histone-fold, and similarly H3 pairs with H4 to form a cross-brace motif. The antiparallel heterodimers thus formed result in the formation of a crescent-shaped heterodimer with the $\alpha 1$ helix and the loops towards the convex surface and the $\alpha 2$ and $\alpha 3$ helices on the concave surface. The convex surface of the histone heterodimers exhibits a strong positive charge and constitutes the DNA binding region of the histone fold (Fig.1.3.1).

The histone octamer is assembled using a four-helix bundle formed through mutual interaction of $\alpha 2$ and $\alpha 3$ helices from the adjacent histone folds. The extended structured N- and C-terminal regions also play a part in the assembly of the octamer. The αN helix of histone H3 which is located between its N-terminal tail and histone-fold is positioned on top of the H4 histone fold and plays an important role in defining the DNA entry/exit site of the NCP. C-terminal extensions of both H2A and H2B also contribute to histone packing and spatio-temporal arrangement within the octamer (McGinty & Tan, 2014).

The β -bridges formed by the loops in the H2A-H2B dimer and H3-H4 dimer interact directly with the negatively charged DNA backbone. Moreover, the histone heterodimers within the core octamer form direct contact with three consecutive DNA minor grooves facing inwards towards the histones. The histones have intrinsically disordered N-terminal tails, which remain exposed out from the octamer and act as target sites for histone modifications and thereby play crucial roles in the recruitment of several chromatin modifying enzymes (du Preez & Patterton, 2013). These flexible histone N-terminal tails are generally of 15 to 36 amino acids length and can interact with DNA from the same nucleosome or DNA and histones from adjacent nucleosomes (Zheng & Hayes, 2003).

Apart from the basic histone N-terminal tails, the nucleosomal acidic patch, composed of acidic residues of H2A and H2B provides another interacting hub for nucleosome binding proteins (Barbera *et al*, 2006; Makde *et al*, 2010; Armache *et al*, 2011; Kato *et al*, 2013; McGinty *et al*, 2014; Morgan *et al*, 2016 and Wilson *et al*, 2016). The acidic patch is located on the H2A/H2B dimer surface and encompasses eight acidic residues in total, six from H2A (E56, E61, E64, D90, E91, and E92) and two from H2B (E102 and E110) (Fig. 1.3.2).

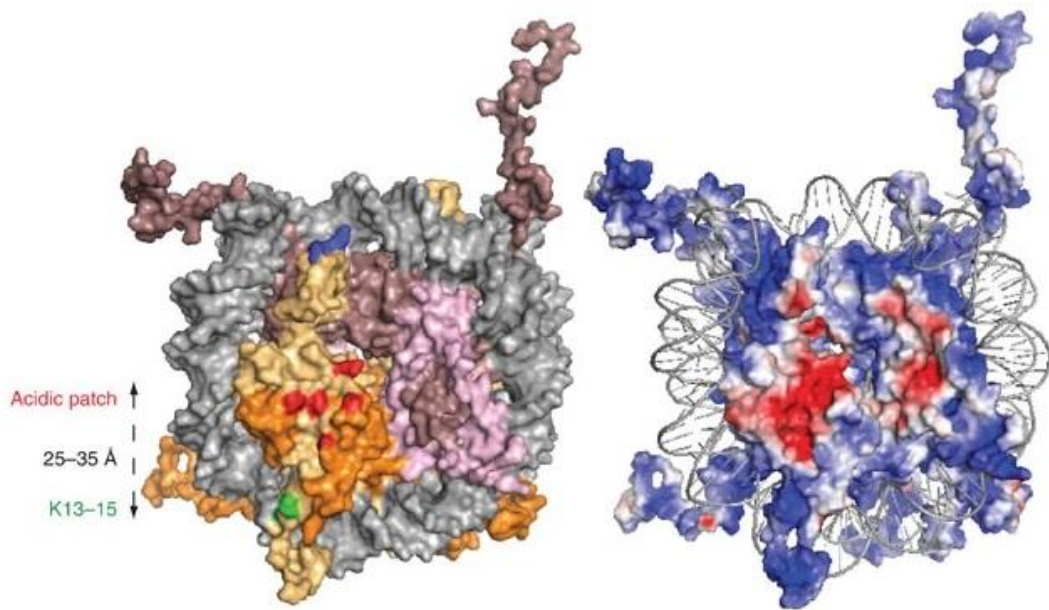


Fig. 1.3.2 The nucleosomal acidic patch. The acidic residues are highlighted in red. H2B, H2A, H3 and H4 are shown in orange, yellow–orange, light pink and violet respectively on the left. On the right, the net basic charge is shown in blue and neutral surfaces are shown in white. DNA is highlighted in grey. (Adapted from Mattioli *et al*, 2014; PDB ID 1KX5)

An intact nucleosome has a diameter of approximately 100 Å and molecular weight of ~200 kDa respectively. The exposed solvent accessible surface of a mononucleosome is about 74 000 Å². From the previously solved crystal structures of reconstituted nucleosomes using histones from Yeast, *Drosophila* and Human (White *et al*, 2001; Clapier *et al*, 2008; Tsunaka *et al*, 2005), the overall nucleosomal architecture was found to remain constant with minor changes in the composition of exposed surfaces due to sequence differences.

1.4 Histone Variants

During the assembly of nucleosome core particles, double-helical DNA is packaged into the compact structure by wrapping around histone proteins. Over time, histone-fold domain-containing proteins have evolved from ancestral archaea into the distinct core octamer subunits of the eukaryotic nucleosome. Within the eukaryotic genome, multiple copies of histone genes are generally present. Expressions of these genes are tightly regulated at multiple levels (Kurat *et al*, 2014). In higher eukaryotes, canonical histone genes remain

clustered and are synthesized for rapid deposition behind the replication fork. Histone variants are paralogs of the eukaryotic core histones which are involved in chromosome segregation, transcription and DNA repair (Talbert & Henikoff, 2010). While synthesis and deposition of canonical histones are generally coupled to DNA replication, incorporation of histone variants is independent of DNA replication and happens throughout the cell cycle. Replacement of canonical histones into noncanonical variants impacts chromatin composition and brings about diversity in nucleosome structure and function (Venkatesh & Workman, 2015; Bushbeck & Sake, 2017).

Of the 4 core histones, H3 and H2A have the most histone variants (Fig.1.4). The centromeric H3 variant, CENP-A (Cse4 in Yeast and CENH3 in plants), when incorporated into the nucleosome, forms the foundation for the assembly of kinetochore (McKinley and Cheeseman, 2016). Another constitutively expressed form of H3, H3.3 acts as substrate for replication-independent nucleosome assembly. H3.3 incorporation occurs at highly active genes and results in subtle changes in the properties of chromatin at actively transcribed loci. Moreover, H3.3 is also associated with repairing the gaps in the chromatin landscape caused due to nucleosome disruption and disassembly (Schneiderman *et al*, 2012; Ray-Gallet *et al*, 2011).

Several H2A variants have also been implicated in the regulation of chromatin. H2A.X is characterized by the presence of a 4 amino acid C-terminal motif where the serine at 139 gets phosphorylated at DNA break sites early in the double-strand break repair process. Phosphorylated H2A.X is denoted by γ H2A.X and it acts as a target substrate for various components of the repair machinery (Podhorecka *et al*, 2010). H2A.X phosphorylation is also involved in mammalian spermatogenesis and has a major role in regulating condensation and pairing. MacroH2A and H2A.B (also called H2A.Bbd) are two vertebrate-specific variants of H2A, which exhibit contrasting features when incorporated within nucleosomes in vitro. H2A.B facilitates transcription while MacroH2A obstructs it.

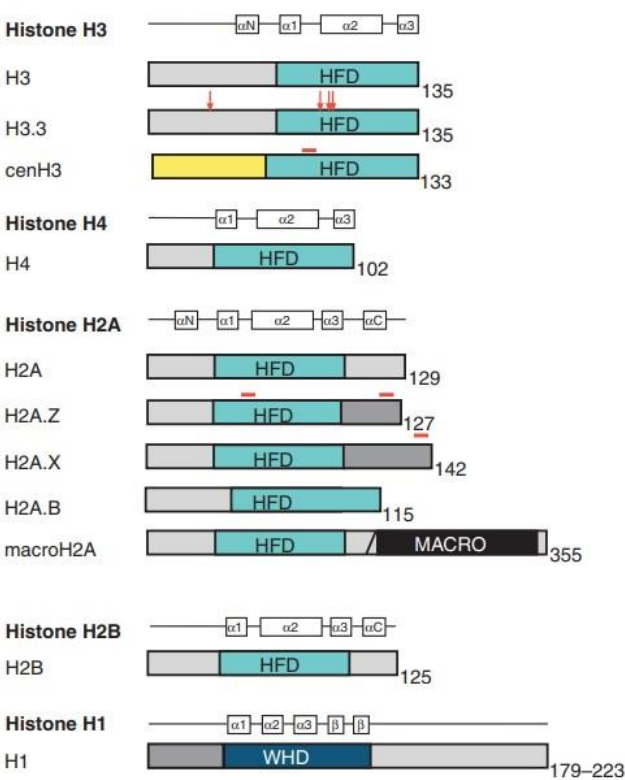


Fig. 1.4 Protein domain architecture for the core canonical histones (H3, H4, H2A, and H2B), linker histone H1, and major variants of histones H3 and H2A. HFD refers to the histone fold domain. The sequence differences between the canonical histones and their variants are highlighted in red. (Adapted from Henikoff & Smith, 2015)

Several structural studies of histone variants have helped in demonstrating some variant-specific roles in the stability of nucleosomes. For instance, CENP-A specific nucleosomes were found to pack about 121 bp of nucleosomal DNA due to a shortened H3 α N helix (Tachiwana & Kurumizaka, 2011). H2A.Z, however, extends the H2A/H2B acidic patch and pushes the H2A/H2B interface across H3/H4 causing subtle destabilization (Dechassa *et al*, 2011). Similar destabilization was reported in the case of testis-specific variant H3T (Tachiwana *et al*, 2010). In addition to eukaryotes, histone-like proteins are also found in archaea. The archaeal histones are smaller than their eukaryotic counterparts despite sharing the histone-fold domain with the same characteristics (Starich *et al*, 1996; Decannaire *et al*, 2000). However, they lack the N- and C-terminal extensions and tails, which in eukaryotic chromatin contributes to nucleosome stability. Archaeal histones are reported to form both hetero- and/or homodimers, which is again in contrast with their eukaryotic counterparts.

1.5 Histone Post-Translational Modifications

Apart from the DNA sequence that wraps around the nucleosome and defines nucleosome positioning, other distinct mechanisms are also involved in modulating nucleosome stability and dynamics. One such mechanism requires chemical alteration of histones which results in changes in the histone – DNA interactions and energy landscape. Those chemical changes that occur in histones post-translationally are referred to as histone post-translational modifications (PTMs). These reversible changes are generally enzyme driven and contribute to dynamic nature of DNA accessibility. The well-studied histone PTMs over the years have been acetylation, methylation, phosphorylation, ubiquitylation, and ADP-ribosylation along with a few more recently reported modifications such as crotonylation, succinylation, and malonylation (Bowman & Poirier, 2014).

The discovery and identification of histone modifications have been largely possible due to the development of mass spectrometric techniques. Different strategies including bottom-up, middle-down, and top-down strategies have been employed from time to time depending on experimental advantages and goals. Bottom-up mass spectrometry can provide the highest accuracy for identifying modifications while analysing trypsin-digested small peptides. Top-down approach deals with intact proteins and can be used to identify a whole set of modifications while the middle-down approach deals with larger peptides obtained from the action of rarer histone cleaving enzymes. With the identification of an entire set of modifications within the same histone, further studies are undertaken to dissect their synergistic or antagonistic roles (Zhao & Shilatifard, 2019).

Few of the above-mentioned modifications occur on the exposed N-terminal tail regions of the histones, which remain accessible on its surface (Fig. 1.5.1). These histone tails after being subjected to these modifications can directly affect inter-nucleosomal contacts and regulate transcription (Akhtar & Becker, 2000) and chromatin compaction positively or negatively (Shogren-Knaak *et al*, 2006; Lu *et al*, 2008) both in vitro and in vivo. Moreover, histone tail modifications are also involved in the direct recruitment of effector proteins for activation of subsequent signalling pathways and indirect roles in recruiting chromatin modifiers and transcription factors (Lawrence *et al*, 2016). They are also reported to restrict association of remodelling complexes (Margueron *et al*, 2005). However, it has also been conversely reported that complete deletion of histone tails does not critically affect

nucleosome stability (Ausio *et al*, 1989). Hence, it is likely that histone tail modifications are not essential for nucleosomal integrity despite being involved in the maintenance of chromatin landscape.

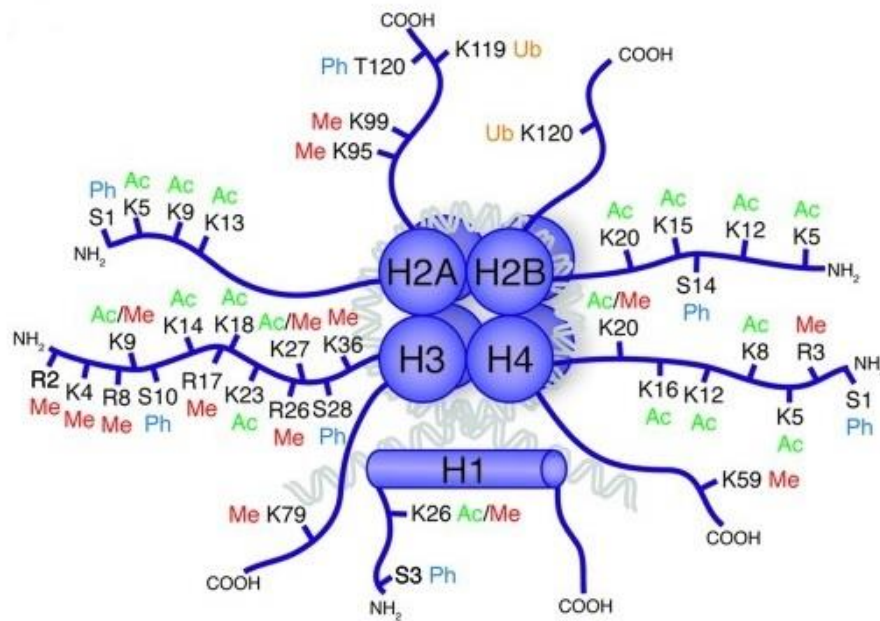


Fig. 1.5.1 Schematic diagram of nucleosome showing all the different histone tail modification sites. The covalent modifications are highlighted on both N- and C-terminal tails of each histone - methylation (Me), phosphorylation (Ph), acetylation (Ac), and ubiquitination. (Adapted from Tollervey & Lunyak, 2012)

Apart from the tails, the central globular domains of the histones, which constitutes the core of the nucleosome, also bear a variety of modification sites (Tropberger & Schneider, 2013). The lateral surface of the histone globular domains, which is located on the surface of histone octamer and directly interacts with DNA, is especially a hub of such modifications (Fig. 1.5.2). Since the discovery of H3 lysine 79 (H3K79) methylation in 2002, many histone modifications localized to the core domains have been identified. These modifications upon being mapped onto the nucleosome crystal structure are classified into three distinct classes - the exposed solute accessible face, the histone-histone interaction interface and the histone lateral surface. Studies suggest that modifications in these classes affect the chromatin structure in distinct ways. Further, the evolutionary conservation of these modifications across species underlines their great physiological relevance.

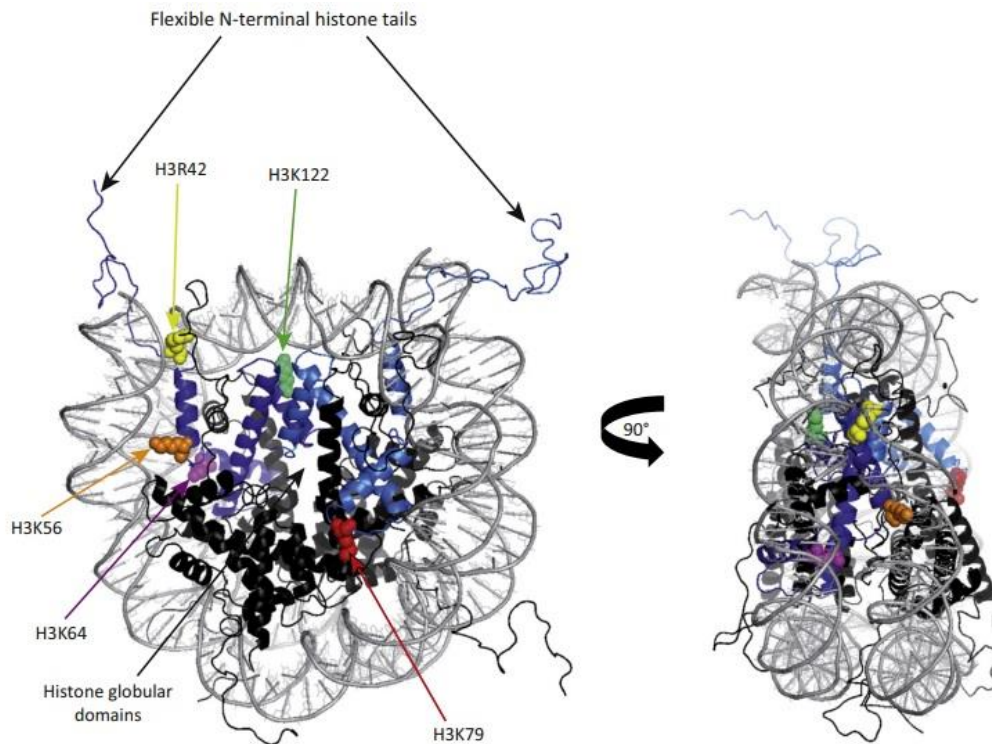


Fig. 1.5.2 Schematic diagram of nucleosome showing front and side views of histone globular domain modifications. DNA is depicted in grey. H3K79, highlighted in red is located on the exposed solute accessible face of the octamer. H3R42 (yellow), H3K56 (orange), H3K64 (pink), and H3K122 (light green) are located on the lateral surface of the histone. (Adapted from Lawrence *et al*, 2016)

1.5.1 Histone Methylation

Histone methylation has been commonly reported at the side chains of lysine and arginine residues. They provide a diverse myriad of complexity since lysine residues may be mono-, di- or tri-methylated, whereas arginine residues are either mono-methylated or symmetrically or asymmetrically di-methylated.

SUV39H1 was the first histone lysine methyltransferase (HKMT) to be identified which was found to target H3K9 (Rea *et al*, 2000). Since then, several HKMTs have been reported, most of which are involved in lysine methylation within the histone N-terminal tails. Interestingly, all these HKMTs share a SET domain which is responsible for the enzyme function. One exception to this general occurrence is the Dot1 enzyme in budding yeast (along with its human homolog, Dot1L) that methylates H3K79 within the histone globular core. Dot1 and its homologs share a catalytic methylase fold resembling that of class I methylases but do not

contain a SET domain. The different HKMTs utilize S-adenosylmethionine (SAM) as a co-factor to mediate the transfer of a methyl group from SAM onto a lysine's ϵ -amino group. In general, the HKMTs tend to be specific towards their target residues. Furthermore, HKMT enzymes also regulate the degree of methylation onto its substrate (i.e., mono-, di- or tri-methyl state). based on their specificity. For instance, the histone KMT5 family in humans comprises of the PR-Set7 and SUV4-20H1/2, which catalyses H4K20 mono methylation and di-/trimethylation, respectively but not the other way around (Jørgensen *et al*, 2013). Elegant structures obtained from X-ray crystallographic studies reveal the presence of key residues within the catalytic domain of the HKMT enzymes that determines the extent of methylation onto the target residue.

In case of arginine methylation, arginine methyltransferase enzymes are classified into 2 types - type-I (which generate Rme1 and Rme2as) and type-II (which generate Rme1 and Rme2s) enzymes. Together, the two classes of enzymes are collectively known as protein arginine methyltransferases (PRMTs). PRMTs mediate transfer of methyl group from SAM to the ω -guanidino group of target arginine residues. PRMT1, 4, 5 and 6 are involved in histone arginine methylations. Histone arginine methylation is considered a high-profile epigenetic mark since it is directly associated with chromatin structure remodelling and regulation of gene transcription (Di Lorenzo & Bedford, 2011). Several sites of arginine methylation have been identified over the years of which, the most well-studied ones are H3R2, H3R17, and H4R3 which are all located on the N-terminal tails of histones H3 and H4.

Histone methylation is a reversible process and involves the participation of demethylases for the removal of methylation marks when needed. Lysine-specific demethylase 1 was the first lysine demethylase to be reported which used FAD as co-factor (Shi *et al*, 2004). LSD1 is active towards mono- and dimethylated lysine substrates. However, LSD1 catalyses demethylation of nucleosomal histones only when it is in complex with the Co-REST repressor complex. Another group of lysine demethylases was found to demethylate tri-methylated lysine residues, using Fe (II) and α -ketoglutarate as co-factors. JMJD2 was the first enzyme to be identified from this group which was found to target H3K9me3 and H3K36me3 (Whetstine *et al*, 2006). The catalytic domain responsible for demethylase activity of JMJD2 was the JmjC jumonji domain which is shared by many other lysine demethylases. Similar to methyltransferases, the demethylases are also very specific towards

their target lysine and sensitive to the degree of methylation involved (Mosammaparast & Shi, 2010). While H3K9me₃, H3K27me₃ and H3R2me have been associated with gene repression, H3K4me_{1/2/3}, H3K36me_{1/2/3}, H3K79me_{1/2/3} alongside H3K9me₁, H3K27me₁, H4K20me₁ have been generally linked to active transcription (Zhao & Shilatifard, 2019).

1.5.2 Histone Acetylation

Histone acetylation was first reported in 1964 by Allfrey *et al.* Since then, several enzymes involved in mediating this modification mark have been identified. Acetylation in histones occurs on the lysine residues and is highly dynamic in nature. They are generally associated with increasing the propensity for gene transcription and are regulated by the mutually antagonistic action of histone acetyltransferases (HATs) and histone deacetylases (HDACs). The HAT enzymes mediate the transfer of an acetyl group to the ϵ -amino group of lysine side chains using acetyl CoA as cofactor. Acetylation of the lysine counters the lysine's positive charge and this leads to weakening of the DNA-histone interactions and helps maintain chromatin in an open state poised for transcriptional elongation. HATs are classified into two major classes: type-A and type-B. The type-B HATs are highly conserved and involved in acetylation of free histones but not nucleosomal histones and thus are predominantly cytoplasmic. They are implicated in acetylation of newly synthesized histone H4 at K5 and K12, thus forming an acetylation pattern which is required for their subsequent deposition onto the DNA, after which the marks are removed (Parthun, 2007). On the other hand, the type-A HATs are more diverse and are classified into three separate groups - GNAT, MYST and CBP/p300 families, depending on the conformational structure and sequence homology of amino acids (Hodawadekar & Marmorstein, 2007). They mediate modification at multiple sites within the histone N-terminal tails, and also within the globular histone core, such as H3K56 (Yang & Seto, 2007; Tjeertes *et al.*, 2009). The type-A HATs have been reported to be a part of large multiprotein complexes wherein the associated component proteins of these complexes are crucial in recruitment, substrate specificity and function. For example, scGCN5 as part of the SAGA complex efficiently acetylates nucleosomal histones but individually can acetylate only free histones (Grant *et al.*, 1997).

HDACs have enzyme activity that reverses that of HATs and removes acetyl groups from already acetylated lysine residues. This helps in the compaction of chromatin and thus

HDACs are associated predominantly with transcriptional repression. They are classified into four classes - classes I and II are members which are closely related to yeast scRpd3 and scHda1, respectively, class IV consists of a single member HDAC11, and class III members (also called sirtuins) are homologs of yeast scSir2. Class III HDACs use NAD⁺ as co-factor as compared to other classes (Yang & Seto, 2007). HDACs have been reported to have relatively low substrate specificity by themselves. The mechanism of HDAC recruitment and specificity is complex as often, HDACs are typically part of distinct complexes which also harbour other HDAC family members. For instance, HDAC1 is part of the NuRD, Sin3a and Co-REST complexes which also have HDAC2 as a constituent member (Yang & Seto, 2008).

1.5.3 Histone Phosphorylation

Similar to histone acetylation, phosphorylation of histones is dynamic and occurs on serine, threonine and tyrosine residues present mostly in the N-terminal histone tails. The levels of the histone phosphorylation are tightly regulated by kinases and phosphatases that are involved in adding or removing the modification respectively (Oki *et al*, 2007). The histone kinases that have been identified so far are generally concerned with transferring a phosphate group onto the hydroxyl group of the target amino-acid side chain from hydrolysis of ATP. This phenomenon adds significant negative charge to the histone that contributes towards alteration of the chromatin structure. However, it is yet unclear in most cases how the kinases are recruited accurately to their active site on chromatin. In case of mammalian MAPK1 kinase, recruitment occurs through its DNA-binding domain which tethers itself to the DNA, like the DNA-binding transcription factors (Hu *et al*, 2009). In some cases, recruitment of kinases might also involve prior presence of chromatin-bound factors before they interact with DNA. Phosphorylated histones have been implicated in regulation of chromatin compaction associated with cell cycle, transcriptional activity and DNA damage repair. Histone phosphorylation has been also reported to work in conjunction with other histone modifications, thereby establishing cross-talk between them. For example, histone H3 phosphorylation at Serine 10 (H3S10ph) can directly affect acetylation levels at H3K9ac and H3K14ac of the same histone. Additionally, it can also interact with H4K16ac which results in transcriptional activation (Alaskhar Alhamwe *et al*, 2018). Although, majority of histone phosphorylation occurs within the N-terminus of histones, in some cases the globular domains may also be targeted. Phosphorylation of H3Y41, mediated by non-receptor tyrosine kinase JAK2 is such an example (Dawson *et al*, 2009). Due to the high phosphatase activity

within the nucleus, the action of the phosphatases like PP1 are crucial in regulating the phosphorylation status of the chromatin. PP1 counters the action of Aurora B kinase and reduces genome-wide H3S10ph and H3S28ph levels during mitosis (Sugiyama *et al*, 2002; Goto *et al*, 2002).

1.5.4 ADP Ribosylation

Histones are targeted by ADP-ribosyltransferases to undergo mono- and poly-ADP ribosylation on glutamate and arginine residues, however not much is known concerning the downstream significance of this modification. Poly-ADP ribosylation process is reversible and is regulated by the mutually counteracting action of poly-ADP ribose polymerase (PARP) family of enzymes, which adds the modification, and poly-ADP-ribose-glycohydrolases, which reverses them. Poly-ADP ribosylated histones have been associated with relatively relaxed chromatin state (Hassa *et al*, 2006). Activation of PARP-1 results in increased levels of core histone acetylation (Cohen-Armon *et al*, 2007). Moreover, PARP-1-mediated ribosylation of KDM5B inhibits H3K4me3 demethylation and blocks its chromatin association (Krishnakumar & Kraus, 2010). Histone mono-ADP-ribosylation has been detected on the linker histone H1 as well as core histones. This is mediated by mono-ADP-ribosyltransferases and has been found to increase during DNA damage response.

1.5.5 Histone Ubiquitination

Among all the different histone modifications, histone ubiquitination stands out due to its size and complexity. This modification pathway involves the addition of ubiquitin, which itself is a 76-amino acid polypeptide, to histone lysine residues via the stepwise action of three different enzymes, E1- activating, E2-conjugating and E3-ligase enzymes (Hershko & Ciechanover, 1998). Substrate specificity (depending on the target lysine residue) and the extent of ubiquitination (mono- or poly-ubiquitinated) are determined by the mediating enzyme complexes. In case of histones, mono-ubiquitination has been mostly reported within H2A and H2B. Monoubiquitination of H2A at lysine 119 (H2AK119Ub) has been commonly implicated in gene silencing (Wang *et al*, 2004), whereas monoubiquitination of H2B at lysine 120 (H2BK120Ub) is associated with transcriptional initiation and elongation and memory. (Lee *et al*, 2007; Kim *et al*, 2009). The histone ubiquitination events and the corresponding enzymes involved in mediating those events have been summarized in table 1.1.

Although H2A was identified as a target for ubiquitination in higher eukaryotes, ubiquitinated H2A was not detected in *Saccharomyces cerevisiae*. Instead, ubiquitinated H2A protein was originally classified as a unique chromosome associated protein called A24. Ubiquitinated H2B was also detected around the same time in mouse cells. Thankfully, it was soon evident that only a single ubiquitin moiety was being attached to both H2A at Lys-119 (H2AK119Ub) and H2B at Lys-120 (H2BK120Ub) in mammals. In yeast, the residue targeted for ubiquitination in H2B corresponds to Lys-123 in *S. cerevisiae* (H2BK123Ub). Unlike polyubiquitination, monoubiquitination is associated with signalling and is therefore reversible. Attached ubiquitin can be removed from target substrates by enzymes called deubiquitinases (DUBs). Among the seven different sub-groups of DUBs, the ones most associated with histone ubiquitination are called ubiquitin-specific proteases (UBPs in yeast; USPs in mammals). The interplay between ubiquitination and deubiquitination is highly dynamic and tightly regulated to mediate various cellular functions. Dysregulation of this balance has been implicated in various diseases, including neurodegenerative diseases and cancer (Weake & Workman, 2004). Apart from transcriptional regulation, histone ubiquitination over the years has been linked to DNA damage response (DDR), DNA replication and maintenance of histone dynamics.

Table 1.1 Summary of histone ubiquitination events and the enzymes involved

Target histone residue	Type of Ub	E3	DUB	Process	References
H2A K13/15	monoUb	RNF168/169	USP51, USP44, USP11, USP3	DDR, DNA replication	Gatti <i>et al</i> , 2012 Wang <i>et al</i> , 2016
H2A K119	monoUb	Ring1A/B (Polycomb repressive complex 1)	BAP1, USP16	Transcriptional regulation	Wang <i>et al</i> , 2004 Uckelmann <i>et al</i> , 2017 Tamburri <i>et al</i> , 2020 Daou <i>et al</i> , 2015
H2A K125/127/129	monoUb	BRCA1/BARD1	USP48	DDR	Kalb <i>et al</i> , 2014 Densham <i>et al</i> , 2016 Saredi <i>et al</i> , 2016

H2B K120 (K123 in yeast)	monoUb	RNF20/40 (Bre1 in yeast)	USP22 (SAGA complex) (Ubp8 in yeast), USP51	Transcriptional regulation, DNA replication, DDR	Nakamura <i>et al</i> , 2011 Zheng <i>et al</i> , 2018 Chernikova <i>et al</i> , 2012 Northam <i>et al</i> , 2016 Atanassov <i>et al</i> , 2016 Ai <i>et al</i> , 2019
H3 K14/18/23	multi-monoUb	UHRF1		Maintenance of DNA methylation during DNA replication	Qin <i>et al</i> , 2015 Li <i>et al</i> , 2018
H3 K14	monoUb	Cul4 (CRLC) in budding yeast		Heterochromatin regulation	Oya <i>et al</i> , 2019
H3 K23/36/37	monoUb	NEDD4		Transcriptional regulation	Zhang <i>et al</i> , 2017
H3 121/122/125	Not specified	Rtt101 ^{Mms1} in yeast and Cul4A ^{DDb1} in human		Histone dynamics during DNA replication	Han <i>et al</i> , 2013
H4 K91	Not specified	BBAP (also known as Dtx3L)		DDR	Yan <i>et al</i> , 2009

1.5.5.1 H2A Ubiquitination

Ubiquitinated histones were originally detected through 2-dimensional gel electrophoresis and H2A was the first histone reported to be ubiquitinated among the core histones. Approximately 5–15% of total histone H2A present in the cell has been found ubiquitinated. However, during the initial years, the E3 ligase enzyme responsible for H2A K119 ubiquitination was not known. In 2003, Yi Zhang and his co-workers successfully fractionated HeLa nuclear proteins and identified Polycomb repressive complex 1 (PRC1)-like as the ubiquitination module for H2A K119 in humans (Wang *et al*, 2004).

Of the constituent subunits of the canonical PRC1 complex, only Ring1B is capable of ubiquitinating H2A *in vitro*, indicating that Ring1B is the E3 ligase responsible for monoubiquitination at H2AK119. From the PRC1-nucleosome core particle complex structure, it is evident that the Ring1B-Bmi1 ubiquitinating module binds E2 UbCH5c and

makes contact with H2AK119 within the NCP in two distinct modes – proximal and distal, with key differences in the UbcH5c-nucleosome interface (McGinty *et al*, 2014) (Fig. 1.5.3).

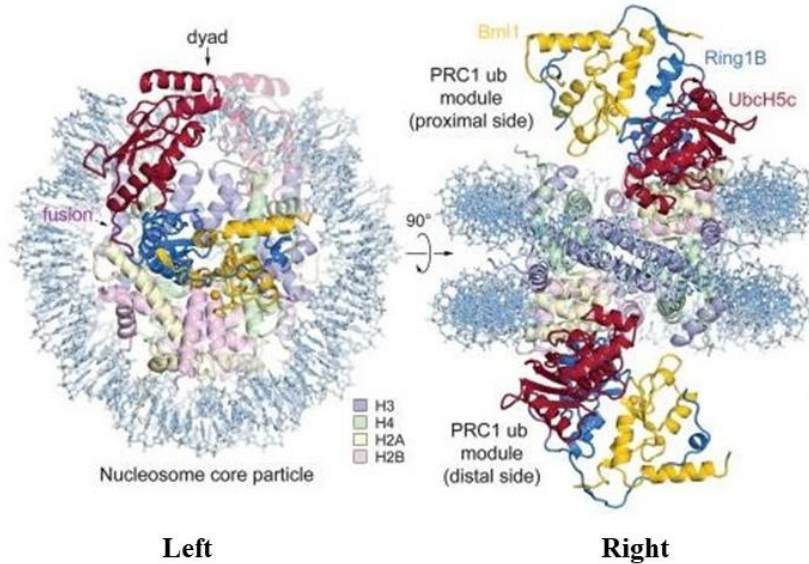


Fig. 1.5.3 Structural insights into the PRC1-nucleosome core particle complex. Left and right denote the upright and orthogonal views of the complex respectively. The orthogonal view allows a differential view of proximal and distal halves of the complex structure. (Adapted from McGinty *et al*, 2014; PDB ID 4R8P)

This has been further confirmed in vivo where upon Ring1B knockdown, H2AK119Ub levels get significantly reduced in *Drosophila* (Wang *et al*, 2004). Among the other constituents, Ring1A and Bmi1 stimulate the E3 ligase activity of Ring1B. Apart from the PRC1 module, 2A-HUB has been identified as another E3 ligase which targets H2AK119. Similar to Ring1B, 2A-HUB also harbours a RING finger and has been shown to ubiquitinate H2AK119 in vitro. In HEK293T cells, overexpression of 2A-HUB results in elevated H2Aub levels (Zhou *et al*, 2008). However, it is not clear how Ring1B and 2A-HUB together mediate H2AK119 monoubiquitination.

Apart from K119, K13 and K15 of the H2A N terminal tail along with K125, K127 and K129 of H2A C-terminus can also be ubiquitinated. K13 and K15 are located fairly distant from H2AK119 and not ubiquitinated by the E3 modules targeting K119. Instead, RNF168 (RING finger protein 168) has been identified as the E3 ligase responsible for H2A ubiquitination at K13 and 15 in response to DSBs (Gatti *et al*, 2012; Mattioli *et al*, 2012). Efforts to co-crystallize E3 ligase RNF168 in complex with its ligand H2AK13/15 have not been

successful over the years. However, amide-backbone based NMR spectroscopy has been used to map the interacting surface between the RNF168^{RING} domain and the H2A-H2B histone dimers (Fig.1.5.4). NMR data, along with crosslinking mass-spectrometry, subsequent mutagenesis and data-driven modelling suggest that RNF168 binds to the nucleosomal acidic patch and directs E2 to the target lysine (Horn *et al*, 2019). Similar structural studies involving NMR spectroscopy show that RNF169 connects histone and ubiquitin surfaces and stabilizes the ubiquitin positioning upon the H2AK15 ubiquitinated nucleosome containing H4K20me2 (NCP-ubme) to form a high-affinity complex. The affinity of RNF169 towards NCP-ubme is much higher than the binding affinity of 53BP1 and thereby results in the displacement of 53BP1 from NCP-ubme (Hu *et al*, 2017).

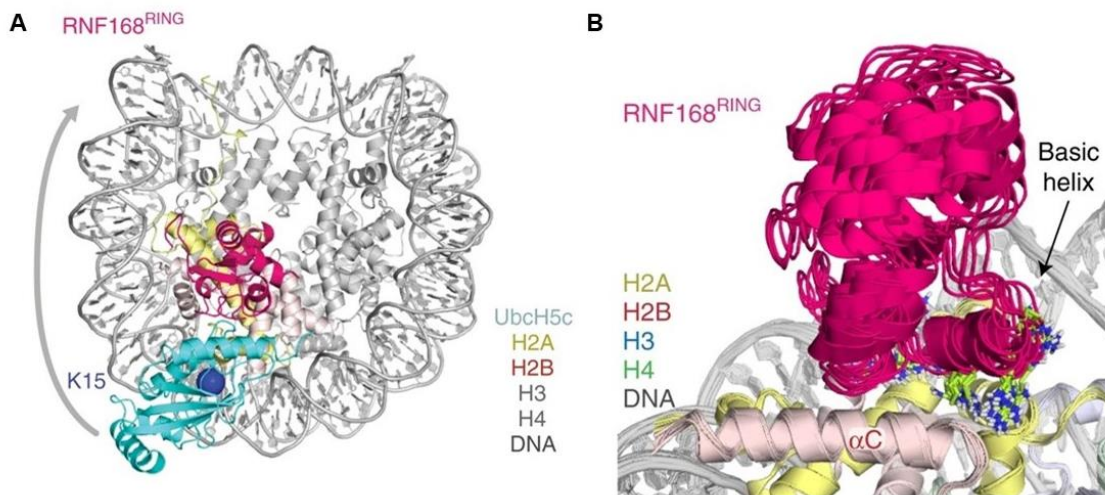


Fig. 1.5.4 Structure of RNF168 RING in complex with nucleosome.

(A) Cartoon representation of the RNF168 RING – nucleosome complex highlighting the 180° rotated position of UbcH5c in contact with RNF168. (B) Superposed scoring solutions of cluster 1 of RNF168 RING contacting the nucleosome as calculated using HADDOCK. (Adapted from Horn *et al*, 2019; PDB IDs 4GB0, 2PYO, 1X23 were used for docking).

Monoubiquitination of H2A is regulated by enzymes which are generally part of large macromolecular complexes and are involved in multiple epigenetic crosstalks within the nucleosome. Both Ring1B and 2A-HUB, the E3 ligase enzymes catalysing H2AK119Ub are linked to gene repression, with Ring1B being shown to associate with repressive complexes such as PRC1 (Weake & Workman, 2008), whereas 2A-HUB interacts with the histone

deacetylase complex N-CoR/HDAC1/3 (Zhou *et al*, 2008). Cumulatively, these data suggest that H2AK119Ub functions as a repressive mark. Moreover, extensive studies show that H2AK119Ub contributes towards Polycomb silencing (Wang *et al*, 2004) and X chromosome inactivation (Fang *et al*, 2004). H2AK119Ub also inhibits RNA Pol II elongation by preventing histone chaperone FACT (facilitate chromatin transcription) to be recruited to the promoters of the repressed genes (Zhou *et al*, 2008).

H1 binding to human nucleosomes is affected by H2AK119 deubiquitination. Thus, by assisting H1 binding, H2A ubiquitination may be involved in maintenance of higher-order chromatin compaction. H2AK119Ub links up with H3K27 trimethylation (H3K27Me3) by acting as a binding scaffold to facilitate the recruitment of Polycomb repressive complex 2 (PRC2), which is responsible for H3K27Me3. (Kalb *et al*, 2014; Barski *et al*, 2007).

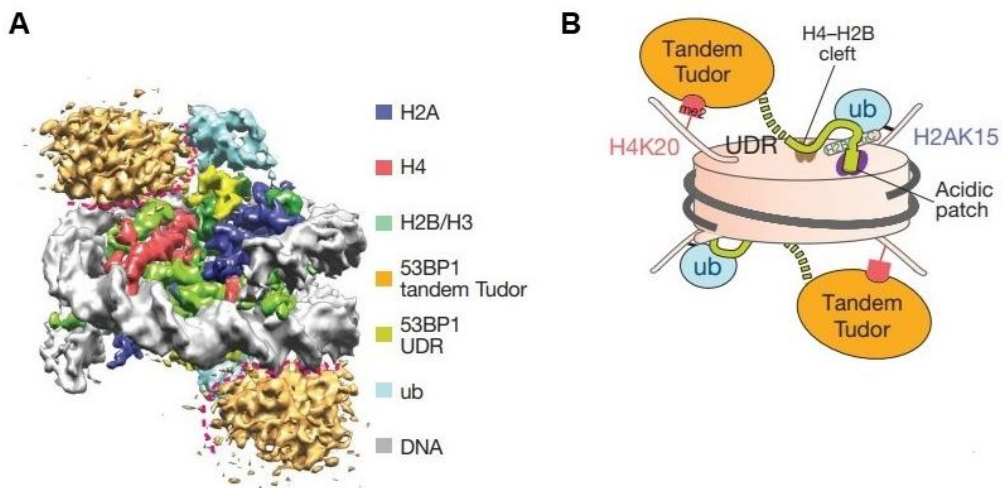


Fig. 1.5.5 Structure of NCP-ubme-GST-53BP1 complex. (A) 3D cryo-EM map of the obtained NCP-ubme-GST-53BP1 complex, where different constituents have been coloured differently. (B) Working model of the NCP-ubme-GST-53BP1 complex. (Adapted from Wilson *et al*, 2016; PDB ID 5KGF)

Based on recent studies, H2AK13/15Ub is considered to be associated with the DNA damage repair pathway (DDR), which is activated in response to double-stranded breaks (DSBs). DSBs trigger release of a whole lot of chromatin modifiers nearby to the damaged site and cause subsequent recruitment of several mediating factors which further activate cell cycle checkpoints and initiate repair pathway. H2AK13/15Ub enables the recruitment of 53BP1

(also known as TP53BP1) which binds specifically to nucleosomes containing H2AK15Ub via the ubiquitination-dependent recruitment motif (UDR). This binding is facilitated by the simultaneous binding of 53BP1 tandem Tudor domain (TTD) to histone H4 lysine 20 dimethylation (H4K20me₂). The cryo-EM structure of 53BP1 with NCP-ubme reveals that 53BP1 makes intimate contacts with the multiple nucleosomal elements including the acidic patch for H4K20me₂ and H2AK15ub binding. UDR motif of 53BP1 was found to be sandwiched between ubiquitin and the NCP surface (Fig.1.5.5). The structure helps to uncover the mechanism of 53BP1 recruitment to DSB sites and demonstrates how multiple epigenetic marks work in combination to provide specific chromatin responses in case of DSBs (Wilson *et al*, 2016).

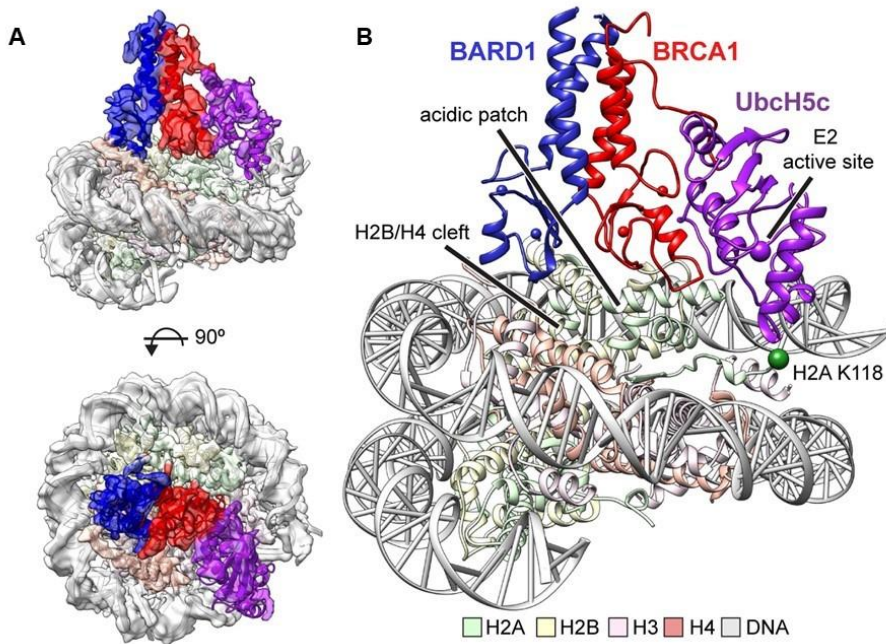


Fig. 1.5.6 BRCA1-UbcH5c/BARD1/nucleosome complex. (A) Cryo-EM density of the BRCA1-UbcH5c/BARD1/NCP complex with semi-transparent surface fitted along with atomic model. (B) Atomic model of BRCA1-UbcH5c/BARD1 ubiquitinating module targeting the nucleosome. (Adapted from Witus *et al*, 2021; PDB ID 7JZV)

The H2A also contains lysine residues in 125,127 and 129th positions of the disordered C-terminal tail which are targeted by BRCA1-BARD1 for ubiquitination. They operate in close

proximity with Ring1b/Bmi1 which targets H2A K118/119 but don't interfere with each other's targets.

The structure of BRCA1/BARD1 RING heterodimer in complex with its E2 conjugate UbcH5c and substrate nucleosome reveals how the complex selectively targets lysines 125/127/129 in the histone H2A C-terminal tail in humans (Fig.1.5.6). In the structure, a novel BARD1-histone interface combines with the UbcH5c being repositioned which is distinct compared to the ubiquitin transfer mechanism of PRC1 E3 ligase that targets H2A Lys119 in nucleosomes. NMR study shows that this distinct mode of E3-mediated substrate regulation is enabled due to the dynamicity in the H2A C-terminal tail. These findings also collectively demonstrate how different E3 ligase complexes preferentially target distinct lysine residues located in proximity to each other by distancing themselves according to the steric landscape of the nucleosome (Witus *et al*, 2021).

1.5.5.2 H2B Ubiquitination

H2B ubiquitination was first reported by Bonner group in 1980 wherein they found that 1–1.5% of total H2B in mouse cells are ubiquitinated (West & Bonner, 1980). Next, in budding yeast, the E2 conjugating enzyme for H2B ubiquitination was identified to be Rad6 which targeted lysine residue 123 (K123) (Robzyk *et al*, 2000). Subsequently, two different groups independently identified Bre1 as the E3 ligase involved in H2B monoubiquitination in *S. cerevisiae* (Wood *et al*, 2003; Hwang *et al*, 2003). Later, Danny Reinberg's lab discovered the human homologues to yeast Bre1 - RNF20 and RNF40, which are the E3 ligases responsible for ubiquitinating H2B at K120, a site analogous to yeast H2BK123. (Zhu *et al*, 2005). They also showed that RNF20/40 along with E2 conjugating UbcH6 can successfully ubiquitinate nucleosomal H2B in vitro. With time, apart from H2BK120, several other lysine residues such as K34, K46 and K108 in H2B have also been found to be ubiquitinated in adult mouse brain (Tweedie-Cullen *et al*, 2009). Several lysine residues on yeast H2B have also been identified which are ubiquitinated in a manner that is independent of Rad6 and Bre1 (Geng & Tansey, 2008).

Ubiquitin is an 8.5 kDa sized protein which when attached to nucleosomal H2B significantly increases the molecular weight of an H2BUb (~22 kDa) as well as the surface area of the ubiquitinated nucleosome. The H2BK120 site which is targeted for ubiquitination is located on the C-terminal α -helix tail and remains exposed on the face of the nucleosome. To test if

the bulky ubiquitin moiety disrupted oligo-nucleosome compaction by hindering inter-nucleosomal interactions, Tom Muir's lab developed a disulfide-linkage method to attach ubiquitin onto a specific target site on H2B and then was able to reconstitute mono and oligo-nucleosomes containing uniformly ubiquitinated H2B species (Chatterjee *et al*, 2010; Fierz *et al*, 2011). Ubiquitinated H2B was found to reduce higher-order chromatin compaction but did not show a significant impact on nucleosome stability in vitro (Fierz *et al*, 2011). However, both H2B copies in the in-vitro system prepared using the disulfide-linkage method were ubiquitinated which was unlike in vivo. Interestingly, ubiquitinated H2B was found to stabilize nucleosomes and promote reassembly of nucleosomes following transcription elongation (Chandrasekharan *et al*, 2009). These results cumulatively suggest that H2B monoubiquitination at K120 has a complex role in the regulation of chromatin structure. On one hand it interrupts inter-nucleosomal interactions, and prevents higher-order chromatin compaction, but at the same time, it facilitates intra-nucleosomal interactions within the cell, with other chromatin-associated factors such as the histone chaperone FACT (Fleming *et al*, 2008).

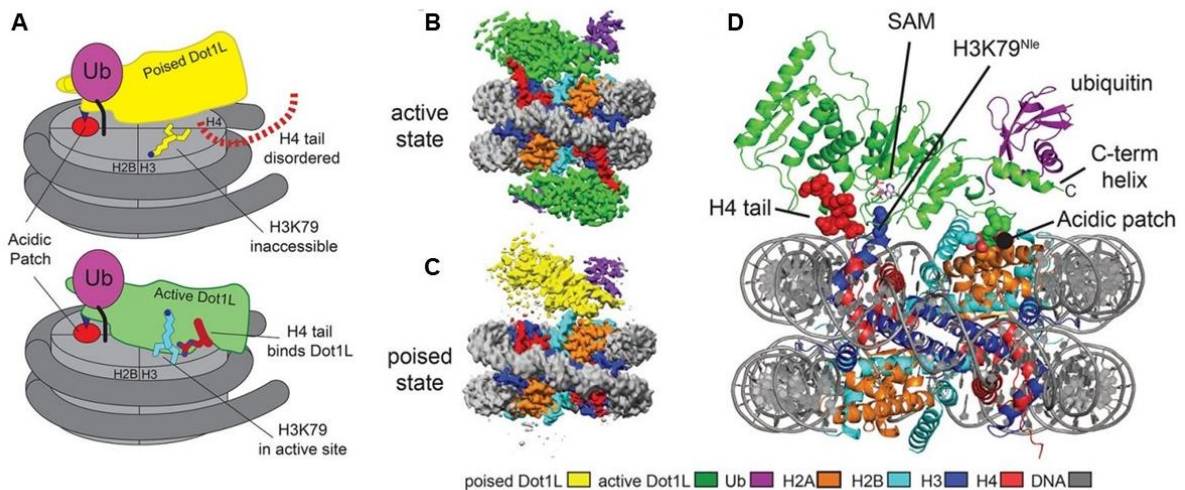


Fig. 1.5.7 Structural insights into the trans-histone crosstalk between H2BK120 monoubiquitination and Dot1L mediated H3K79 trimethylation. (A) Cartoon representation of the mechanism of stimulation of histone H3K79 methylation by H2BK120 ubiquitination. (B, C) Cryo-EM reconstruction map of the Dot1L-ubiquitinated NCP 2-to-1 active state and poised state complex respectively. (D) Atomic

model of the active state complex between Dot1L and the ubiquitinated nucleosome. (Adapted from Worden *et al*, 2019; PDB IDs 6NQA, 6NOG, 6NJ9)

H2B monoubiquitination has been reported to promote di- and trimethylation of H3K4 and H3K79 (Schneider *et al*, 2005; Shahbazian *et al*, 2005; Sun & Allis, 2002). The H2BUb-H3 methylation epigenetic crosstalk remains conserved from budding yeast to humans. H3K79 di- and trimethylation in humans is mediated through the action of Dot1L methyltransferase enzyme. From the cryo-EM structure of Dot1L bound to ubiquitinated nucleosomes, it is now known how H2BUb stimulates Dot1L methyltransferase activity and writer function upon H3 trimethylated at K79 (Fig. 1.5.7). It also reveals how the histone H4 tail plays a crucial role in positioning Dot1L upon the ubiquitinated nucleosome and induces a conformational change in the H3 globular core domain which rearranges itself to make the K79 site accessible for catalysis (Worden *et al*, 2019). Further structural studies from independent labs have reported that Dot1L binds to ubiquitinated H2B through the hydrophobic helix located at its C-terminal end. Dot1L also recognises the nucleosome acidic patch using an arginine anchor (Anderson *et al*, 2019; Valencia-Sánchez *et al*, 2019; Jang *et al*, 2019).

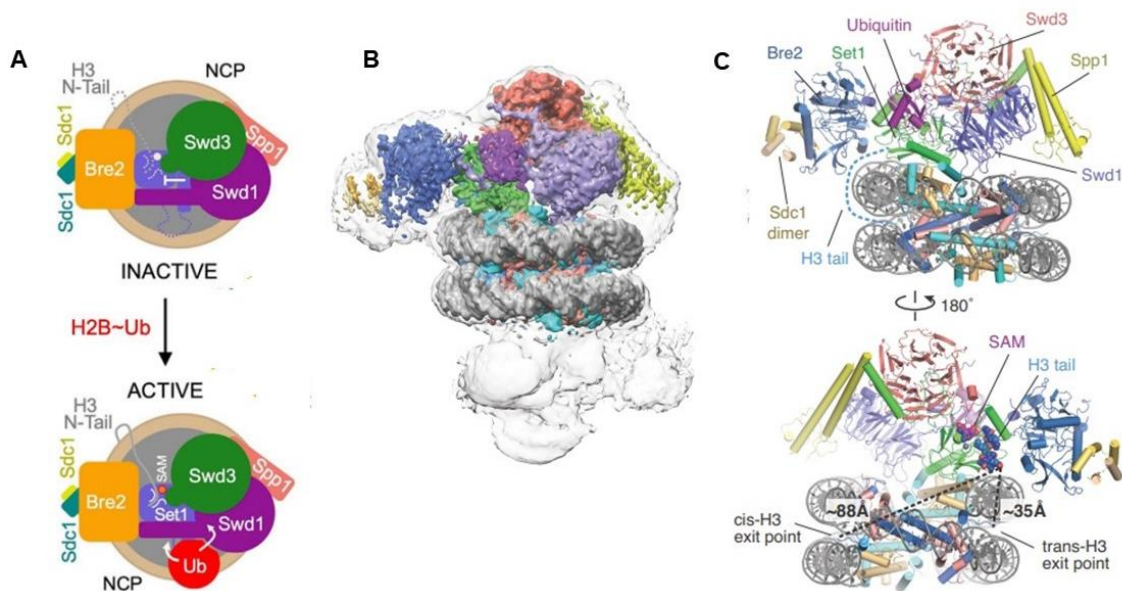


Fig. 1.5.8 Structural basis of the crosstalk between H2BK120 monoubiquitination and COMPASS mediated H3K4 trimethylation. (A) Cartoon representation of the mechanism of activation of COMPASS by H2BK120 ubiquitination. (B) Unsharpened cryo-EM density map of the COMPASS-ubiquitinated nucleosome complex. (C) The model of the complex view as aligned in the previous panel (top) and from the dyad axis

(bottom). (Adapted from Hsu *et al*, 2019; Worden *et al*, 2020; PDB IDs 6NQA, 6NOG, 6NJ9)

H3K4 trimethylation in yeast is mediated by Set1 which is responsible for the methyltransferase activity of COMPASS. From the cryo-EM structure of COMPASS bound to a ubiquitinated nucleosome, it is evident that COMPASS utilizes the whole face of the ubiquitinated nucleosome, binding ubiquitin on one side and subsequently methylating H3 on the alternate side. The ubiquitin-COMPASS interaction enables Set1 to reorient itself to stabilize the interactions with both ubiquitin and nucleosome (Fig. 1.5.8). The critical Set1 arginine-rich motif, which is autoinhibitory when unoccupied, gets packed into the acidic patch, thereby countering the inhibitory effect and activating enzymatic assembly (Hsu *et al*, 2019; Worden *et al*, 2020).

H2B K120 ubiquitination, along with H3K4 and K79 trimethylation has been long associated with transcriptional elongation. Upon deletion of the H2B deubiquitinase enzymes Ubp8 and Ubp10, H2BUb was found localised in genomic sites with enriched trimethylated H3 K4me3 and H3K79me3 respectively, indicating that Ubp8 and Ubp10 probably target different pools of nucleosomal H2Bub (Schulze *et al*, 2011). H2B K120 ubiquitination has also been implicated in DNA damage-induced cell cycle arrest and double-stranded break (DSB) repair pathways. Rad6, which is the E2-conjugating enzyme for H2BK123 in yeast, is associated with several DNA damage signalling pathways. Upon drastic reduction of H2BK123, phosphorylation of yeast Rad9 (a homolog of human 53BP1) is disrupted thereby compromising the DNA damage checkpoint response. DNA damage response in yeast was also found compromised when Dot1 was absent, indicating that crosstalk between H2BUb and H3K79me3 was essential for the cell cycle checkpoint signalling (Giannattasio *et al*, 2005). Similar to RAD51 in yeast, human RNF20-dependent ubiquitination of H2B is also required for homologous recombination (HR) and H2BUb was found to be involved in recruitment of BRCA1 and RAD51 to the damage sites (Nakamura *et al*, 2011). RNF20 and RNF40 both are targeted for phosphorylation by ATM at DSB sites. Moreover, H2BUb also promotes non-homologous end joining (NHEJ) repair factors to be associated with DSBs. These results collectively indicate that H2B monoubiquitinating enzymes, along with the epigenetic mark itself plays a very important role in regulating both DSB repair pathways (Moyal *et al*, 2011). During transcription, the Rad6-Bre1 complex in yeast is associated with RNA polymerase II (RNA Pol II) elongation along with H2B ubiquitination. In budding

yeast, H2BUb is found enriched within the coding regions of actively RNA Pol II-transcribed genes (Schulze *et al*, 2011). Moreover, in an *in vitro* study using reconstituted enzymes, mammalian histone chaperone FACT was found to be involved in the recruitment of the transcription elongation complex PAF, which further enabled the association of E3 ligase enzymes to ubiquitinate H2B. FACT chaperone activity is stimulated by ubiquitinated H2B which results in H2A/H2B dimer displacement, thereby allowing RNA Pol II to proceed along with nucleosomal DNA (Pavri *et al*, 2006). The RNA Pol II stalling is caused due to DNA lesions along the nascently transcribed DNA strand, which is countered by the transcription-coupled repair (TCR) pathway. UV-induced RNA Pol II stalling is triggered by the reduction in H2BK120Ub level and is regulated by H2B-specific deubiquitinase enzymes Ubp8 and Ubp10. In absence of both Ubp8 and Ubp10, TCR gets suppressed. This was further confirmed by the reduced TCR obtained in a *ubp8Δubp10Δ* mutant which coincided with an increase in nucleosome occupancy, suggesting that H2BUb has to be removed for nucleosome destabilization near the stalled RNA Pol II for the TCR to occur smoothly (Mao *et al*, 2014).

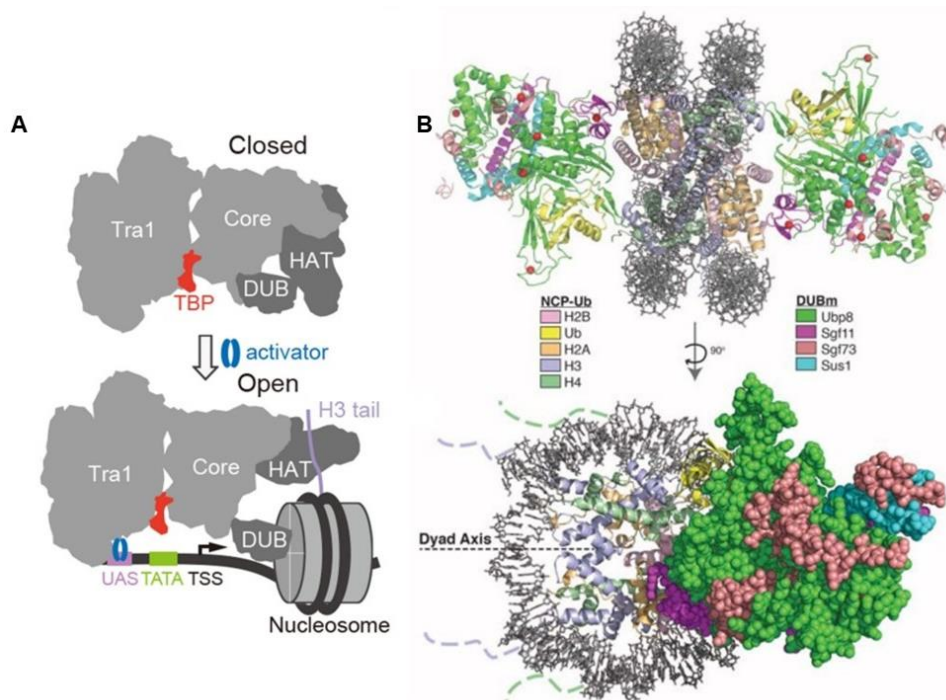


Fig. 1.5.9 Overview of the SAGA DUB module containing the Ubp8 subunit bound to the monoubiquitinated nucleosome. (A) Model showing re-orientation of SAGA DUB module upon nucleosome binding. (Adapted from Wang *et al*, 2020). (B) Crystal

structure showing NCP-Ub being sandwiched within two DUB modules through the catalytic Ubp8 and Sgf11 subunits of the SAGA (top) and its perpendicular view showing the nucleosomal dyad axis (bottom). (Adapted from Morgan *et al*, 2016)

Of the two deubiquitinase enzymes mentioned previously which were involved in H2B deubiquitination, Ubp8 is part of the Spt-Ada-Gcn5 acetyltransferase (SAGA) coactivator complex which catalyses its deubiquitinase function via the SAGA DUB module. The SAGA DUB module in yeast contains Sgf11, Sus1, and Sgf73 apart from the catalytic subunit Ubp8 (Henry *et al*, 2003). The DUB module is connected to rest of the SAGA complex by the C-terminal extension of Sgf73 and the basic features of the catalytic module remain conserved from yeast to humans. From the crystal structure of the DUB – nucleosome complex, two DUB hetero-tetramers were found contacting the nucleosome core particle (NCP) from both faces. The NCP in the complex harboured two copies of ubiquitinated H2B at K120 (Fig. 1.5.9). The structure further reveals that the DUB module primarily interacts with the acidic patch of H2A/H2B through the arginine cluster on the Sgf11 zinc finger domain. Moreover, the Ubp8 catalytic domain also interacts with both ubiquitin and other parts of H2B to ensure re-orientation of the DUB module for efficient DUB activity (Morgan *et al*, 2016; Wang *et al*, 2020).

1.6 Ubiquitin Pathway

PTMs generally involve reversible covalent attachment of molecules such as phosphate, methyl or acetyl groups to protein substrates. As previously mentioned, protein ubiquitination is a far more dynamic and complex modification than most other PTMs as it involves covalent attachment of a small protein - ubiquitin (Ub), or in some other cases -SUMO (Small ubiquitin-related modifier) in the case of SUMOylation, or Nedd8 – in case of Neddylation.

Ubiquitin (Ub) is a stable 76 amino acid protein found conserved from yeast to humans. Structurally, it has a compact β -grasp fold and a flexible six-residue C-terminal tail that ends with two glycine residues (Vijay-Kumar *et al*, 1987). Attachment of ubiquitin to target proteins is referred to as ubiquitination and involves formation of an isopeptide bond, between the ϵ -amino group of a substrate lysine residue and the carboxy-terminus of Ub

(G76) (Hershko *et al*, 1983). Ubiquitin has remained mostly conserved across all eukaryotes. However, yeast and human ubiquitin sequences differ by three amino acids. Primary sequence of ubiquitin contains seven lysine residues which act as secondary Ub accepting sites during the formation of polyubiquitin chains.

The ubiquitin pathway which operates as part of the Ubiquitin Proteasome pathway (UPP) has been previously associated with protein degradation. Polyubiquitination of proteins that are targeted for degradation occurs through a cascade of ATP-dependent enzymatic steps which are then recognized by the 26S proteasome and processed for degradation. However, degradation – independent roles of ubiquitin transfer, especially in monoubiquitination and multi-monoubiquitination of substrates have been since reported and studied. The Nobel prize in chemistry, 2004 was given to Aaron Ciechanover, Avram Hershko and Irwin Rose for their discovery of the ubiquitin-mediated proteolysis.

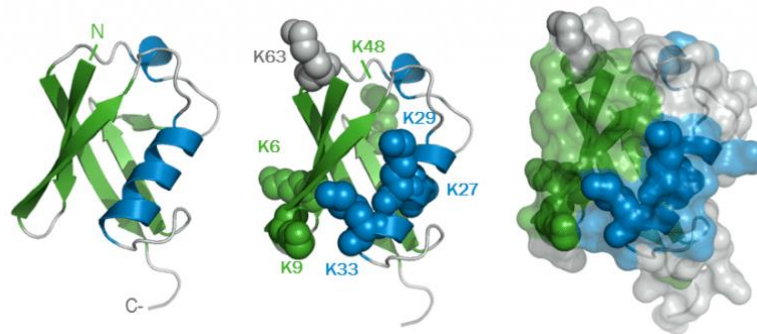


Fig. 1.6.1 Crystal structure of ubiquitin. (PDB ID: 1UBQ) (left) Figure highlighting all the seven lysine residues (middle) which are involved in ubiquitin chain formation. (Adapted from Vijay-Kumar *et al*, 1987)

1.6.1 Types of Ubiquitination

Protein substrates that are targeted for ubiquitination can be modified by attachment of a single Ub molecule (monoubiquitination), or multiple single Ub molecules, each attached to a different lysine residue on the substrate (multi-monoubiquitination) or by Ub chains attached to any one of the lysine residues (polyubiquitination). The diversity and extent of this modification often dictate its functional cue (Fig. 1.6.2). Monoubiquitination is often associated with transcriptional regulation and protein trafficking unlike polyubiquitination which leads to proteolytic downstream events. Most well studied homotypic polyubiquitin

chain formations are possibly Lys48-linked and Lys63-linked Ub-chains. Lys48-linked polyubiquitination is the most prevalent form of ubiquitination and amounts to about half of all Ub chains found in cells (Kwon & Ciechanover, 2017). Lys48-linked chains target a large number of cellular proteins to the 26S proteasome for degradation. These chain formations are dictated by various E3s and the specificity is determined by their E2 conjugate enzymes. Lys63-linked chains, apart from autophagic degradation are also implicated in DNA repair (Hoege *et al*, 2002; Watanabe *et al*, 2004) and activation of protein kinases (Yang *et al*, 2010).

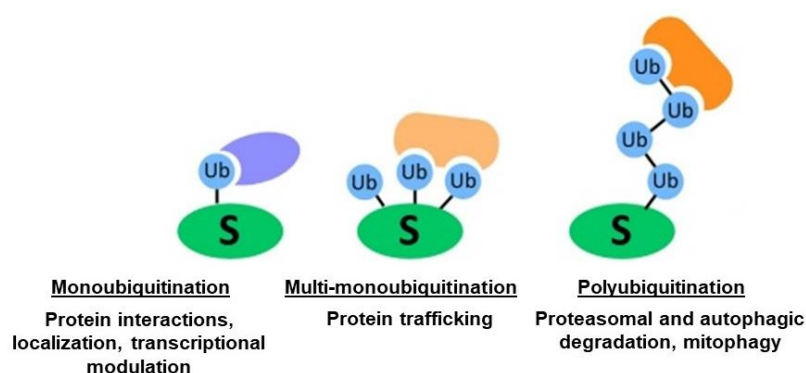


Fig. 1.6.2 Types of ubiquitination events (mono, multi-mono and poly-ub) and their physiological function. (Adapted from Kwon & Ciechanover, 2017)

1.6.2 Types of Enzymes Involved in Ubiquitin pathway

The initial step ubiquitin enzyme cascade is the activation of Ub at its C-terminus by the Ub-activating enzyme E1. Following activation, the ubiquitin that is bound to E1 *via* the thioester linkage is transferred to a sulfhydryl group of Ub conjugate proteins or E2s which contain the cysteine that forms a thioester linkage with the activated Ub. The large number of E2s contributes towards the specificity of the ubiquitination system because different E2s conjugate with one or more E3 ligases, which mediate the final step of ubiquitin transfer onto the highly specific protein substrates (Komander & Rape, 2012). Apart from proteasomal degradation, the removal of ubiquitin from the substrate involves the action of deubiquitinases (Fig. 1.6.3).

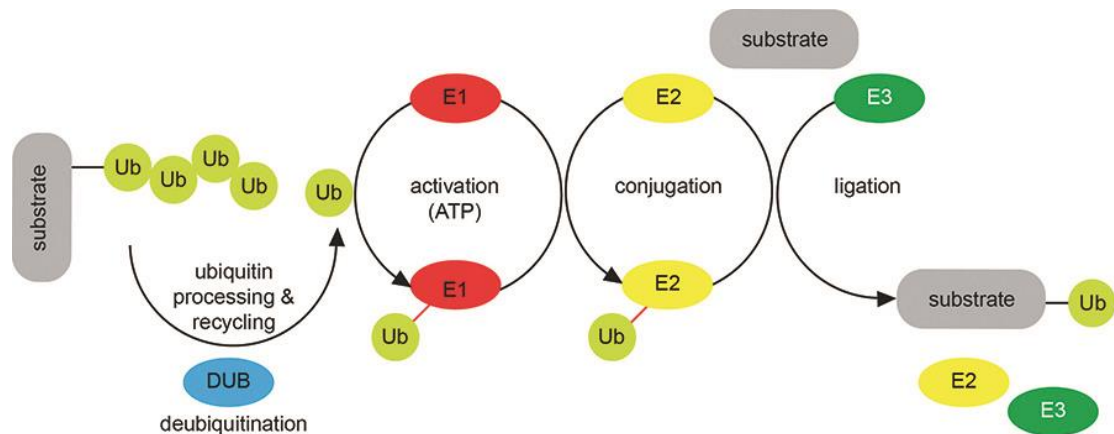


Fig. 1.6.3 Schematic representation of ubiquitin pathway. (Adapted from Kliza & Husznak, 2020)

1.6.2.1 Ubiquitin Activating Enzyme (E1)

Activation of ubiquitin is the first step in the ubiquitin cascade wherein ubiquitin is loaded onto an enzyme called ubiquitin-activating enzyme (E1) (Ciechanover *et al*, 1981). This protein remains highly conserved from yeast to humans (Handley *et al*, 1991; McGrath *et al*, 1991), indicating its essential role in ubiquitin pathway. E1 is a single polypeptide of ~120 kDa in eukaryotes and Ube1 (also called Uba1) is the principal E1 enzyme in eukaryotes. Eukaryotic Ube1 has multiple domains which include an active adenylation domain (AAD), the catalytic cysteine residue containing domain, a four-helix bundle which is an insertion in the inactive adenylation domain and a C-terminal Ub-fold domain (UFD) (Fig. 1.6.3). The AAD is involved in ATP/Ub binding and Ub activation while UFD participates in the recruitment of E2 (Huang *et al*, 2007; Lois & Lima, 2005).

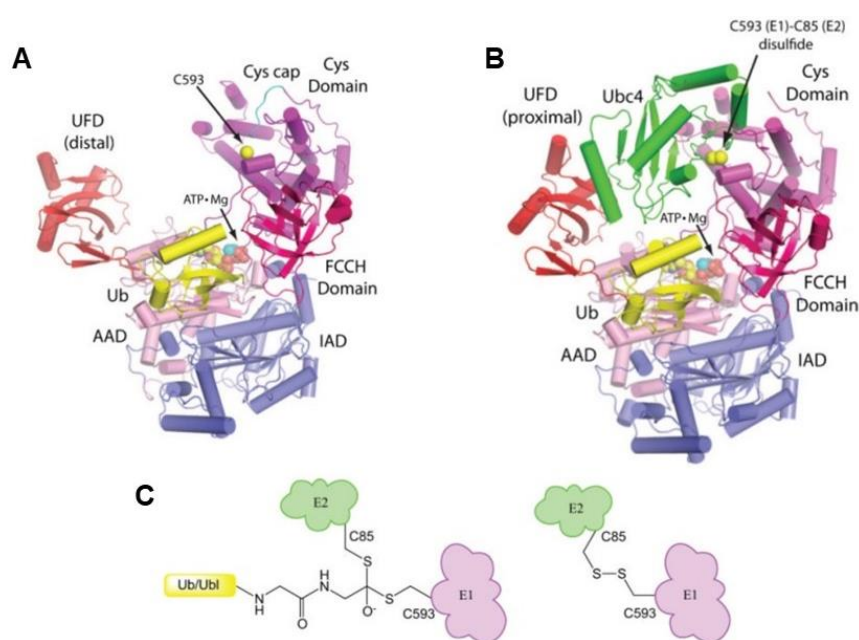


Fig. 1.6.4 Structural insights into Ube1-Ub complex. (A) Cartoon representation of the Uba1/Ub/ATP-Mg complex, with Ube1 domains labelled. AAD and IAD denote active and inactive adenylation domains respectively. (B) Cartoon representation of the Uba1-Ubc4/Ub/ATP-Mg complex. (C) On the left, chemical intermediate of Ub/Ubl during E1-E2 thioester transfer and on the right, E1-E2 disulfide-linked complex. (Adapted from Olsen & Lima, 2013)

During ubiquitin activation, AAD of the Ube1 binds Ub, ATP, and Mg^{2+} while the other adenylation domain stabilizes the complex (Lee & Schindelin, 2008). The adenylation of the Ub C-terminus is accompanied by the release of pyrophosphate (PPi) (Haas & Rose, 1982). Subsequently, Ub is transferred to the active site of E1 forming a thioester bond with the release of AMP while AAD binds to a second Ub, ATP, and Mg^{2+} creating a doubly loaded E1 complex (Pickart *et al*, 1994). This ternary complex next transfers the Ub from the E1 catalytic cysteine to an E2 catalytic cysteine residue. Structures of individual Ube1-Ub-ATP· Mg^{2+} and in complex with E2 suggest that E1 undergoes specific conformational changes to enable Ub transfer to E2s (Olsen & Lima, 2013).

1.6.2.2 Ubiquitin Conjugating Enzyme (E2)

The ubiquitin-conjugating enzymes or E2s are the enzymes which conjugate with the activated ubiquitin after being handed over by the E1 enzyme. There are about 40 different E2 enzymes in humans. They harbour a core catalytic domain called the UBC domain which

is about 150 amino acids long. Although the UBC domain is sufficient for E2 conjugating activity, many E2s contain either an N- and/or C-terminal extension attached to the UBC domain. These extensions sometimes play crucial roles in regulating the catalytic activity of the UBC core domain. Based on the presence of these additional regions, E2s have been divided into four classes:

class I- which contains the UBC domain only – such as Ube2D1-4 (UbcH5a-c), Ube2N.

class II- UBC domain with C-terminal extension – such as Ube2R1-2, Ube2S.

class III- UBC domain with N-terminal extension – such as Ube2e1-3 (UbcH6-8), UbcH10.

class IV- UBC domain with both N- and C-terminal extensions – such as Ube2O, BIRC6.

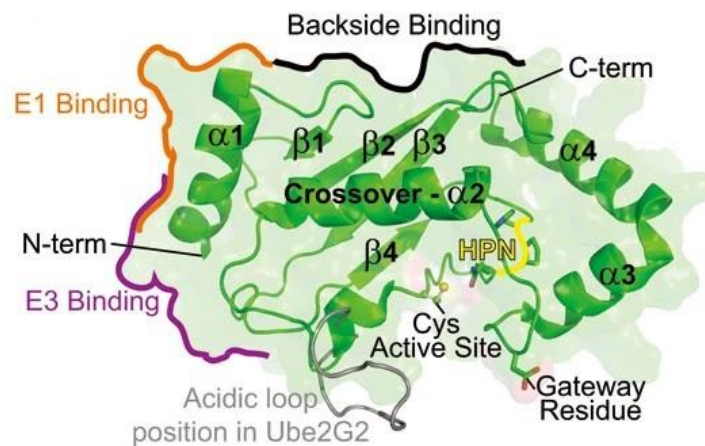


Fig. 1.6.5 Overview of E2 structure (Ube2D3 and Ube2G2 superimposed) with important structural features and binding surfaces highlighted.

(Adapted from Stewart *et al*, 2016; PDB IDs 2FUH, 2CYX).

The cysteine residue, which constitutes the active site of the enzyme, is nested in a narrow groove within the UBC domain. The active cysteine is nearby to the ‘HPN’ motif, which mediate thioester formation. This motif consists of Histidine (H), Proline (P) and Asparagine (N) residues and is usually found in ~10 residues to the N-terminal side of the catalytic cysteine (Figure 1.6.5). The histidine is involved in the stabilization of the E2 enzyme whereas the asparagine is involved in the catalysis of the isopeptide bond formed between ubiquitin and lysine residue of the substrate (Wu *et al*, 2003).

1.6.2.3 Ubiquitin Ligase Enzyme (E3)

More than 600 E3 ligase enzymes exist in humans which are responsible for substrate recognition and specificity and subsequent ubiquitination (Deshaies & Joazeiro, 2009). Based on their ubiquitin transfer mechanism, E3s are categorized into four classes -

- (1) homologous to the E6AP carboxyl terminus (HECT)-type
- (2) Really Interesting New Gene (RING)-type
- (3) U-box
- (4) RING-in-Between-RING (RBR)

HECT E3 ligases were named after the first member of this family to be identified and stand for Homologous to the E6-AP Carboxyl Terminus (Huibregtse *et al*, 1995). These E3s directly accept ubiquitin from E2 conjugate enzymes by forming their own E3~Ub thioester intermediates. In humans, 28 HECT E3 ligase enzymes have been reported (Rotin & Kumar, 2009). Most of the HECT E3 enzymes contain an N-terminal substrate-binding domain and a C-terminal HECT domain. The HECT domain in turn is formed of two lobes bridged together by a flexible hinge loop (Fig. 1.6.6).

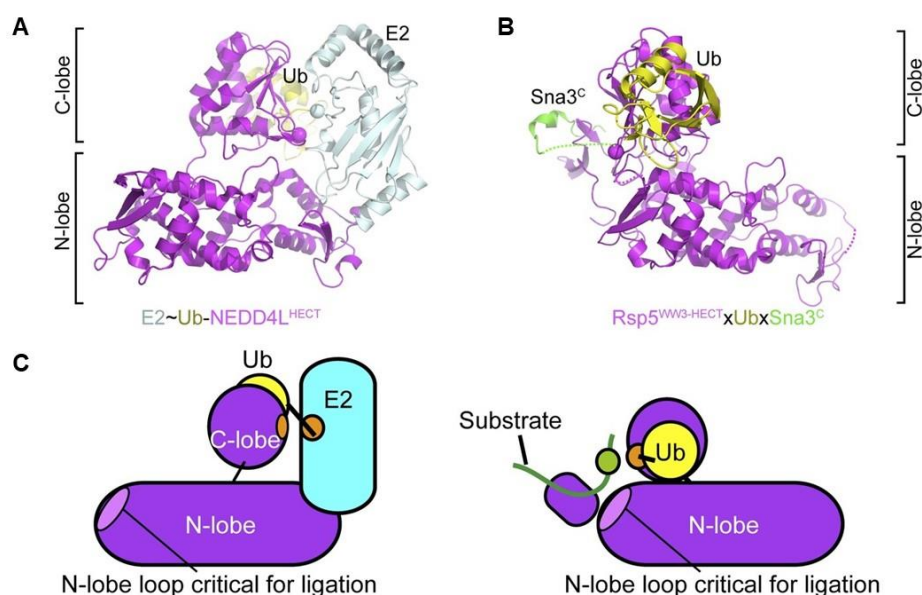


Fig. 1.6.6 Mechanism of ubiquitin transfer by HECT E3 Ligases. (A) Structure of E2 (pale cyan)~Ub (yellow)-NEDD4L HECT (violet) in the complex. (B) Rsp5 WW3-

HECT domain (violet) in complex with Ub (yellow) and Sna3^C (substrate, in green) aligned over the N-lobes of the NEDD4L and Rsp5 HECT domains. (C) Schematic model of E2-to-HECT ubiquitin handover and E3-to-substrate ubiquitin ligation. (Adapted from Kamadurai *et al*, 2013, PDB IDs 3JW0, 4LCD)

The N-terminal lobe interacts with E2~Ub whereas the catalytic cysteine involved in E3~Ub thioester formation is present within the C-terminal lobe (Huang *et al*, 1999). The flexible loop in between enables the lobes to be dynamic and carry out Ub transfer (Verdecia *et al*, 2003; Kamadurai *et al*, 2013). HECT E3s are involved in several cellular pathways like downregulation of cell surface receptors (Rotin & Kumar, 2009), biogenesis of Golgi apparatus (Tang *et al*, 2011), and maintenance of homeostasis in B lymphocytes and pancreatic β cells (Hao *et al*, 2012).

RING E3 ligases all share a conserved RING domain which is generally 60-100 amino acid residue long. Structural studies of RING domains have revealed that conserved cysteine and histidine residues present within the domain stretch are involved in Zn²⁺ coordination through a rigid, globular cross-braced structure. The sequence of canonical RING domain is Cys-X2-Cys-X(9-39)-Cys-X(1-3)-His-X(2-3)-CysX2-Cys-X(4-48)-Cys-X2-Cys, where X can be any variable amino acid (Deshaies & Joazeiro, 2009) (Fig 1.6.7A, B). RING E3s can either be monomeric (CBL), homo- or hetero-dimeric (RNF4, BRCA1/BARD1, RNF20/RNF40), or multi-subunit proteins (Cullin RING Ligase superfamily or CRLs) depending on the number of RING domain(s) involved.

RING domain was first identified in 1991 as a domain that was likely to bind zinc ions based on its eight conserved Cys and/or His residues (Freemont, 1993). Since then, over 600 RING E3 ligases have been reported in humans. The initially characterized RING E3s, such as c-Cbl, Ubr1, AO7, Rbx1/Roc1 (Joazeiro *et al*, 1999; Kamura *et al*, 1999; Lorick *et al*, 1999; Ohta *et al*, 1999), showed that RING-type E3s contain a RING domain and a substrate-binding domain and are actively involved in ubiquitination of the substrate. However, unlike HECT E3s, they mediate ubiquitin transfer directly from E2~Ub onto a substrate(s) (Fig. 1.6.7C)

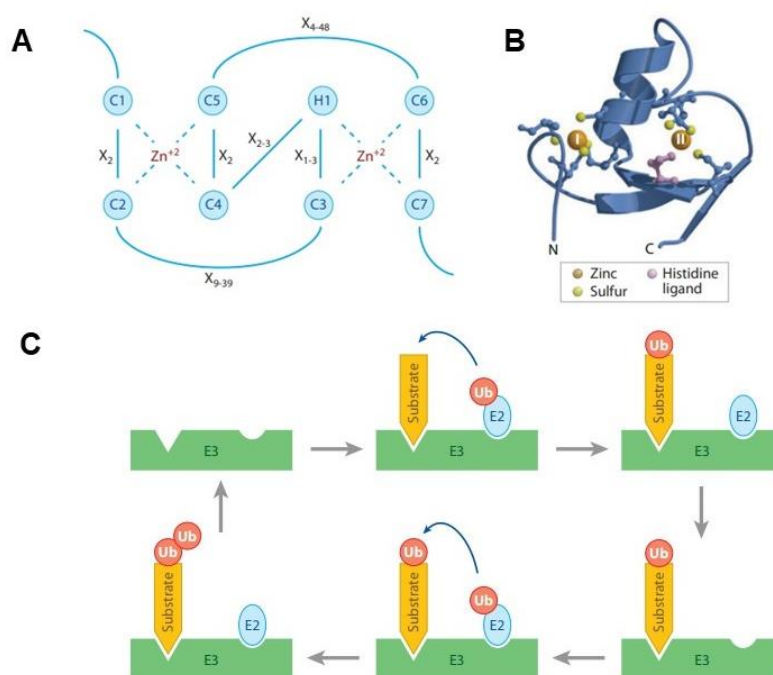


Fig. 1.6.7 Overview of the RING finger domain and its mechanism of ubiquitin transfer. (A) Sequence organization of the Zn²⁺-coordinating residues within the RING domain. Labelling starts from the first cysteine that coordinates zinc – hence denoted as C1, and so on. H1 is the histidine residue in the 5th position in the C3-H-C4 arrangement. Any amino acid in between the Zn²⁺ is represented by X. (b) Ribbon diagram depicting the structure of the RING domain from c-Cbl. The zinc atoms in sites I and II are represented as numbered spheres. (C) Schematic diagram of RING mediated E2~Ub catalysis and substrate ubiquitination. (Adapted from Deshaies & Joazeiro, 2009)

However, not every single RING domain can individually exhibit E3 activity. For example, the RING domains of BARD1, Bmi1, and MdmX can't mediate ubiquitin transfer by themselves. Interestingly, each of those RING domains interacts with a second RING domain protein (BRCA1, Ring1b, and Mdm2, respectively) and this resultant heterodimer formation greatly enhances E3 activity of the latter. Extensive studies suggest that RING domains are involved in recruitment of ubiquitin-conjugating enzymes, however we still do not fully understand if ubiquitin ligase activity correlates to the RING domain's ability to bind E2 with high affinity (Lorick *et al*, 1999). For example, BRCA1-BARD1 when bound to E2 UbCH7, remains inactive for ubiquitin transfer (Christensen *et al*, 2007). But then again, some highly active E2-E3 enzymes do not form a stable complex together. In most cases, the affinity of

individual RING domains towards their corresponding E2 partners is usually low, with K_d values typically in the low micromolar range.

The first crystal structure of a RING E3/E2-Ub complex reported was that of RNF4 in complex with ubiquitin-loaded Ubch5a through an isopeptide bond (Plechanovova *et al*, 2012). RNF4 is a dimeric RING E3 ligase which associates with proteins that are already SUMOylated and ubiquitinates them which leads to degradation (Plechanovova *et al*, 2011; Galanty *et al*, 2012). The structure reveals an RNF4 dimer, with each RNF4 RING bound to a single Ubch5a-Ub conjugate (Fig 1.6.8). The structural details collectively establish that dimeric RING E3s are involved in the activation of the E2~Ub thioester whereby the previously closed conformation of E2~Ub is stabilized through interactions of ubiquitin with both RING domains of a dimer which enables the E2~Ub thioester to be ready for catalysis and subsequent ubiquitin transfer onto the substrate.

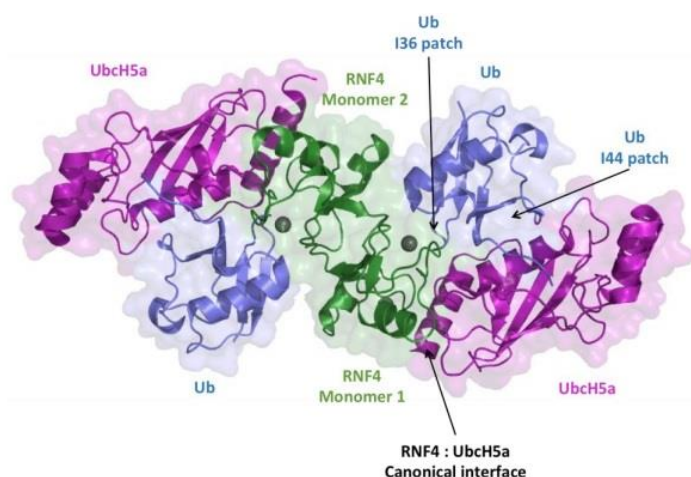


Fig. 1.6.8 Structural insights into an E3-E2-Ub complex. The dimeric RING E3 ligase RNF4 (green) is bound to Ubch5A (magenta) charged with ubiquitin (blue). Each RING binds to an E2~Ub intermediate where ubiquitin packs against the proximal RING domain through the I36 patch and the E2 through the I44 hydrophobic patch. (Adapted from Plechanovova *et al*, 2012)

U-box E3 ligases contain the signature U-box domain and operate similar to the RINGs in terms of mechanism of ubiquitin transfer. Although the U-box domain is distantly related in terms of sequence to the RING finger, it does not contain any conserved zinc coordinating residues. Hence, the RING-like structure in U-box domains is held together by the hydrophobic and salt-bridge interactions. The first U-box protein to be identified to have a

role in ubiquitination was UFD2 (Koegl *et al*, 1999). Six U-box proteins have been reported in humans (Hatakeyama *et al*, 2001).

RBR ligases or RING-between-RING E3s are an interesting class of ubiquitin ligases which project the properties of RING and HECT-type E3s and are regulated through post-translational modifications and autoinhibition for their E3 ligase activity. Interaction with their binding partners is also crucial for them to mediate their function. There are about 14 RBR E3s in humans. Structures of well-characterized RBR E3s such as PARKIN, HHARI, and HOIP reveal the presence of three distinct domains- RING1, IBR and RING2, along with other domains (Beasley *et al*, 2007; Duda *et al*, 2013; Kumar *et al*, 2015; Lechtenberg *et al*, 2016). RING1 is structurally similar to the canonical RINGs whereas IBR and RING2 adopt similar RING like structures. However, RING1 recruits E2~Ub and then transfers the Ub onto the catalytic cysteine on RING2 to form a RING2~Ub intermediate, which resembles the HECT E3 mediated ubiquitin transfer mechanism. RBRs are associated with activation of NF- κ B pathways (Tokunaga *et al*, 2009), mitochondrial damage rescue (Clark *et al*, 2006), and translational regulation (Tan *et al*, 2003).

1.7 Chromatin Modifying Proteins

The dynamic chromatin state is maintained and modulated by chromatin-regulating proteins, which are involved in alteration of chromatin structure, facilitate release of DNA by nucleosome destabilization, and regulation of gene expression by mediating histone modification or mobilizing histone-DNA complex. Chromatin-modifying proteins are divided majorly into two distinct classes,

1. Enzymes that are involved in energy-driven remodelling of DNA-histone structure with the help of ATP hydrolysis – chromatin remodelers.
2. Enzymes that target histones for post-translational modifications – histone-modifying enzymes (writers, readers, erasers).

Both these classes of proteins play a crucial role in nuclear processes, including gene expression and DNA replication, recombination and repair.

1.7.1 Chromatin Remodelers

They are large macromolecular complexes which utilise the energy yielded from ATP hydrolysis to alter the physical state of chromatin by sliding, evicting, or reorienting nucleosomes. Remodelers work in concert with other chromatin associated factors to modulate nucleosome packaging to provide accessibility to the DNA that harbours promoters, enhancers and origin of replication sites for readout by DNA and RNA polymerases. This function is crucial to execute various DNA mediated processes like transcription, DNA replication and DNA repair (Clapier & Cairns, 2009).

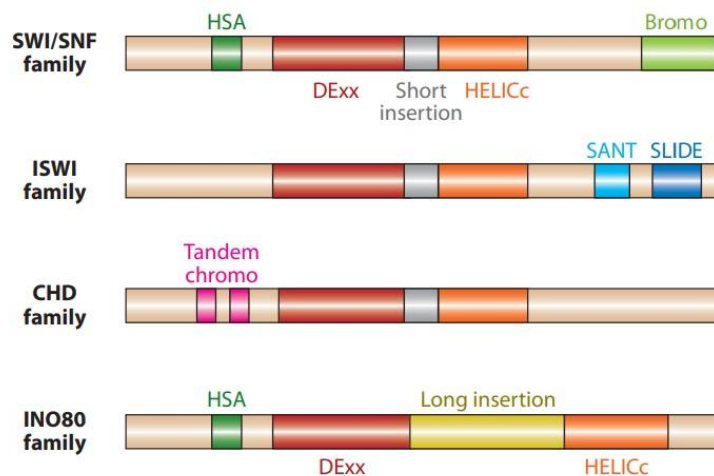


Fig. 1.7.1 Domain architecture of the chromatin remodeler families, with their ATPase sub-unit highlighted. All remodeler families contain an ATPase subunit harbouring an ATPase domain that is divided into two parts: DExx (red) and HELICc (orange). They are distinguished from each other by the presence of mutually distinct domains residing within, or adjacent to, the ATPase domain. (Adapted from Clapier & Cairns, 2009)

They are currently classified broadly into four different families – SWI/SNF, ISWI, CHD and INO80, of which all four are dependent on ATP hydrolysis for the energy required for remodelling and for altering histone-DNA contacts. (Fig. 1.7.1). However, they also have their specialized roles which they perform by means of distinct domains residing within their catalytic ATPases and their signature subunits. The families remain conserved from yeast to humans. Together, these macromolecular complexes are collectively involved in nucleosome recognition, engagement, and remodelling.

1.7.2 Histone Modifying Enzymes

The histone code hypothesis, put forward by CD Allis and his co-workers, predicts that the post-translational modifications of histones, individually or combinatorially, operate to direct specific and distinct DNA-templated processes leading to activation or silencing of regions of the genome for efficient regulation of gene expression. The addition, interpretation and removal of these histone modifications constitute the basis of the histone code (Strahl & Allis, 2000).

Modulation and maintenance of the chromatin landscape through histone PTMs occurs through the cumulative action of histone modifying enzymes - writers, readers and erasers (Fig. 1.7.2). Of these, ‘writers’ are those enzymes that add the specific PTMs to the target histone residues, are divided into classes based on the specific type of modification they write – methyltransferases for methylation, acetyltransferases for acetylation and E3 ligases for ubiquitination. Similarly, enzymes which mediate removal of specific PTMs from histone substrates are called ‘erasers’ and are divided based on the modification they target – demethylases for methylation, deacetylases for acetylation and deubiquitinases for ubiquitination. Finally, ‘readers’ denote the class of specialized protein factors that recognize specific histone post-translational modifications individually or in combination and thus brings about recruitment of various other macromolecular complexes which leads to different transcriptional outcomes.

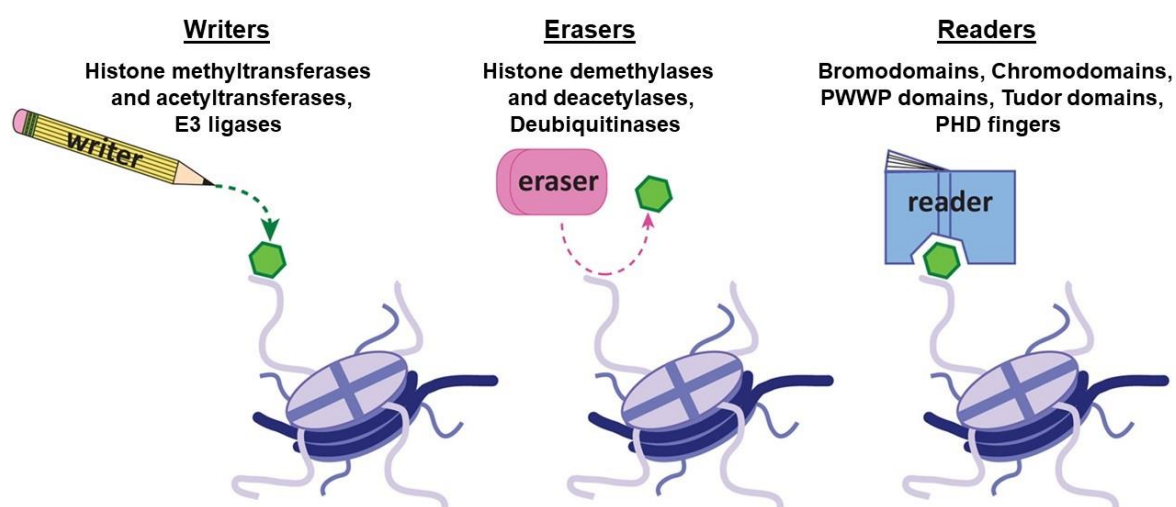


Fig. 1.7.2 Schematic diagram depicting the action of histone-modifying enzymes (writers, erasers and readers).

1.7.2.1 RING domain as writer of histone PTMs

As previously mentioned, the RNF20/40 heterodimeric complex is the E3 ligase module involved in mediating H2B K120 monoubiquitination in humans. The RING domain of RNF20 has been found essential for monoubiquitination of H2B. However, the extended N-terminal region of the RNF20 RING domain also significantly contributes towards E2 recruitment of RNF20/RNF40 complex (Kim & Roeder, 2009). RNF20 overexpression causes enhanced levels of H2B monoubiquitination and increases transcription of a set of HOX genes (Zhu *et al*, 2005). Similarly, depletion of RNF20 causes the H2BK120Ub levels to drop which leads to suppression of p53 targeting genes (Kim *et al*, 2005; Shema *et al*, 2008).

Although RNF20 has been extensively studied due to its catalytic role in H2B monoubiquitination, studies suggest that both RNF20 and RNF40 are active E3 ligases in-vitro. The importance of RING fingers of the RNF20/40 complex in this regard has also been firmly established. Both RNF20 and RNF40 RING domains individually form stable dimers in solution and are found capable of hydrolysing the E2-Ub thioester bond in-vitro (Foglizzo *et al*, 2016).

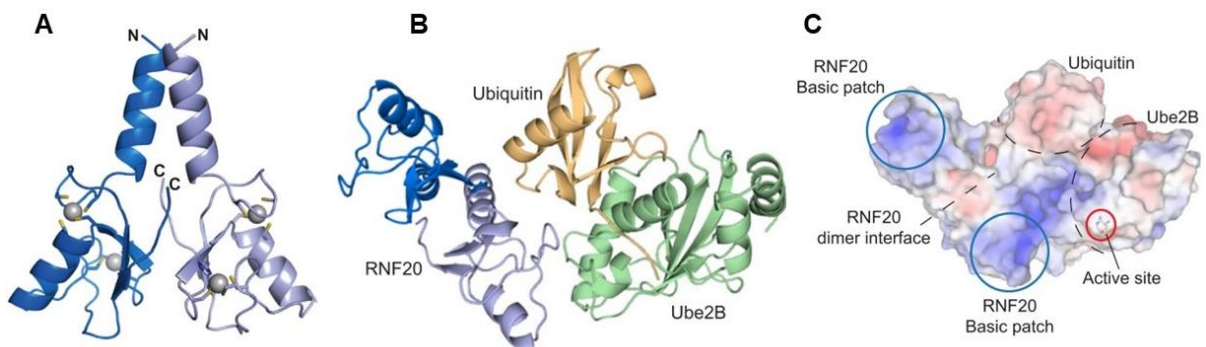


Fig. 1.7.3 Structural insights into RNF20 RING domain. (A) Cartoon representation of the RNF20 RING homodimer. (B) Model of the docked RNF20 RING-Ube2B-Ub complex. (C) The surface representation of RNF20 RING-Ube2B-Ub complex showing electrostatic potential. (Adapted from Foglizzo *et al*, 2016, PDB ID 5TRB)

The structure of the RNF20 RING domain reveals the characteristic cross-braced arrangement holding the two zinc ions in place through one histidine and seven cysteine residues (Fig. 1.7.3). The dimer interface is formed by the mutually interacting residues

within the core RING domain, the N-terminal α -helix, and the C-terminal tail and the RNF20 RING dimer is similar in structural features with dimeric structures of XIAP (PDB ID: 4IC2), BIRC7 (PDB ID: 4AUQ) and yeast Bre1 (PDB ID: 4R7E).

Apart from RNF20/40, RING domain containing proteins RNF168 and CUL4 have been found to mediate ubiquitination of H2AK13/15 and H3K14 ubiquitination respectively (Gatti *et al*, 2012; Oya *et al*, 2019).

1.7.2.2 PHD finger as reader of histone PTMs

Plant homeodomain (PHD) zinc fingers form one of the largest group of epigenetic readers which are associated with recognizing and binding methylated, acetylated or unmodified histone tails. PHD zinc fingers are distributed widely across the eukaryotic genome and play a crucial role in the regulation of gene transcription and chromatin dynamics. Apart from their individual reader function, they participate in the combinatorial readout and interpretation of histone patterns operating tandemly with one or more epigenetic reader domains. This idea is bolstered by the fact that the majority of human PHD fingers are present in proteins that contain other histone reader modules, including bromodomains which bind to acetylated histones.

The PHD finger was first identified in *Arabidopsis thaliana* as a 'Cys4HisCys3' zinc finger motif in the homeodomain protein HAT3.1 (Schindler *et al*, 1993). However, the function of PHD finger was still quite unknown. The identification and characterization of PHD fingers from human bromodomain and PHD finger transcription factor (BPTF) and inhibitor of growth-2 (ING2) for the first time described PHD fingers as a novel class of histone effector module capable of binding and interpreting H3K4me3 epigenetic mark (Li *et al*, 2006; Wysocka *et al*, 2006; Peña *et al*, 2006; Shi *et al*, 2006). Since then, several PHD fingers have been reported to exhibit methylated and acetylated histone-binding activities, summarised below.

Table 1.2 Summary of PHD containing proteins and their target histone modifications

Protein	Histone PTM	Protein function	Biological outcome	References
BPTF	H3K4me3	ATP dependant chromatin remodeler	Nucleosome mobility, Transcriptional activation	Li <i>et al</i> , 2006 Wysocka <i>et al</i> , 2006
ING1	H3K4me3	Histone deacetylase	Transcriptional repression	Peña <i>et al</i> , 2006 Peña <i>et al</i> , 2008
ING2	H3K4me3	Histone deacetylase	Transcriptional repression	Peña <i>et al</i> , 2006 Shi <i>et al</i> , 2006
ING3	H3K4me3	Histone acetyltransferase	Transcriptional activation	Peña <i>et al</i> , 2006
ING4	H3K4me3	Histone acetyltransferase	Transcriptional activation	Peña <i>et al</i> , 2006 Palacios <i>et al</i> , 2008 Hung <i>et al</i> , 2009
ING5	H3K4me3	Histone acetyltransferase	Transcriptional activation	Peña <i>et al</i> , 2006 Champagne <i>et al</i> , 2008
JARID1A (KDM5A)	H3K4me3	Histone demethylase	Transcriptional repression	Wang <i>et al</i> , 2009
MLL1	H3K4me3	Histone methyltransferase	Transcriptional activation	Wang <i>et al</i> , 2010 Chang <i>et al</i> , 2010 Park <i>et al</i> , 2010
PHF2	H3K4me3	Histone demethylase	Transcriptional activation	Wen <i>et al</i> , 2010
PHF8	H3K4me3	Histone demethylase	Transcriptional activation	Horton <i>et al</i> , 2010
PYGO1/2	H3K4me3	Transcription factor, Wnt Signaling	Transcriptional activation	Fiedler <i>et al</i> , 2008 Miller <i>et al</i> , 2010
RAG2	H3K4me3	Recombinase	Recombination	Matthews <i>et al</i> , 2007 Ramon-Maiques <i>et al</i> , 2007 Liu <i>et al</i> , 2007
TAF3	H3K4me3	Transcription factor	Transcriptional activation	Vermeulen <i>et al</i> , 2007 Van Ingen <i>et al</i> , 2008
AIRE	H3K4	Transcription factor	Transcriptional activation	Org <i>et al</i> , 2008 Koh <i>et al</i> , 2008 Chignola <i>et al</i> , 2009 Chakravarty <i>et al</i> , 2009

Chapter 1: Review of Literature

ATRX	H3K4	ATP dependant chromatin remodeler	Chromatin remodelling, Heterochromatin formation	Iwase <i>et al</i> , 2011
BHC80	H3K4	Histone demethylase	Transcriptional repression	Lan <i>et al</i> , 2007
CHD4	H3K4	ATPase, ATP dependant chromatin remodeler, Histone deacetylase	Transcriptional repression, Chromatin remodelling	Musselman <i>et al</i> , 2009 Mansfield <i>et al</i> , 2011
DNMT3A	H3K4	DNA methyltransferase	Transcriptional repression	Otani <i>et al</i> , 2009
DPF3	H3K4	Chromatin remodelling	Transcriptional activation	Lange <i>et al</i> , 2008 Zheng <i>et al</i> , 2010
JADE1	H3K4	Histone acetyltransferase	Transcriptional activation	Saksouk <i>et al</i> , 2009
TRIM24	H3K4	Transcriptional intermediary factor	Both transcriptional activation and repression	Tsai <i>et al</i> , 2010
CHD4	H3K9me3	ATPase, ATP dependant chromatin remodeler, Histone deacetylase	Transcriptional repression, Chromatin remodelling	Musselman <i>et al</i> , 2009 Mansfield <i>et al</i> , 2011
SMCX	H3K9me3	Histone demethylase	Transcriptional repression	Iwase <i>et al</i> , 2007
ECM5 (Yeast)	H3K36me3	Putative histone demethylase		Shi <i>et al</i> , 2007
NTO1 (Yeast)	H3K36me3	Histone acetyltransferase		Shi <i>et al</i> , 2007
DPF3	H3K14ac	Chromatin remodelling	Transcriptional activation	Lange <i>et al</i> , 2008 Zheng <i>et al</i> , 2010

Several PHD fingers have been reported to target the histone H3 N-terminal tail, notably the ‘NH2-ARTKQTARK’ motif. Histone H3K4me3 is a well-known epigenetic mark associated with transcriptional activation and is often enriched at the start sites of actively transcribed genes (Ruthenburg *et al*, 2007). Although PHD fingers are fairly diverse and share low

similarity within their primary amino acid sequences, they all possess a conserved interleaved zinc finger arrangement that rearranges into a compact module consisting of an antiparallel two-stranded sheet within its core. Structural studies of BPTF and ING2 PHD fingers, both of which are associated with H3K4me3 readout reveal a lot of common structural features. In case of BPTF binding, the H3 peptide conforms into an induced β -strand on the existing antiparallel β -sheet core, and embeds itself within the surface groove of BPTF PHD finger, establishing a robust interaction. The free N-terminus of H3, along with the extended side chains of both R2 and K4me3 contribute to the specificity of H3K4me3 binding (Fig. 1.7.4). Many PHD fingers like UHRF1-PHD utilize H3R2 as a docking site for their targeting of ARTK segment of H3. Recent structural studies also indicate that the first zinc coordinating cysteine residue within N-terminal stretch of PHD fingers plays a critical role in histone peptide recognition. Since PHD fingers are very divergent in terms of their sequence and targets, it is plausible that these different PHD fingers operate under different contexts towards their common histone peptide recognition and hence differ in their downstream recruitment signalling.

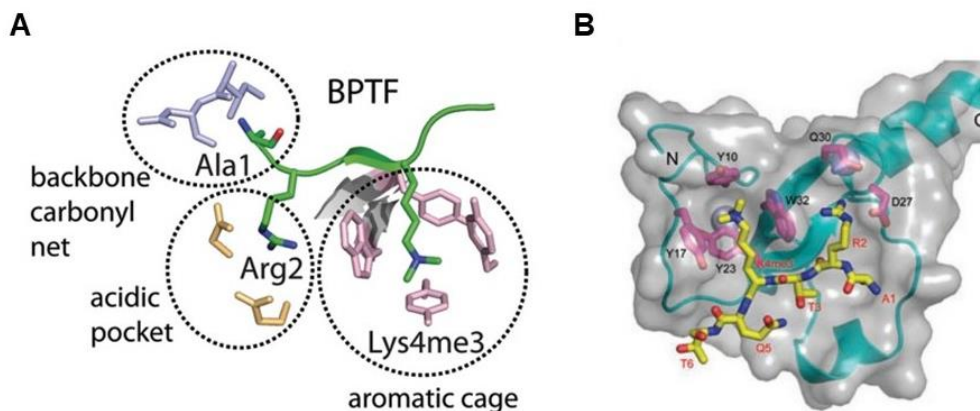


Fig. 1.7.4 Overview of H3K4me3 binding by BPTF PHD finger. (A) The histone-binding sites of the BPTF-PHD are highlighted. (B) Structure of the PHD finger of BPTF in complex with H3K4me3 peptide. (Adapted from Musselman & Kutateladze, 2016; Zhang, 2006 PDB ID 2F6J)

PHD fingers of different families have been reported to be involved in mediating histone crosstalk. Depending on their local binding pocket composition, PHD fingers from different families often display different histone PTM cross-talk sensitivity. For example, H3R2

asymmetrical dimethylation (H3R2me2a) is negatively correlated with H3K4 trimethylation in eukaryotes (Kirmizis *et al*, 2007). PHD fingers of proteins like BPTF, ING4, TAF3, UHRF1 are sensitive to this crosstalk while RAG2, DNMT3a and PYGO are not. These differences indicate variations in the R2-binding pocket of each PHD finger.

Another important feature of PHD modules is that they often exist in pairs and likely operate in a combinatorial manner (Wang & Patel, 2011). PHD fingers have also been shown to be frequently present with other signature reader modules like bromo, chromo, tudor or PWWP fingers. The presence of additional reader modules has been reported to significantly enhance binding affinity and target specificity of PHD fingers. BPTF PHD– bromodomain is involved in such combinatorial readout of H3K4me3 and H4K16ac marks at the mononucleosome level. PHD fingers have also been discovered to mediate SUMOylation - KAP1 PHD finger was found to target adjacent bromodomain which resulted in the repressive function of KAP1 (Zeng *et al*, 2008). PHD fingers are varied and quite abundant in eukaryotic genome and have the potential to drive many more different functions which hopefully would be uncovered as research progresses.

1.8 Role of H2B Monoubiquitination in Cancer

Epigenetic modifications have been well known to play key roles in cancer progression which is characterized by aberrant gene expression. Among the different histone modifications expressed differentially in the malignant cells, H2BK120 monoubiquitination (H2BK120Ub) was identified as a tumour suppressor. Loss of global H2BK120Ub levels, as detected by immunohistochemistry, was found in case of several cancers including breast (Prenzel *et al*, 2011), lung and colorectal (Urasaki *et al*, 2012) and parathyroid (Hahn *et al*, 2012). Loss of H2BK120Ub is often accompanied by mutations in tumour suppressor CDC73 in parathyroid cancers, which disrupts the stability of the RNA polymerase II-associated factor 1 (PAF1) transcriptional complex and brings down the ubiquitination of H2B (Hahn *et al*, 2012). CDC73 is a member of the human PAF1 transcriptional complex that is associated with the RING finger E3 ubiquitin ligases RNF20–RNF40. In benign parathyroid tumours, CDC73 was found to be unaltered and H2BK120Ub levels were concomitantly unchanged.

In breast cancer, basal levels of H2BK120Ub were found to remain unchanged in normal mammary epithelium as well as benign breast tumours; however, H2BK120Ub is significantly reduced in malignant and metastatic breast cancer cells (Prenzel *et al*, 2011). H2BK120 E3 ligase RNF20 itself has been shown to act as a tumour suppressor. RNF20 promoter is found hypermethylated in primary breast cancer cells, which likely leads to disruption of E3 ligase function of RNF20 (Shema *et al* 2008, 2011). RNF20 transcript levels are found reduced in metastatic prostate cancer cells with respect to benign disease cells (Varambally *et al*. 2005). Several mutations have been reported in other H2BK120Ub-associated E3 ubiquitin ligases and DUBs in primary tumours, including DUBs USP22 and USP36 which get overexpressed in multiple aggressive cancers and bring about rapid breast cancer progression through dysregulation of H2BK120Ub (Glinsky *et al*, 2005; Li *et al*, 2008).

1.9 UBR Ubiquitin Ligase Family

The UBR protein family is a group of E3 ubiquitin ligases which share the evolutionally conserved UBR box domain among themselves. Seven members of this group has yet been identified in mammals - UBR1 to UBR7. UBR1 (ubiquitin–protein ligase E3 component N-recognin 1) was the first member of the family to be identified whose molecular weight is around 225 kDa. It was reported to act as an N-recognin (recognises and binds N-degrons) in yeast (Hochstrasser & Varshavsky, 1990). Later, all seven UBR box-containing proteins were identified in mammals. Apart from the UBR box, the mammalian UBR members contain various domains and motifs generally associated with other E3 ubiquitin ligases and hence were classified as E3 ligases (Fig. 1.9.1A).

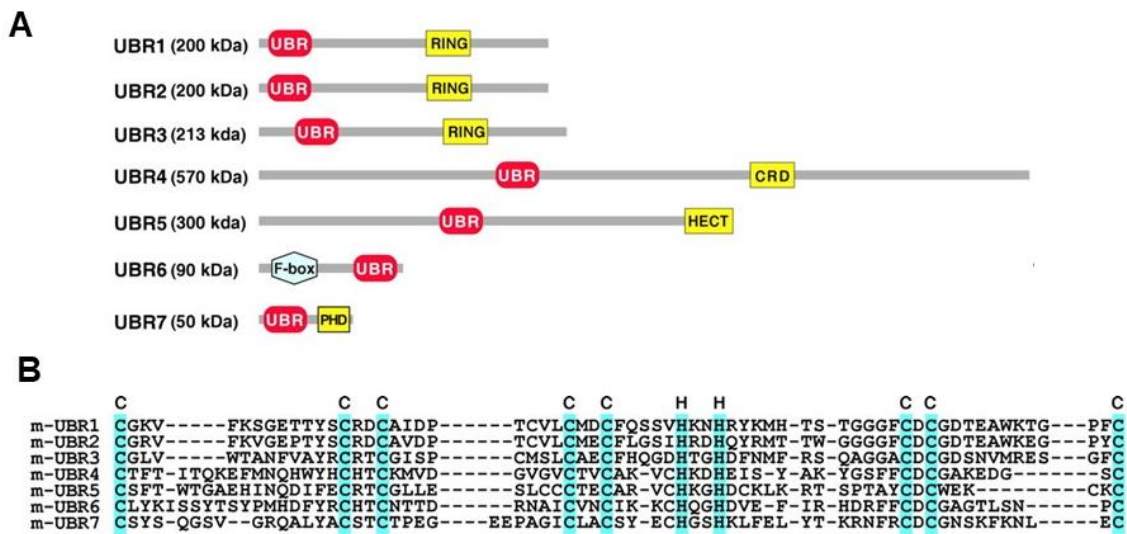


Fig. 1.9.1 The UBR protein family. (A) Domain architecture of the members of UBR protein family with their signature domains highlighted. (B) Multiple sequence alignment showing conservation of zinc coordinating residues within UBR box shared by the different members of UBR protein family in mammals. (Adapted from Tasaki *et al*, 2005)

The shared UBR domain where the zinc coordinating residues remain conserved among the different members of the family, has been reported to be involved in substrate recognition in certain cases (Tasaki *et al*, 2005) (Fig 1.9.1B).

In the Ubiquitin Proteasome pathway (UPS), N-degrons refer to the degradation signals whose reside within the sequence of the N-terminal residues of certain cellular proteins. N-degrons are characterized by adjoining sequence motifs and internal lysine residues which have the potential to get polyubiquitinated. N-degrons are classified mostly into 2 types – Type I and II. Type I N-degrons can occur as primary destabilizing residues, which are directly recognized by N-recognins and composed of positively charged amino acids, or as secondary and tertiary destabilizing amino acids, which has to be conjugated to a primary destabilizing residue. On the other hand, type II destabilizing residues are mostly composed of bulky hydrophobic amino acid residues. N-degron-binding UBR box proteins are called N-recognins and UBR box of UBR1, UBR2, UBR4, and UBR5 has been shown to bind type I N-degrons. UBR1 and UBR2 can also bind type II N-degrons through their corresponding N domains. (Tasaki *et al*, 2009).

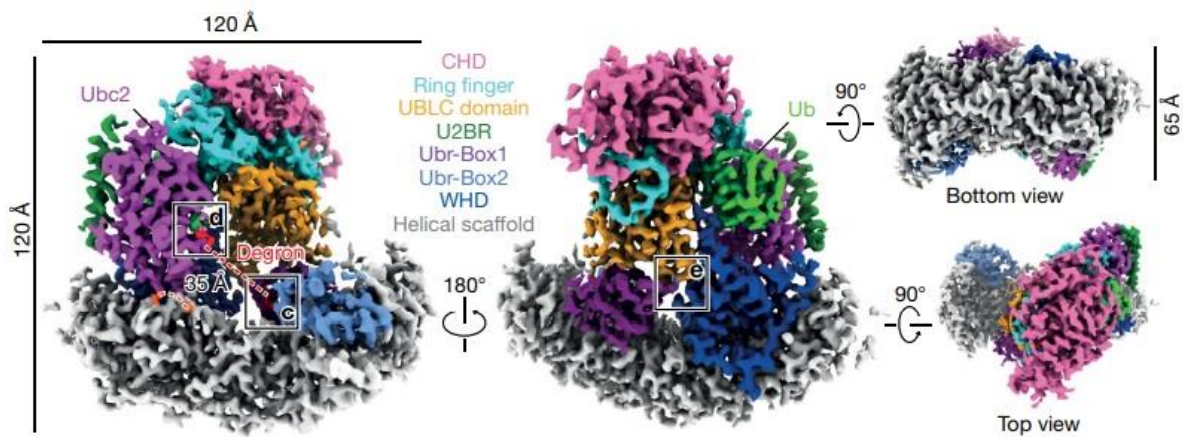


Fig. 1.9.2 Cryo-electron structure of the initiation complex comprising of UBR1, Ub, Ubc2 and degron peptide. The UBR box 1 and 2 and the RING are highlighted in violet, royal blue and cyan respectively. (Adapted from Pan *et al*, 2021; PDB ID 7MEX)

UBR1, UBR2, and UBR3 harbour RING domains, which mediate direct transfer of ubiquitin from the E2 enzyme to the target protein and do not directly interact with ubiquitin itself. They, like most RING E3s possess conserved seven cysteine and a single histidine in between the 3rd and the 4th cysteine residues which eventually coordinate and bind two zinc ions among themselves, thus forming the overall structure (Deshaies & Joazeiro, 2009). Although the RING participates in the recruitment of its partner ubiquitin-conjugating enzyme, its association with E2 is not always robust and this transiency doesn't affect its E3 ligase activity. In case of UBR1, the RING domain by itself has low affinity for its E2 partner Ubc2 (an E2), and the actual E2 binding region is located in the basic region N-terminal to that of UBR1-RING domain (Fig. 1.9.2). However, the RING is still essential for E3 ligase activity of UBR1 (Xie & Varshavsky, 1999). UBR1 has been reported to mediate K-48 linked polyubiquitination of N-degrons in the Arg/N-degron pathway (Pan *et al*, 2021). UBR2, which also harbours a RING finger, has been reported to localize to meiotic chromatin regions to mediate transcriptional silencing by histone H2A ubiquitination. In order to carry out ubiquitin transfer onto H2A, UBR2 interacts with the E2 enzyme HR6B promotes the HR6B–H2A interaction. UBR2 mediated H2A ubiquitination leads to transcriptional silencing during meiosis (An *et al*, 2010).

UBR5 contains the HECT domain which consists of two different lobes, the N-terminal lobe (N-lobe) and C-terminal lobe (C-lobe), connected by a flexible linker. As previously mentioned, N-lobe is responsible for E2 recruitment and binding, whereas the C-lobe

harbours the catalytic cysteine that is involved in ubiquitin transfer. The HECT domain of UBR5 is present in its C-terminal tail (Qian *et al*, 2020). UBR5 has been reported to play a key role in regulation of proliferation, apoptosis and metabolism (Henderson *et al*, 2002). Recent studies connect UBR5 to the ubiquitin–proteasome system (UPS) in cancer and suggest that UBR5 might be involved in driving the proliferation and metastasis of triple-negative breast cancer (Shearer *et al*, 2015; Liao *et al*, 2017; Matsuura *et al*, 2015). Being an E3 ligase, UBR5 contributes to polyubiquitination and degradation of target proteins in many cancers (Zhang *et al*, 2014; Yang *et al*, 2016; Wang *et al*, 2017).

UBR6, also called FBXO11, is distinct from the other members since it carries a F-box and is part of SCF ubiquitin ligase complex consisting of Cullin 1 (CUL1), Skp1, RBX1, and the corresponding F-box protein. The F-box protein is responsible for directly binding to substrates to initiate ubiquitination and subsequent proteasomal degradation (Zheng *et al*, 2002). The F-box domain by itself is a 40-amino-acid region which is involved in Skp1 binding (Bai *et al*, 1996).

Last but not the least, UBR7, the smallest member of the group, contains the plant homeodomain (PHD) finger, a 50–80 amino acid protein domain which is mainly associated with binding of methylated and acetylated histones. Proteins with this domain have been extensively reported to regulate gene expression within the nucleus. However, in spite of not having any signature domains which are commonly associated with E3 ligase activity, UBR7 PHD has been recently demonstrated to monoubiquitinate histone H2B at lysine 120 (H2BK120Ub). The loss of UBR7 was also found to correlate with the development of triple-negative breast cancer and metastatic tumours (Adhikary *et al*, 2019). Apart from H2B, UBR7 has also been shown to associate abundantly with histone H3.1 (Campos *et al*, 2015). Within the eukaryotic nucleus, UBR7 has been shown to play an important role in re-incorporation of NASP-bound post-nucleosomal histone H3 (Hogan *et al*, 2022). Although UBR7 has a conserved UBR box (44–116), it somehow demonstrates a rather uncharacteristic behaviour and does not recognise and bind N-degrons (Tasaki *et al*, 2009). UBR7 has also been implicated in nucleotide biosynthesis and identified as strong interactors of phosphoribosyl pyrophosphate synthetases (PRPSs). In this regard, UBR7 protects the PRPS catalytic subunits by promoting polyubiquitination of PRPS-associated protein (PRPSAP) which acts as a negative regulator of PRPS (Srivastava *et al*, 2021). UBR7 also

interacts with tumor-suppressor TP53 in breast cancer cells, resulting in p53 transcriptional activation and cell death (Isobe *et al*, 2020).

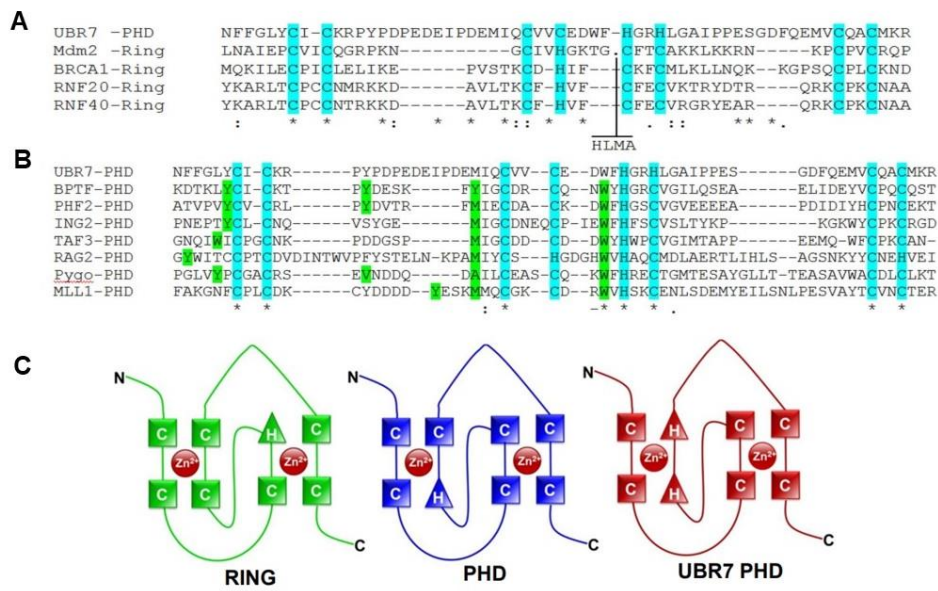


Fig. 1.9.3 UBR7 PHD finger is distinct from other canonical RING and PHD fingers. (A,B) Multiple Sequence alignment of UBR7 PHD with (A) other RING fingers (B) and canonical PHD fingers. (C) Cross brace topology depicting the differential arrangement of zinc coordinating residues in canonical RING, PHD and UBR7 PHD. (Adapted from Adhikary *et al*, 2019)

Despite functioning similar to a classic RING finger, UBR7 PHD has a distinct arrangement of zinc coordinating residues which is distinguishable among other RING and PHD fingers. When compared to the Cys3-His-Cys4 arrangement of the RING fingers, or the Cys4-His-Cys3 arrangement of the PHD fingers, the atypical PHD finger of UBR7 has a Cys4-His2-Cys2 arrangement whereby both the histidine residues are engaged with the first zinc atom. It is thus very interesting that UBR7, in absence of any significant size or sequence similarity with its fellow family members, and lacking any canonical E3 ligase domain, still plays the role of a ubiquitin ligase module and catalyse degradation-independent monoubiquitination of its substrate as compared to other UBR E3s which generally mediate polyubiquitination and proteasomal degradation (Adhikary *et al*, 2019).

Chapter 2

Objectives

Background

UBR7 is the smallest member of the UBR protein family of E3 ligases, but it distinguishes itself from the other members in two major aspects – firstly, it doesn't have any signature domains associated with E3 ligase activity, and secondly, it doesn't bind to N-degrons like the other UBR proteins. The functional role of UBR7 was relatively unknown when a collaborative study involving our lab reported for the first time that UBR7 catalyzes H2BK120 monoubiquitination and plays a tumour suppressor role in countering proliferation, invasion, epithelial-to-mesenchymal transition (EMT) and metastasis in triple-negative breast cancer (TNBC) (Adhikary *et al*, 2019). The study also identified the E2 conjugating enzyme that was responsible for UBR7 mediated H2B ubiquitination to be UbcH6. UbcH6 (also called E2E1) has previously been reported as the common E2 enzyme in both H2A and H2B monoubiquitination (Zhu *et al*, 2005; Wheaton *et al*, 2017). However, the mechanism of UBR7 mediated ubiquitin transfer was still not well understood.

TNBC is one of the most aggressive forms of breast cancer with a poor prognosis than other forms of cancer (Onitilo *et al*, 2009; Dent *et al*, 2007, Haffty *et al*, 2006). Currently, there is no efficient targeted therapy for TNBC (Vagia *et al*, 2020), which emphasizes the dire need to understand the mechanism of UBR7 function and its association with histones in the chromatin context. Here, we intended to characterize the association of UBR7 with chromatin by employing cell-based, biophysical, and biochemical approaches and thus tried to elucidate the molecular mechanism behind the E3 ligase function of UBR7.

Objectives of the Present Study

Objective 1: Preparation of E2-Ub thioester

- 1a. Expression and purification of Ubiquitin, Ube1(E1), and UbcH6(E2).
- 1b. Ubiquitin loading of UbcH6 in presence of Ube1.
- 1c. Purification of UbcH6~Ub using size exclusion chromatography.

Objective 2: Identification and characterization of the E2-E3 interaction interface between E3 ligase UBR7 and E2 conjugate UbcH6

- 2a. Identification of the region of UBR7 involved in E2 binding by *in vitro* and *ex vivo* interaction studies.
- 2b. Determination of the role of the E2 interacting region of UBR7 concerning its ubiquitin-binding and E3 ligase activity when compared to the UBR7 full-length protein.
- 2c. Generation of E2 mutants spanning mutations on probable E3 binding residues.
- 2d. To ensure that the mutations designed in the E2 UbcH6 don't disrupt its folding, E1 interaction, and ubiquitin loading ability.
- 2e. Mapping of the region of UbcH6 involved in UBR7 binding by comparing the E3 binding ability of the mutants with UbcH6 wildtype through protein-protein interaction studies.
- 2f. Determination of the role of the residues encompassing the E3 binding surface of UbcH6 in H2BK120 monoubiquitination.

Objective 3: Characterization of H2B monoubiquitination by UBR7-UbcH6 complex

- 3a. Characterization of the ubiquitin transfer rate of UBR7-UbcH6 complex onto substrate histone H2B.
- 3b. Identification of the minimum stretch of substrate required for UBR7 mediated ubiquitin transfer using histone peptides.
- 3c. Finding the affinity of UBR7 PHD towards its substrate histone H2B using Fluorescence spectroscopy.
- 3d. Detection of monoubiquitinated H2B species using antibody-independent mass spectrometric analysis.

Objective 4: Determination of oligomeric status of UBR7 and its role in E3 ligase function

- 4a. To identify the region of UBR7 involved in self-association through cell-based expression and co-immunoprecipitation study.

- 4b. Prediction of UBR7 structure using molecular modelling technique.
- 4c. Confirmation of oligomeric status of UBR7 using the chemical crosslinking assay.
- 4d. Determination of the molar mass of UBR7 in solution using Size-exclusion chromatography coupled with multiple angle light scattering (SEC-MALS)
- 4e. Identification of possible residues involved in oligomerization and exploring their role in E3 ligase function, E2 and substrate binding.

Objective 5: Comparison of the mechanism of UBR7 mediated ubiquitin transfer with other H2BK120 monoubiquitin ligases.

- 5a. Exploring the potential of ubiquitin discharge of UBR7 PHD finger and RNF20 RING finger from E2-Ub thioester in the presence and absence of substrate.
- 5b. Detection of conformational alteration in UBR7 PHD upon interaction with substrate H2B by Circular dichroism studies.

Chapter 3

Materials and Methods

3.1 Chemically Competent Cell Preparation

Superior quality competent cells are required for transformation during cloning and sub-cloning, which involves the imbibition of extracellular plasmid DNA into the cell. To make the bacterial cells chemically competent, they are treated with a high concentration of divalent cations (like Ca^{2+} and Mg^{2+}). This enables the cell wall and cell membrane of bacteria to become permeable for uptake of DNA. In our study, we used different strains of chemically competent *E. coli* cells for sub-cloning or DNA propagation and then subsequent expression, either DH5 α , or BL21 StarTM (DE3) (both from Invitrogen), and RosettaTM (Novagen). Briefly, LB agar plates without any antibiotics (for DH5 α and BL21) or with Chloramphenicol (for Rosetta) were streaked with previous stocks of corresponding competent cell strains to obtain single colonies and incubated at 37° C overnight. The next afternoon, a single colony was picked and seeded into 10ml Luria Bertani (LB) broth medium overnight at 37° C with shaking at 180 rpm, as pre-culture. On the following morning, 1ml of overnight culture was inoculated into 200 ml LB medium, which was then grown at 37° C with shaking, till optical density (O.D.) at 600 nm reached a range of 0.3 - 0.4. At this point, further bacterial growth was arrested by incubating the culture flask in ice and occasional swirling. All the subsequent steps of this protocol were strictly performed on ice or 4° C. The cells were harvested by centrifuging at 6000 rpm for 15 minutes, followed by washing with chilled 100 mM Magnesium Chloride (MgCl_2) followed by one more round of centrifuging. The cell pellets were this time resuspended in chilled 100 mM Calcium Chloride (CaCl_2) and incubated on ice for 30 minutes. The cell suspension was again spun down and gently dissolved in an ice-cold solution of 75 mM CaCl_2 and 20% glycerol. The cell suspension was further aliquoted in 1.5 ml microtubes, flash-frozen in liquid nitrogen, and stored at -80° C. The freshly prepared competent cells were later tested for their transformation competency and/or presence of contamination if any before being used for transformation in sub-cloning or expression.

3.2 Cloning of Expression Constructs and Site-Directed Mutagenesis (SDM)

Ubiquitin (#10861), Ube1 (#34965), and UbcH6 (#15787) cloned in bacterial expression constructs were purchased from Addgene. GST-tagged RNF20 RING construct (in pGEX-6p-1 vector) and recombinant human histone H2A and H2B constructs (in pET 28A vector) were kindly gifted by Dr. Catherine L. Day (University of Otago, Dunedin) and Dr. Shin-ichi Tate

(Hiroshima University) respectively. GST tagged UBR7-full length (1–425) and the UBR domain (41-123) cloned in Gateway vector pDEST15 and EGFP tagged UBR7- full length cloned in Gateway vector pcdDNA-EGFP were available in the lab from a previous study. UBR7-full length, UBR7 Δ UBR (Δ 44-116), and UBR7 Δ PHD (Δ 132-188) were cloned in mammalian vector pcDNA3.1+/C-(K)DYK for use in cell-based overexpression and immune-pulldown experiments.

Table 3.1 Summary of cloned bacterial constructs used in this study

Gene cloned	Vector	Restriction enzymes used	Primers (Forward and Reverse)
UbcH6 core domain (39–193)	pET28A	NdeI/XhoI	F: ATATCATATGAGCAAAAACTCCAAACTCCTCT R: TATACTCGAGTTATGTAGCGTATCTCTTGGTCCA
UBR7 PHD (132–203)	pGEX6P-1	BamHI/XhoI	F: ATATGGATCCGGCCTGTATTGCATTTGTAAAC R: TATACTCGAGTTATTATTTGGTAACAGCCAGCTGAG
UBR7 CTD (325–425)	pGEX6P-1	BamHI/XhoI	F: ATATGGATCCGACCTGGATGTGCTGTTTCCT R: TATACTCGAGGCTGCAATAGTATTGCATGCC

The different cloned constructs used in this work were cloned mostly using restriction enzyme digestion and T4 DNA ligase mediated ligation strategy. Gene sequences from existing clones were PCR amplified using Taq Polymerase (Thermo) and the obtained PCR product was run on a 0.8-1% Agarose gel and purified using Gel Elution Kit (Qiagen). The PCR products and vectors were then digested using the same set of restriction enzymes and subsequently incubated together in presence of T4 DNA Ligase (NEB) followed by transformation. In case of cloning of the UBR7-PHD domain and its triple mutant (E145K/E151A/R195A) into pETite N-His SUMO vector (Lucigen), a ligase-independent strategy was employed as the purified insert DNA was directly added to the N-His-SUMO- KAN vector in absence of any ligase enzyme and the entire reaction mixture was transformed.

Transformation of ligated products (in case of cloning) or already cloned plasmid (in case of expression) was performed using chemically competent cells. Briefly, 50-100ng of plasmid DNA or the entire 20 μ l of ligated product was added to the competent cell suspension and incubated on ice for 30 minutes. Heat shock was applied for 90 seconds at 42°C followed by cold shock on ice for the next 5 minutes. The cells were next allowed to revive by adding 1ml of LB broth media and incubated for 1hr at 37°C with constant shaking at 180 rpm. The cells

were spun down at 4000 rpm for 5 minutes at room temperature (RT). The resuspended pellet was plated on respective antibiotic-containing LB agar plates. The agar plates were kept at 37° C overnight to obtain appropriate bacterial colonies.

Table 3.2 Summary of mutants generated in this study

Parent construct	Mutations Generated	Primers
		Forward primer (F) and Reverse Primer (R)
UbcH6 core pET28A	Δ107-111	F: TTGTCCGAAATGTAACTTATATTCTGGTGTAAGGTGATATCGAGAAA R: TTTCTCGATATCACTTTTACACCAGAATATAAGGTTACATTTCGGACAA
UbcH6 core pET28A	Δ138-142	F: AAGGAGGACTTTAGAAATGGTTAGATCTTCAATATGTCCAAGCAAAT R: ATTTGCTTGGACATATTGAAAGATCTAACCATTTCTAAAGTCCTCCTT
UBR7 PHD pGEX 6P-1	E145K	F: CGGAATTTTCGTCTTTCGGGTCCGGGTACG R: CGTACCCGGACCCGAAAGACGAAATTCCG
UBR7 PHD pGEX 6P-1	E151A	F: CGACGCACTGGATCATTGCGTCCGGAATTTTCTTT R: AAAGAAAATTCCGGACGCAATGATCCAGTGCGTCG
UBR7 PHD pGEX 6P-1	R165A	F: GGATTGCACCCAGGTGAGCGCCATGAAACCAATCTT R: AAGATTGGTTTCATGGCGCTCACCTGGGTGCAATCC

Point mutations on existing clones were introduced by Stratagene QuikChange site-directed mutagenesis kit (Liu & Naismith, 2008). The PCR products containing the targeted mutations were obtained by setting up PCR reactions using PfuUltra High-Fidelity DNA Polymerase. The template DNA was digested using Dpn1 restriction endonuclease that specifically cleaves methylated and hemimethylated parental plasmid DNA. The digested product was then transformed using ultra-competent DH5α cells. The cloned constructs and their corresponding mutants were sequenced and confirmed.

3.3 Recombinant Protein Expression and Purification

The plasmid DNA constructs were transformed into Escherichia coli BL21 (DE3) or Rosetta, while being incubated at 37° C overnight. The bacterial colonies were flushed into a pre-culture and grown for 1-2 hrs. Next, it was added to larger LB media volumes in the baffled flasks and

grown until an O.D. of 0.6 was achieved which was measured using a spectrophotometer at 600 nm, followed addition of 1 mM Isopropyl β -D-1-thiogalactopyranoside (IPTG) for induction of recombinant proteins at 18 °C for 18 hours at 180 rpm shaking. During the expression of zinc finger proteins, 50 μ M ZnCl₂ was added with IPTG. The next day, cultures were centrifuged at 6000 rpm for 10 minutes to obtain cell pellets, which were either subjected to purification steps on the same day, or flash-frozen and stored at - 80° C for future use.

In case of GST-tagged proteins, bacterial cell pellets were resuspended in lysis buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM sodium chloride (NaCl), 5% glycerol, 2 mM Dithiothreitol (DTT) and 0.05% Nonidet-40 (NP-40) detergent and sonicated until the texture of the suspension turned darker and translucent. The sonicated suspension was next centrifuged at 28000 rpm for 45 min at 4 °C using a Sorvall RC-5C Plus ultracentrifuge. The supernatant obtained was filtered with a 0.22 μ m filter, and subsequently incubated with pre-equilibrated glutathione Sepharose resin (GE Healthcare) in a rotor-mixer for 2 hrs at 4 °C. Following bead binding, the beads were washed with the wash buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5% glycerol, and 2 mM DTT. Bound proteins were eluted with 25 mM reduced glutathione containing buffer and 1 mM phenylmethylsulphonyl fluoride (PMSF) further added to obtain GST-tagged proteins. Alternatively, proteins were cleaved using PreScission protease (GE Healthcare) to obtain untagged proteins.

GST-tagged ubiquitin, which is cloned in a pGEX 2TK vector, has a thrombin cleavage site in between the GST tag and the protein. Thrombin enzyme was used to obtain untagged ubiquitin which was used for enzymatic assays and MALDI-TOF experiments. Thrombin cleavage was achieved using 1X PBS buffer containing 140mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, and 1.8mM KH₂PO₄, pH 7.3, and the cleavage reaction was allowed to proceed at RT for 16 hrs using mild shaking. No reducing agent was added to any buffers during the purification of ubiquitin as it reduces thrombin cleavage activity.

In case of purification of His-tagged proteins, pellets were resuspended in buffer containing 20 mM Tris-HCl (pH 8.0), 25 mM imidazole, 350-400 mM NaCl, 2 mM 2-mercaptoethanol (β -Me), 5% glycerol, and 0.05% NP-40. Similar to GST protein purification, the lysed suspension was sonicated and subjected to ultracentrifugation at 28000 rpm for 45 min at 4 °C. The soluble fraction was then allowed to bind to TALON resin (Clontech) for 2 hours at 4 °C. The protein-bound beads were subjected to repeated washing steps with buffer containing 500 mM NaCl.

The His-tagged proteins were further eluted using with freshly made elution buffer containing 300 mM imidazole (pH 8.5). Desalting of eluted proteins was achieved using Hi-Prep 26/10 Desalting column (GE Healthcare) to remove imidazole.

The purification of His-Ube1 involved one further step of anion exchange chromatography using a 5-mL HiTrap Q HP column (GE Healthcare). Untagged proteins cut from their GST tag were subjected to size exclusion chromatography using HiLoad Superdex 75 pg column connected to an ÄKTA pure chromatography system (Both GE Healthcare). Eluted protein fractions were confirmed on SDS, blotted and detected by specific antibodies. Then they were pooled together to be concentrated using Amicon Ultra Centrifugal Filter units (Millipore) of appropriate cut-offs.

3.4 Purification and Reconstitution of H2A-H2B Dimers

His-tagged histones when expressed individually were found in inclusion bodies after sonication. Thereafter, histone purification from inclusion bodies was done as described previously (Dyer *et al*, 2004). Briefly, the sonicated pellet was suspended in an ice-cold inclusion body wash buffer containing 6M urea. The inclusion body suspension was then centrifuged at 18,000 rpm for 30 min. The supernatant was made to bind to TALON beads and later eluted with buffer containing 300 mM imidazole along with 6M urea. Eluted histones H2A and H2B were mixed in equimolar ratios and imidazole was removed by dialyzing the protein against a buffer lacking imidazole (20 mM Tris-HCl pH 8.0, 6M Urea, 2 M NaCl, 5% glycerol, 1 mM DTT) for 4 hrs in RT followed by a buffer change and another 4 hrs at 4 °C. Refolding and reconstitution of dimers were subsequently achieved by dialyzing against 20 mM Tris-HCl pH 8.0, 2 M NaCl, 5% glycerol, 1 mM DTT, 1 mM EDTA and 1 mM PMSF overnight. Histone refolding was confirmed using CD spectroscopy. Three dialysis steps were performed back-to-back to ensure proper refolding of histones. The dialyzed histones were aliquoted, flash-frozen, and stored at - 80 °C.

3.5 E2 Charging Assay and Purification of UbcH6~Ub Thioester

E2~Ub thioester was prepared using activated ubiquitin in presence of Ube1 (E1) and UbcH6 core domain (39-193) in the absence of any E3 or substrate. UbcH6 core domain is hereafter referred to as UbcH6c. The charging reaction contained 75 μ M UbcH6c, 150 μ M ubiquitin and 0.3 μ M His-Ube1 (E1) in a charging reaction buffer containing 50 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 3 mM ATP and 2 mM DTT. The reaction was allowed to go on for 16 hours following which the UbcH6~Ub thioester species was purified using size-exclusion chromatography. The eluted fractions from a Superdex 75 pg column (GE Healthcare) corresponding to the UbcH6~Ub peak were pooled, concentrated, and stored at - 80 °C for future use.

For the time-point E2 charging experiment involving UbcH6c mutants, the reactions were stopped at different time points – 1 hr, 8 hr, and 16 hr to explore the rate of their thioester forming ability.

3.6 *In vitro* Ubiquitination Assay

Ubiquitination assays were performed as described previously (Wheaton *et al*, 2017; Plechanovova *et al*, 2012). In a 25- μ L volume reaction mixture, His-Ube1 (0.3 μ M), His-UbcH6c (15 μ M), ubiquitin (35 μ M), GST tagged UBR7-PHD (0.15 μ M) WT or mutants and H2A/H2B dimer (5 μ M) were incubated together in 50 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 2mM ATP and 2 mM DTT. The reaction was allowed to proceed at 30 °C for 30 min following which, the reaction was quenched by adding 5 μ L of 5 \times Laemmli buffer and thereafter loaded onto a 15% SDS/PAGE gel for separation and western blot. Ubiquitinated proteins were detected after western blot using anti-H2BK120Ub antibody. Substrate H2B was detected with α -H2B antibody while α -GST and α -His horseradish peroxidase (HRP) antibodies were used to probe for E3s and E2s respectively. When H2B peptides were used as substrate in the ubiquitination assay, the assay time was increased to 6 hours.

The single-turnover substrate assay was performed using the purified E2~Ub thioester. 0.15 μ M UBR7-PHD was added to 10 μ M of UbcH6c~Ub and 5 μ M H2A/H2B dimer in ubiquitination buffer containing 50 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$ and 2 mM DTT. The reactions were stopped at specific time points (0, 2, 4, 6, 8, 10, 15, 20 minutes) and the eventual

product formation was analyzed by western blotting. The experiment was done in three replicates and the intensity of product formation in each case was noted. Due to the overlapping molecular weight of GST and H2BK120Ub species, separate gels were run for blotting H2BK120Ub and GST tagged E3s. The blots were subsequently quantified using ImageJ and the average intensity of product (H2BK120Ub) formation corresponding to every time point was plotted in the product vs time plot and fitted using Hill's equation for dose-dependent ligand binding.

Sequential competitive pulldown and ubiquitination assay involved the sequential addition of UbcH6 (E2) and H2B (substrate) to E3 ligase UBR7-PHD in a different order and intermittent incubation. The addition and incubation steps in between were orchestrated in the following combinations –

1. E3 + E2 + substrate incubated together followed by E1 and Ub, for the reaction to proceed.
2. (E3 + E2) incubated together followed by the introduction of the substrate, E1, and Ub.
3. (E3 + substrate) incubated together followed by the introduction of E2, E1, and Ub.

The different reactions were quenched by Laemmli buffer and thereafter loaded onto a 15% SDS/PAGE gel for separation and western blot analysis. Ubiquitinated H2B species in each case was detected and compared.

3.7 Thioester Hydrolysis Assay

Pre-purified UbcH6c~Ub thioesters were used in the single-turnover thioester hydrolysis assay which was performed as described previously (Foglizzo *et al*, 2016). Briefly, 10 μ M UbcH6c~Ub thioester was incubated with GST-tagged E3 enzymes – namely UBR7-PHD or RNF20 RING (2 μ M each) in the presence or absence of H2B C-terminal peptides as substrate. The reaction was set at RT and was allowed to proceed up to certain time points (0, 0.5, 1, 2, and 4 hours). The reactions were quenched by 5 μ L of 5 \times Laemmli buffer and thereafter loaded onto a 15% SDS/PAGE gel loaded onto a 15% SDS gel, followed by a western blot. The rate of ubiquitin discharge was explored using an α -ubiquitin antibody.

3.8 GST Pull-down Assay

GST pull-down assays were performed according to standard protocols (Busschots *et al*, 2005; Malovannaya *et al*, 2011). Briefly, GST-tagged proteins were added as per equimolar ratio to its potential interacting partner protein and incubated together in pulldown buffer (50 mM Tris-HCl, pH 7.5, 1 mM DTT, 0.1% NP-40, 150 mM NaCl) at 4 °C overnight. Glutathione Sepharose beads pre-blocked with a binding buffer containing 5% BSA were subsequently added and incubated for 2 hours. The protein-bound beads were washed thrice and Laemmli buffer was added. The samples were then run on 15% SDS PAGE and further subjected to western blot analysis.

3.9 Hexa-histidine (6xHis) Pull-down Assay

For hexahistidine tagged protein pull-down, FLAG-UBR7 WT or corresponding deletion constructs were overexpressed in HeLa cells. The cell suspension was spun down and resuspended with lysis buffer containing 50mM HEPES, pH 7.4, 150 mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 1% Nonidet P-40, 0.5% Sodium Deoxycholate, and 1X proteinase inhibitor cocktail (Takara). The resuspended lysate mixture was briefly sonicated, clarified using high-speed centrifugation, and estimated. His-tagged UbcH6- core domain (39-193) was extramurally added to the cell lysate and incubated at 4 °C overnight. This was followed by nickel-nitrilotriacetic acid (Ni-NTA) pull-down. Subsequently, the beads were washed thrice with 25 mM Tris-HCl, pH 7.4, 20 mM imidazole, 150 mM NaCl and 0.1% NP-40, run on SDS PAGE and later western blotted. Twenty percent of inputs were loaded and indicated in the figure.

The Ni-NTA pull-down assays with UbcH6c and untagged PHD WT and mutant proteins were performed similarly.

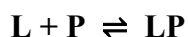
3.10 Peptide Pull-down Assay

Peptide pull-down assays were performed according to the previously established protocols (Wy *et al*, 2006). The biotinylated histone H2B C-terminal peptides (114–125) wild type and K120R mutant used in this study were synthesized and procured from GL Biochem Ltd. In

brief, GST-tagged UBR7-PHD was added to equimolar biotinylated H2B peptides in a pulldown assay buffer containing 50 mM Tris-HCl, pH 7.5, 0.05% NP-40, 150 mM NaCl, 1 mM DTT at 4 °C, overnight. The reactions were then allowed to bind streptavidin sepharose beads (GE Healthcare) and washed three times with the same assay buffer. The beads were then prepared and loaded onto a 12% SDS PAGE gel and subjected to western blot analysis.

3.11 Fluorescence Spectroscopy

The fluorescence measurements were performed in an RF-5301 PC spectrofluorometer (Shimadzu) at 25 °C. Since histone H2A and H2B C-terminal peptides do not have any tryptophan residues, hence they were synthesized from GL Biochem with a FAM fluorophore-conjugated at the 5' end. Fluorescein amidite (FAM) is a water-soluble fluorophore which when excited at 494 nm, exhibits an emission maximum of 520 nm. The fluorescence binding experiments were performed using 0.3 µM of the FAM-conjugated histone H2A/H2B peptides dissolved in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl and 1 mM β-Me and quenching was monitored by adding increasing concentration of UBR7- PHD finger protein. Samples were excited at 494 nm and emission was recorded at 520 nm with a scan speed of 200 nm/s. The data were recorded in triplicates and then averaged. The data analysis was performed according to a previously described method (Adhikary *et al*, 2016; Chakrabarti *et al*, 2000).



From the above state of equilibrium of protein-ligand binding, UBR7- PHD is the ligand (L) in this case while FAM-conjugated H2B peptide is the protein (P). Non-linear curve fitting analysis (Eq. 1) was used here to determine the apparent dissociation constant (K_d). The experimental data points of the obtained isotherms were fitted by least-squares analysis as shown in Eq. 1:

$$C_0 (\Delta F/\Delta F_{\max})^2 - (C_0 + C_p + K_d) (\Delta F/\Delta F_{\max}) + C_p = 0 \quad (\text{Eq. 1})$$

where change in emission intensity for each point of the titration curve is denoted by F while F_{\max} is the change in emission intensity when the histone peptide is saturated by the ligand (UBR7-PHD) and there is no further change in emission intensity upon further addition of the ligand. C_p is the ligand concentration and C_0 is the initial concentration of the histone peptide.

F_{\max} was determined from the double-reciprocal plot:

$$1/\Delta F = 1/\Delta F_{\max} + K_d [\Delta F_{\max} (C_p - C_0)] \quad (\text{Eq. 2})$$

3.12 SEC-MALS

Purified untagged UBR7-PHD WT or mutant proteins were used to determine their individual molecular mass using size exclusion chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) experiments. The experiments were performed using an AKTA-MALS system (GE Healthcare). A Superdex 200 Increase 10/300 GL column (GE Healthcare) was pre-equilibrated in 10 mM Tris (pH 7.5) and 200 mM NaCl at a flow rate of 0.25 mL/min overnight. About 200 μg of purified protein was injected in each case and a DAWN system (Wyatt Technology) was used to analyze the light scattering. The concentrations for protein run were determined using an Opti-lab differential refractometer (Wyatt Technology). The experimentally obtained molar masses of proteins were calculated using ASTRA, version 7.3.2 (Wyatt Technology), and compared to the expected molar mass.

3.13 DFDNB-mediated Cross-linking Assay

1,5-difluoro-2,4-dinitrobenzene (DFDNB) is a chemical cross-linker containing two reactive fluorine atoms that couples with amine-containing molecules, resulting in the formation of stable arylamine bonds. DFDNB is soluble in dimethyl sulfoxide (DMSO), acetone, and most other water-miscible organic solvents. It has a molecular weight of 204.09 and a short spacer arm of 3.0Å which is especially useful for intramolecular crosslinking.

Prescission protease cleaved UBR7-PHD wild type (WT) and triple mutant proteins were purified in Hepes-Na (pH 7.5) buffer. The working stock of DFDNB (Thermo Scientific) was prepared at a concentration of 500 μM by dissolving in DMSO. 5 μg of UBR7-PHD WT and mutant protein was taken for crosslinking, in crosslinking buffer containing 20 mM Hepes-Na (pH 7.5) and 150 mM NaCl. DFDNB was added to make a final concentration of 25 (1 \times) and 50 (2 \times) μM and the reaction mixture was incubated at RT for 30 minutes. The crosslinking reaction was finally quenched by adding 5 μL of 5 \times Laemmli buffer and separated on a 15%

SDS/ PAGE gel. The gel was stained using Coomassie Brilliant blue and visualized through the iBright™ FL1500 imaging system (Thermo).

3.14 Circular Dichroism Spectroscopy

A J-815 spectropolarimeter (Jasco) was used to record the circular dichroism (CD) spectra which was equipped with a Peltier temperature control unit set to 25 °C. Before recording the CD spectra, UBR7- PHD or RNF20 RING were added to H2B C-terminal (114–125) peptide in the ratio of 1: 1 or 1: 2 and incubated on ice for 2 hours. Far-UV CD spectra of the buffer were recorded and subtracted from all the subsequent CD spectra. The CD spectra of proteins in the presence or absence of the peptides were recorded from 195 to 250 nm at RT using a 1-mm path quartz cuvette, a resolution of 0.2 nm, and a scan rate of 50 nm/min. Three consecutive scans were performed for each of the spectra and averaged. CD spectra of UbcH6c wild type and its loop deletion mutants were recorded similarly.

3.15 Cell Culture

Complete aseptic conditions were maintained inside the cell culture room to avoid contaminations. Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% FBS (Gibco, Amarillo, TX, USA), 1% antibiotic-antimycotic (Gibco), and 1% essential amino acid was used to grow and maintain HeLa cells at 37 °C in a Steri-Cycle™ (Thermo) incubator supplied with 5% CO₂.

3.16 Gene Overexpression

EGFP-UBR7 full-length construct were individually or co-transfected with FLAG-UBR7 full-length or the deletion constructs in HeLa cells using Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) following the manufacturer's instruction using 60 mm culture dishes. Cells were incubated at 37 °C with a supply of 5% CO₂ for 24 hours. Following 24 h of transfection, cells were harvested for further experiments.

3.17 Co-immunoprecipitation and Western Blot Analysis

After harvesting the over-expressed cells, whole-cell lysates were prepared for co-immunoprecipitation assay, following previously standardized protocols (Mondal *et al*, 2020). Briefly, the cells were kept on ice and resuspended within the cell lysis buffer containing 50 mM Hepes (pH 7.5), 150 mM NaCl, 1.5 mM MgCl₂, 1mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.5% sodium deoxycholate, 5% glycerol, 1 mM DTT and Protease inhibitor cocktail (Takara). Following resuspension, they were incubated on ice for 1 hour and subsequently sonicated briefly. After centrifugation at 13000 rpm for 10 min at 4 °C, the lysates were incubated with normal sheep serum for pre-clearing. The serum-treated lysate was then incubated overnight with the EGFP antibody (GenScript) followed by incubating with Protein A Sepharose beads (GE Healthcare). The beads were then subjected to extensive washing steps with the same binding buffer and samples were run on 8 or 12% SDS/PAGE for western blot analysis.

For the efficient transfer of proteins, an immobilon nitrocellulose membrane (Millipore) was used with the transfer buffer containing 20% methanol. The transfer was done for 120 min at 4 °C. Membranes were pre-blocked in 5% BSA and then incubated with specific primary antibodies overnight at 4 °C with gentle shaking. The next morning, blots were washed thrice with Tris-buffered saline containing 0.1% Tween-20 followed by incubation with appropriate secondary horseradish peroxidase (HRP) antibodies (Bethyl) for 1 hour at RT. Blots were developed after three more washing steps with Tris-buffered saline containing 0.1% Tween-20. For imaging of the blots, a C400 ChemiDoc imager (Azure Biosystems) was used while Super Signal West Pico Plus (Thermo Scientific) was used as chemiluminescent western blotting substrate. The antibodies and their corresponding dilutions that were used in this study have been summarised in the table below.

Table 3.3 Summary of antibodies used in this study

Sl. No.	Antibody	Company (Cat. No.)	Dilution used
1.	Anti-Ubiquitylated H2B (H2BK120Ub)	Cell Signaling Technology (#5546)	1:2000
2.	Anti-H2B	Abcam (ab1790)	1:10000

3.	Anti-EGFP	GenScript (#A01704)	1:1000
4.	Anti-ubiquitin	Cell Signaling Technology (#43124)	1:2000
5.	Anti-GST	Bethyl	1:50000
6.	Anti-His HRP	Abcam (ab1187)	1:10000
7.	Anti-FLAG	Abcam (ab236777)	1:2000
8.	Anti-Rabbit IgG-HRP	Sigma (A1949)	1:10000
9.	Anti-Mouse IgG-HRP	Promega (W402B)	1:10000

3.18 Protein Structural Modelling

The crystal structure of the BPTF-PHD finger of NURF (PDB ID: 2F6N), which was one of the first reported structures of PHD fingers was obtained from the RCSB Protein Data Bank and was used as a template for structural modeling of the UBR7-PHD finger. Using the PHYRE2 algorithm, the amino acid sequence of UBR7-PHD was threaded onto the structure of the BPTF-PHD template (Kelley *et al*, 2015). PROQ2 tool was employed to evaluate the quality of the model obtained in PHYRE2.0 (Ray *et al*, 2012). The cartoon representation of the structure was created using PYMOL (Schrodinger). The 2 constituent monomers of the UBR7-PHD dimer were coloured differentially and the amino acid residues located at the dimerization interface were highlighted.

3.19 MALDI-TOF Mass Spectrometry Analysis

In-vitro ubiquitination assays were performed using UBR7-PHD finger as E3 and H2B C-terminal WT and K120 mutant peptides as substrates. Thrombin cleaved ubiquitin was used for the assay. Ubiquitinated samples were spotted in triplicates on an MTP 384 BC target. A saturated solution of CHCA (α -cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 0.1% trifluoroacetic acid was used to carry out mass analysis in a 4700 Proteomics Analyzer with

TOF/ TOF optics (Applied Biosystems) set in reflector mode. The peaks obtained in the range of 8.5–11 kDa were analyzed using PEAK EXPLORER (Agilent Technologies). Prominent peaks were annotated. All the annotated peaks indicating the molecular weight of the ubiquitinated species obtained were subsequently compared to the expected molecular weight of ubiquitin and ubiquitinated peptides to check for the evidence of product formation.

Chapter 4

Results

4.1 Preparation of E2-Ub Thioester

The current study aims to understand the mechanism by which UBR7 catalyzes the E2~Ub thioester bond to transfer the ubiquitin onto the substrate. In any *in vitro* ubiquitination assay system, ubiquitin and the activating enzyme Ube1 are common, irrespective of the E2 and E3 enzymes involved. Since the E2 conjugate partner for UBR7 is already known to be UbcH6, it was important to purify the UbcH6~Ub thioester as a reagent for single turnover ubiquitination assays and ubiquitin discharge assays. This is ideally done in an *in vitro* condition using purified recombinant proteins.

For this purpose, I individually expressed and purified Ubiquitin, Ube1 (E1), and UbcH6 Core Domain (E2) consisting of 39-193 residues for subsequent ubiquitin charging of UbcH6.

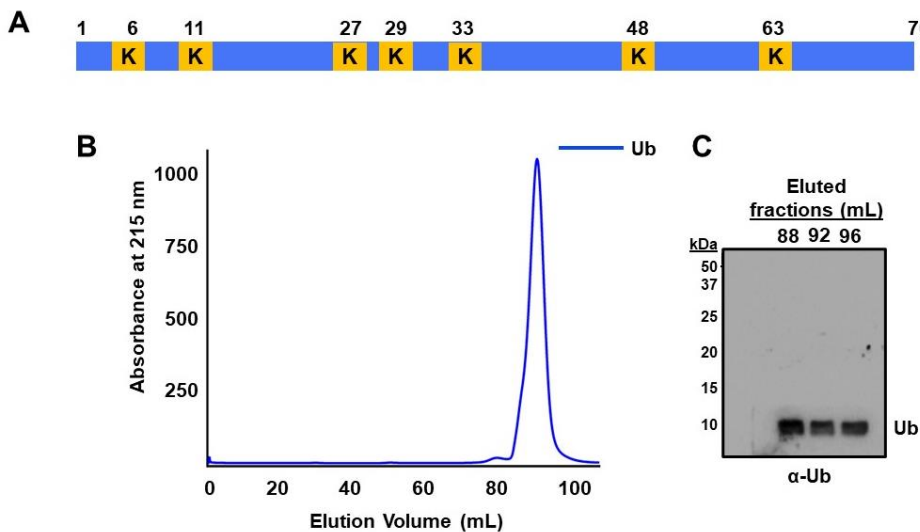


Fig. 4.1 Purification of untagged ubiquitin. (A) Domain architecture of ubiquitin with the seven lysine residues highlighted. (B) Size exclusion chromatogram of untagged ubiquitin as obtained after eluting through Superdex 75 16/60 prep-grade column. (C) Western blot of eluted fractions, probed by anti-ubiquitin antibody.

GST-tagged ubiquitin was expressed and cleaved with thrombin overnight to obtain untagged ubiquitin (9.2 kDa). After performing size exclusion chromatography, the eluted fractions were run on a 15% SDS PAGE followed by a western blot. Purified ubiquitin protein was detected by probing with anti-ubiquitin antibody (Fig. 4.1B and C).

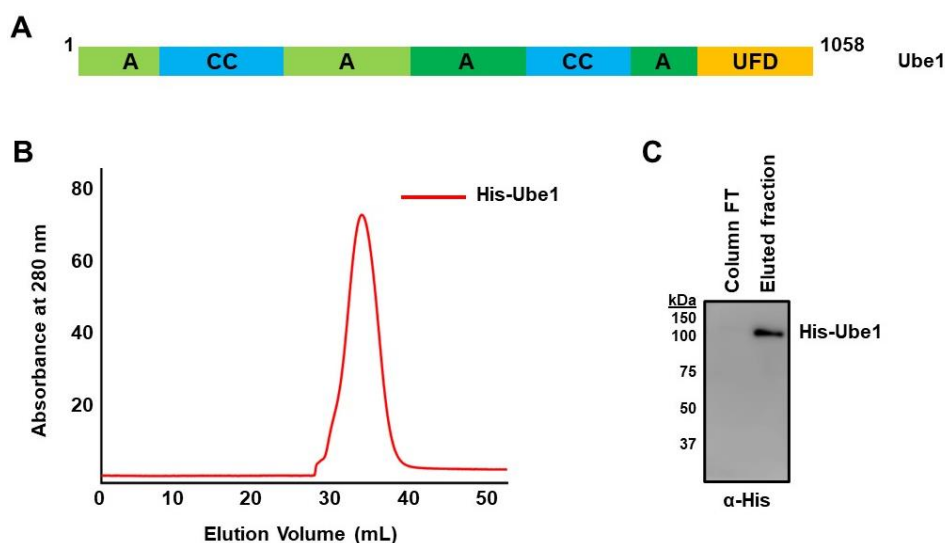


Fig. 4.2 Purification of Ube1. (A) Domain architecture of Ube1 with the functional domains highlighted in separate colours. A - adenylation domain, CC - catalytic cysteine domain, UFD- Ub-fold domain. (B) Elution profile of Ube1 following anion exchange chromatography with a HiTrap Q HP column. (C) Western blot of eluted purified protein probed by anti-His antibody, FT- Flowthrough.

Purification of His-tagged Ube1 was challenging since the protein was highly sensitive to temperature and was susceptible to protease cleavage. Hence, to solve this issue, all buffers involved in the purification protocol were supplemented with PMSF and the protein samples and buffers were maintained at 4°C throughout the procedure. I also opted for a rapid purification protocol so that the entire procedure could be completed within a day. Ube1, after being affinity-purified, was further subjected to size exclusion chromatography followed by anion exchange chromatography. The purified Ube1 was eluted with the help of NaCl gradient and dialyzed in storage buffer (Fig. 4.2B). Dialysed proteins were run on a gel, transferred into nitrocellulose membrane and probed with anti-His antibody to check purity (Fig. 4.2C). Purified Ube1 was aliquoted, flash frozen and stored at -80°C.

UbcH6 belongs to class III E2 enzymes which are characterized by the N terminal extension added to the core UBC domain (Fig.4.2A). The full-length UbcH6 has a tendency to cleave into the core UBC domain in solution when purified *in vitro*. However, the UBC domain has been found sufficient for ubiquitin conjugation activity. Therefore, I subcloned UbcH6 core domain (39-193) into a pET28A vector and did all the experiments in this study with this

UbcH6 core domain only, which I will hereby denote as UbcH6c. UbcH6c elutes as a monomer following size exclusion chromatography and yields a molecular weight of about 17 kDa (Fig. 4.2B and C).

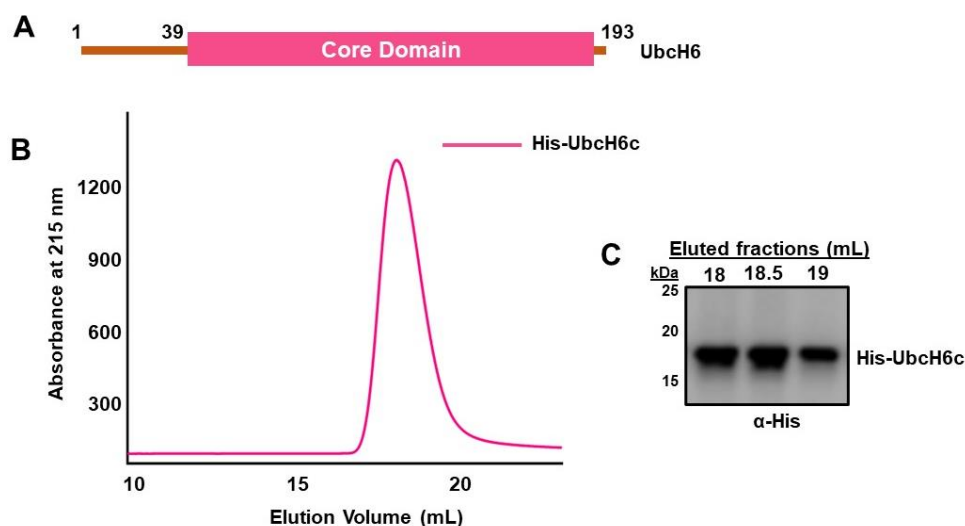


Fig. 4.3 Purification of UbcH6 core domain (UbcH6c). (A) Domain architecture of UbcH6 with the UBC core domain highlighted. (B) Elution profile of UbcH6c following size exclusion chromatography with a Superose 6 Increase 10/300 GL column. (C) Western blot of eluted purified protein probed by anti-His antibody.

Preparation of E2~Ub thioester had been described previously where the ATP dependant ubiquitin loading of E2 in presence of Ube1 was stopped using an ATP-diphosphohydrolase enzyme called Apyrase, which depleted the ATP present in the reaction mixture and arrested E2 charging (Plechanovová *et al*, 2012). However, instead of arresting the E2 charging reaction, I purified the ubiquitin loaded UbcH6c species by injecting the reaction sample into a size exclusion column and isolating it from the uncharged UbcH6c.

From the size exclusion chromatography, the peak obtained is mixed and comprises of ubiquitin loaded UbcH6c and only UbcH6c (not loaded), as indicated in Fig. 4.4A. The fractions from the tip of the peak are collected, concentrated and checked for impurities (Fig. 4.4B). The samples are run on a 15% SDS PAGE, blotted onto a membrane and probed with anti-ubiquitin antibody whereby purified UbcH6c~Ub is detected.

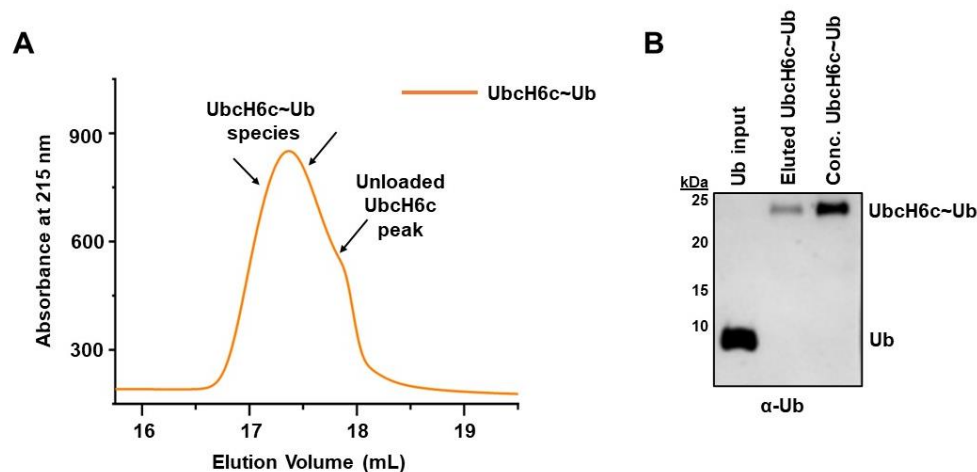


Fig. 4.4 Purification of UbchH6c~Ub thioester. (A) Size exclusion chromatogram of ubiquitin loaded UbchH6c following E2 charging assay using a Superose 6 Increase 10/300 GL column. **(B)** Western blot of eluted purified protein probed by anti-Ub antibody.

4.2 Identification and Characterization of UBR7-UbchH6 Interacting Region

As previously mentioned, among the members of the UBR7 protein family, UBR7 alone contains a PHD finger that is lacking in the rest of the family members (Tasaki *et al*, 2005). In 2019, our work helped establish UBR7 as an H2BK120 monoubiquitin ligase which suppresses tumour metastasis (Adhikary *et al*, 2019).

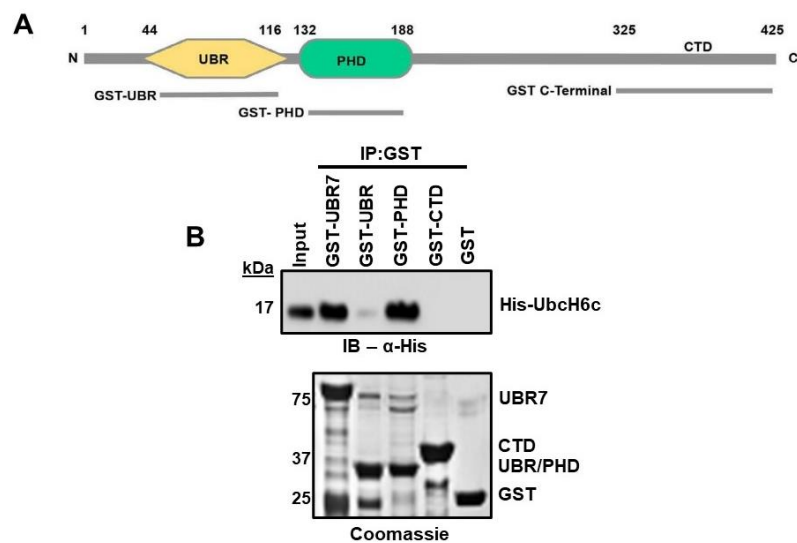


Fig. 4.5 Identification of E2 interacting domain of UBR7 *in vitro*. (A) Human UBR7 full-length protein with the N-terminal UBR box (yellow) and PHD finger (green) highlighted.

The GST-tagged constructs used in the study are indicated by grey bars. They represent the encompassing amino acid residues that span those GST constructs. (B) GST pull-down assay involving UBR7 and its different domains (UBR, PHD, C-terminal domain) with UbcH6c, blots probed with α -GST and α -His antibodies.

UBR7 harbours two functional domains, UBR box (41-123) and PHD finger (132-188), which are located closer to the N-terminal part of UBR7, whereas the C-terminal region is largely unstructured (Fig. 4.5A). I purified the GST tagged UBR7 full-length and domain constructs individually and assessed their UbcH6c binding ability through a GST pulldown assay. UbcH6c was found to strongly interact with wild-type (WT) and the PHD finger of UBR7. However, UBR box or the C-terminal region of UBR7 didn't exhibit any significant binding (Fig. 4.5B).

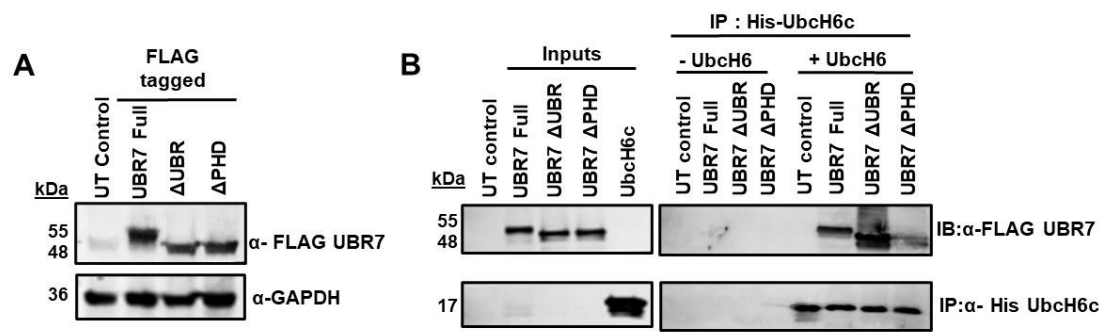


Fig. 4.6 Mapping the E2 interacting domain of UBR7 *ex-vivo*. (A) Expression levels of FLAG-tagged UBR7-full length, Δ UBR and Δ PHD constructs in HeLa cells. (B) Nickel-NTA pull-down assay of His-tagged UbcH6c with whole-cell lysates of overexpressed UBR7-full length, Δ UBR and Δ PHD, blots probed with α -His and α -FLAG antibodies.

In further check for UbcH6 interaction *ex-vivo*, I expressed FLAG-tagged full-length or truncated constructs of UBR7 in HeLa cells. The truncated constructs involved the deletion of either UBR box (44–116) or PHD finger (132–188). All the constructs were found to have comparable expression levels upon overexpression (Fig. 4.6A). Upon extramural addition of recombinant His-UbcH6c, a pull-down assay was performed using nickel-nitrilotriacetic acid (Ni-NTA) beads, it was observed that full-length UBR7 showed a strong interaction with UbcH6c. However, deleting the PHD finger resulted in a significant reduction in UbcH6c interaction when compared to the deletion of UBR box which showed no significant change

(Fig. 4.6B). This confirms that the UbcH6 interaction of UBR7 is accomplished through its PHD finger.

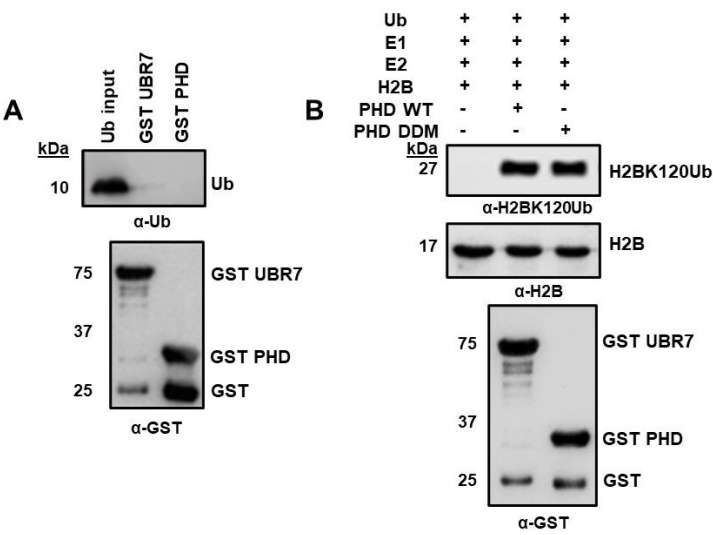


Fig. 4.7 UBR7 does not interact with ubiquitin *in-vitro* in absence of E2 and its PHD finger is sufficient to mediate ubiquitin transfer. (A) GST pull-down assay of UBR7-full length and PHD finger with untagged ubiquitin. (B) *In vitro* ubiquitination assay with UBR7-full length and PHD finger. H2BK120Ub was detected by α -H2BK120Ub antibody.

Since UBR7 PHD finger was found to interact with E2 enzyme UbcH6, I next tried to explore its interaction with ubiquitin *in vitro*. UBR7 doesn't have any canonical ubiquitin-binding motif and neither UBR7-full length nor its PHD finger was found to interact with ubiquitin in absence of its E2 partner, UbcH6 (Fig. 4.7A). However, UBR7 PHD was found sufficient for H2BK120 monoubiquitination as it catalyzed comparable levels of H2BK120Ub with respect to its full-length counterpart (Fig. 4.7B).

Since the crystal structure of UbcH6c encompassing the amino acid residues 39–193 (PDB ID: 5LBN) has been previously reported (Anandapadamanaban *et al*, 2019), I attempted to identify the probable residues of UbcH6 which could be involved in UBR7 interaction. Surprisingly, while looking into amino acid residues around the catalytic cysteine C131, the surface charge potential of loop 4 (residues 107–111) and loop 7 (residues 138–142) of UbcH6 seemed like possible regions involved in UBR7-PHD binding (Fig. 4.8).

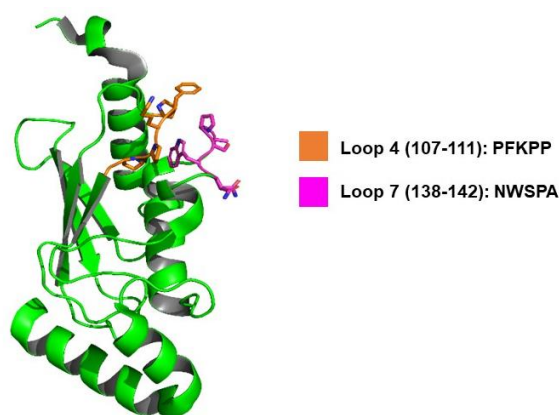


Fig. 4.8 Crystal structure of UbcH6. (PDB ID: 5LBN) Figure highlighting the loops adjacent to catalytic cysteine residue C131, loop 4 (orange) and loop 7 (magenta). The cartoon representation was generated using PyMOL, version 2.1.1.

In order to explore if the aforementioned loops of UbcH6 were indeed binding with UBR7, I deleted loop 4 and loop 7 within the UbcH6c background and obtained two-loop deleted mutants – Δ loop4 and Δ loop7. I then checked for their protein folding in solution. Upon CD spectroscopy with WT and loop deleted UbcH6c mutants (Fig. 4.9A), no significant structural change was observed in the loop deleted mutants. In an effort to investigate whether the loop deletions have any effect on Ube1 interaction or ubiquitin loading, I performed an E2 charging experiment with different time points with the UbcH6 WT and loop deletion mutants using Ube1 enzyme. No significant alteration in the UbcH6~Ub level was observed in any of the loop deletion E2 mutants when compared to the WT UbcH6 at any given time point (Fig. 4.9B). These results collectively suggest that deletion of loops 4 and 7 in UbcH6 doesn't alter the conformation of the UBC fold and doesn't have any inhibitory effect on E2-Ub charging.

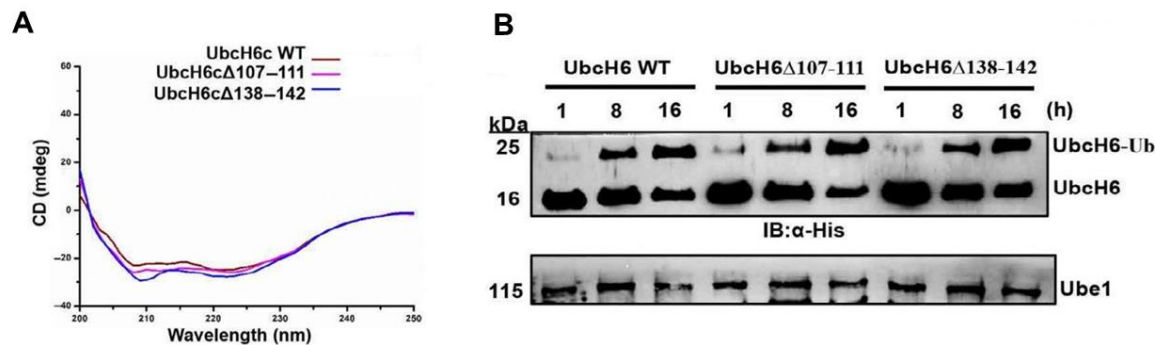


Fig. 4.9 Exploring the role of UbcH6 loops in stabilizing the UBC fold and mediating ubiquitin charging. (A) Far-UV CD spectra with UbcH6c WT and mutants. (B) Time-

course experiments with UbcH6c WT and loop deletion mutants to compare ubiquitin charging. UbcH6c-Ub species, UbcH6c and Ube1 proteins were detected by the α -His antibody.

Next, I compared the ability of UbcH6 loop mutants to bind to UBR7-PHD when compared to UbcH6c WT protein. Upon GST pulldown, both loop-deleted mutants of UbcH6 showed reduced interaction with GST tagged UBR7-PHD finger (Fig. 4.10A). I then performed an *in vitro* ubiquitination assay to explore if loop 4 and loop 7 of UbcH6 contribute significantly to UBR7-mediated histone ubiquitination. When compared to WT UbcH6c, loop 4 and loop 7 deleted UbcH6c mutants showed significant reduction in H2BK120 monoubiquitination levels (Fig. 4.10B). These results, taken together, suggest that UBR7 PHD finger interacts with loop 4 and loop 7 of UbcH6 and this interaction is crucial for ubiquitin transfer onto substrate H2B.

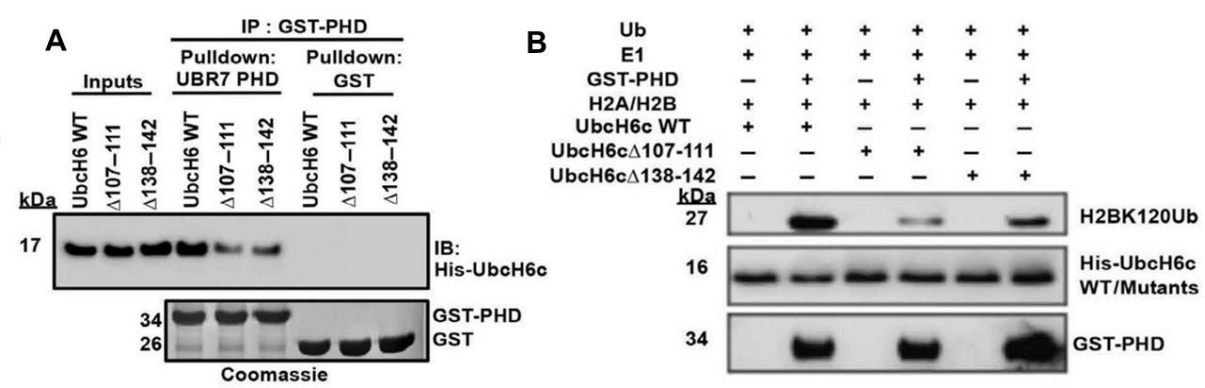


Fig. 4.10 UbcH6 loops 4 and 7 are involved in UBR7 binding. (A) GST pull-down assay with UBR7-PHD and UbcH6c WT loop deleted mutants. (B) *In vitro* ubiquitination assay of UBR7 with UbcH6 WT and loop deleted mutants. H2BK120Ub antibody was used to detect the ubiquitinated H2B species.

4.3 Characterization of H2B Ubiquitination by UBR7-UbcH6 Complex

The RING E3 ligase enzymes target the substrate protein and the cognate E2-ubiquitin thioester both as co-substrates during the ubiquitin transfer reaction. Thus, it is only natural that both substrate and the E2 have a corresponding K_m associated with their respective binding to the ligase. Such binding typically resembles a hyperbolic concentration dependence on the rate of substrate ubiquitination, measured as the function of ubiquitinated product formation. To study this, I performed a time course ubiquitination assay experiment where the formation

of the ubiquitinated product - H2BK120Ub was measured over several time periods ranging between 0–20 minutes. During the duration of this assay, concentrations of the substrate, E3 ligase enzyme and E2~Ub thioester were kept constant. Upon assessing the product formation through an antibody-dependent technique, I obtained a time-dependent increase in H2BK120 ubiquitination up to 15 minutes timepoint post which the E3 ligase enzyme is likely saturated (Fig. 4.11A).

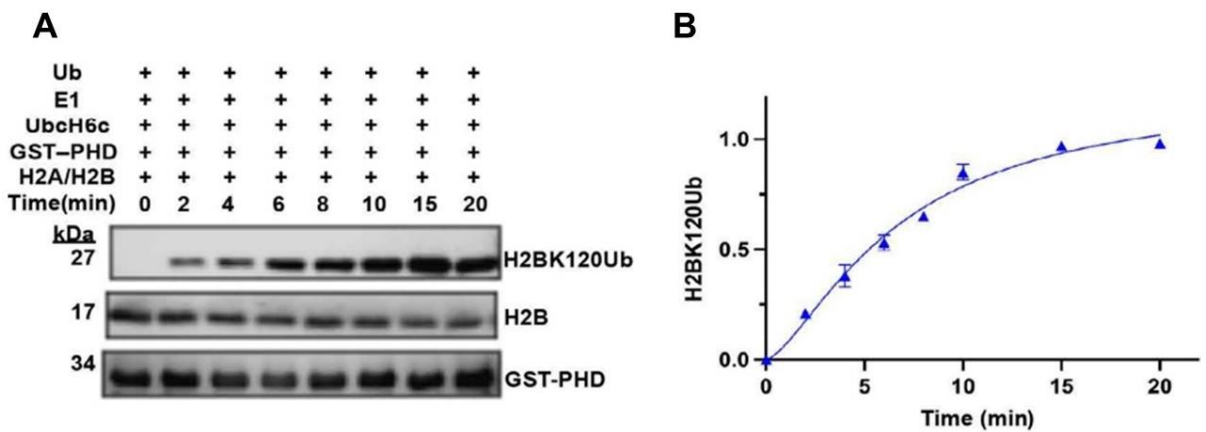


Fig. 4.11 UBR7/UbcH6 enzyme complex catalyzes ubiquitin transfer onto histone H2B C-terminus. (A) UBR7 mediated single-turnover substrate ubiquitination assay with different reaction time points (0–30 min). The experiment was performed in triplicates. (B) The H2BK120 monoubiquitination band intensity obtained was quantified corresponding to each time point plotted in a product versus time graph. The average H2BK120Ub intensity (n= 3 repeats) at the corresponding time point is represented as a triangle. The error bars indicate the standard deviation for each time point. Hill equation was used for data fitting.

I performed the single-turnover substrate ubiquitination assay in triplicates and averaged the obtained values of H2BK120Ub formation corresponding to each time point. Using the Hill equation, these values were plotted in a product versus time graph (Fig. 4.11B).

From the previous study, we already knew that UBR7 interacts with histone H2B both as free dimers and as nucleosomes in vitro and can efficiently ubiquitinate either of them (Adhikary *et al*, 2019). The next question which I tried to address was if the histone fold of H2B had a role to play in H2BK120 monoubiquitination. However, to answer that query, it was first required to identify the minimum stretch of histone H2B sequence that was required as a substrate for ubiquitin transfer. To address this, I generated a biotin-conjugated H2B peptide

encompassing the H2B C-terminal tail (114–125) and a K120R mutant peptide. The mutation is within the same C-terminal tail and is on the same lysine residue that gets ubiquitinated.

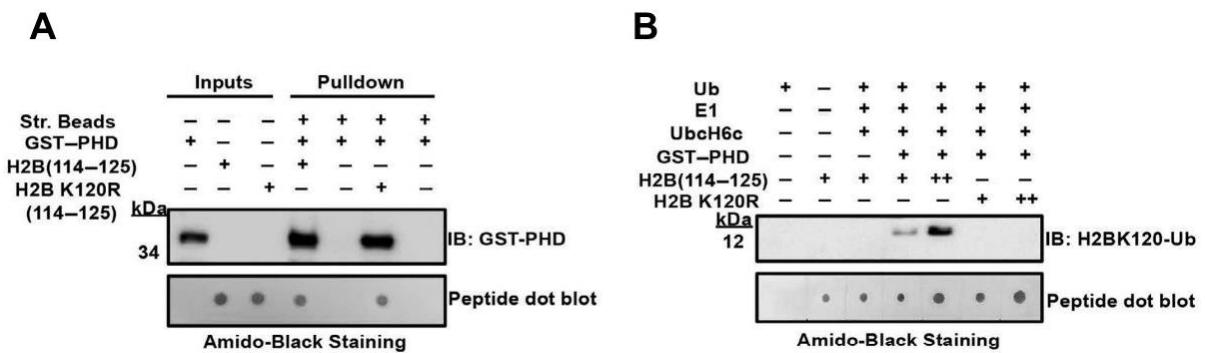


Fig. 4.12 H2B C-terminal tail is sufficient to act as substrate in UBR7 mediated ubiquitin transfer. (A) Streptavidin pull-down assay using biotinylated H2B C-terminal WT (114–125) and the K120R mutant peptide to test binding with UBR7-PHD. α -GST antibody was used to detect GST tagged UBR7-PHD. (B) In vitro ubiquitination assay with H2B WT and K120R mutant peptides as substrates.

I started by checking the ability of these peptides to bind UBR7-PHD with the help of a peptide pulldown assay. It was observed that both wildtype and K120R mutant peptides interacted strongly with UBR7-PHD finger (Fig. 4.12A). However, in an *in vitro* ubiquitination assay involving UBR7-PHD as E3, the H2B WT peptide was efficiently ubiquitinated and was thus detected using anti-H2BK120Ub antibody. The corresponding K120R mutant peptide, where the lysine residue that was to be ubiquitinated was replaced by arginine, showed no such ability to get ubiquitinated. (Fig. 4.12B).

Within the nucleus, H2B exists in the nucleosomal and no-nucleosomal pools being bound to its partner, histone H2A. Since H2A accompanies H2B forming a heterodimeric complex, it was crucial to examine if the interaction of UBR7-PHD with its substrate H2B was specific. Keeping this in mind, I performed a fluorescence spectroscopy experiment to quantify the affinity of UBR7-PHD towards FAM- conjugated H2B peptide. The binding data obtained indicated strong binding and the binding affinity (Kd) was calculated to be around 0.3 ± 0.02 μ M. However, no such binding of UBR7-PHD was observed with the corresponding H2A C-terminal peptide (118–129) (Fig. 4.13).

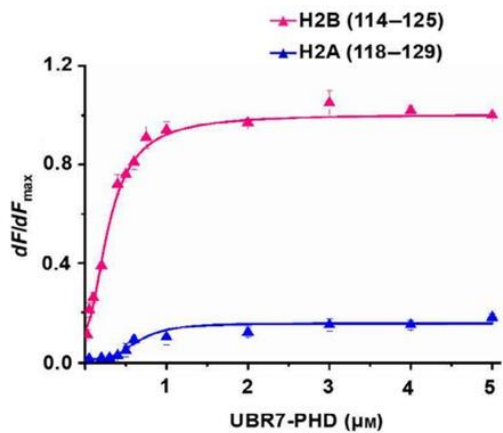


Fig. 4.13 The affinity of UBR7-PHD is specific towards H2B C-terminal tail. The binding isotherms of UBR7-PHD obtained from fluorescence spectroscopy using FAM-conjugated H2B C-terminal peptide (GTKAVTKYTSSK), trace in pink and H2A C-terminal peptide (KKTESHHKAKGK), trace in blue (n = 3 repeats; SD is denoted by the error bars).

Histone ubiquitinations so far have been reported in the lysine residues of the dynamic N-terminal or the C-terminal tails. In addition to the lysine residue at 120th position, C-terminal of histone H2B harbours two more lysine residues at 116, and 125 positions, which could be ubiquitinated. To accurately determine the exact site of ubiquitination, I used mass spectrometry for the analysis of ubiquitinated samples from an assay with UBR7-PHD and the H2B WT and mutant peptides.

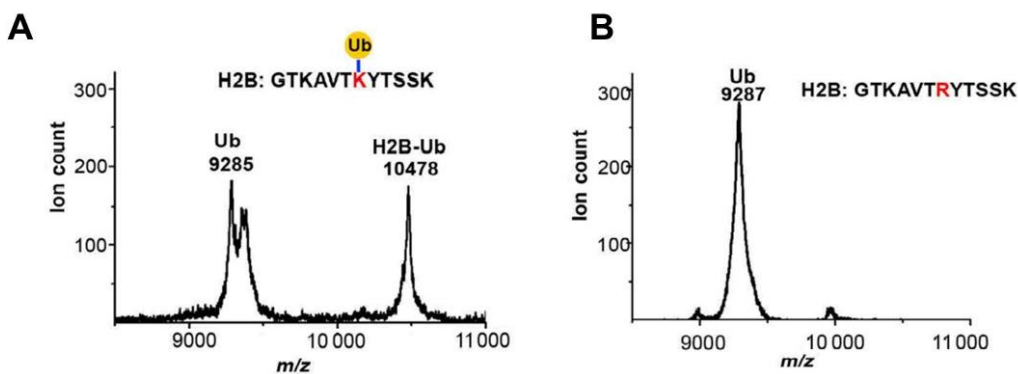


Fig. 4.14 Detection and determination of molecular mass of ubiquitinated H2B peptide using MALDI-TOF mass spectra. (A) Mass spectra of ubiquitination assay sample using histone H2B WT peptide. From the left, free ubiquitin (Ub) of mass 9.285 kDa followed

by the ubiquitinated H2B peptide (114–125) with mass of 10.478 kDa are labelled. (B) Mass spectra of assay sample using the H2B K120R mutant peptide as substrate.

In case of the reaction with H2B WT peptide, a distinct peak was obtained which was determined to be 10.478 kDa. This mass was consistent with the mass of the H2B WT peptide (~1.2 kDa) and monoubiquitin (~9.2 kDa) taken together. As expected, no such peak was observed in the case of H2B K120R mutant peptide (Fig. 4.14). The control run in absence of UBR7-PHD enzyme did not yield any ubiquitinated H2B peptide either. These results together suggest that the C-terminal tail of histone H2B is a specific substrate and is sufficient by itself for UBR7-mediated ligase activity.

4.4 Determination of Oligomeric Status of UBR7 in Solution and its Role in E3 Ligase Function

RING fingers of several E3 ligases tend to form homo- or heterodimers in solution. Their ability to form dimers has been shown to play an important role in attenuating their E3 ligase activity. Previously, no knowledge was available regarding the oligomeric status of UBR7 in solution.

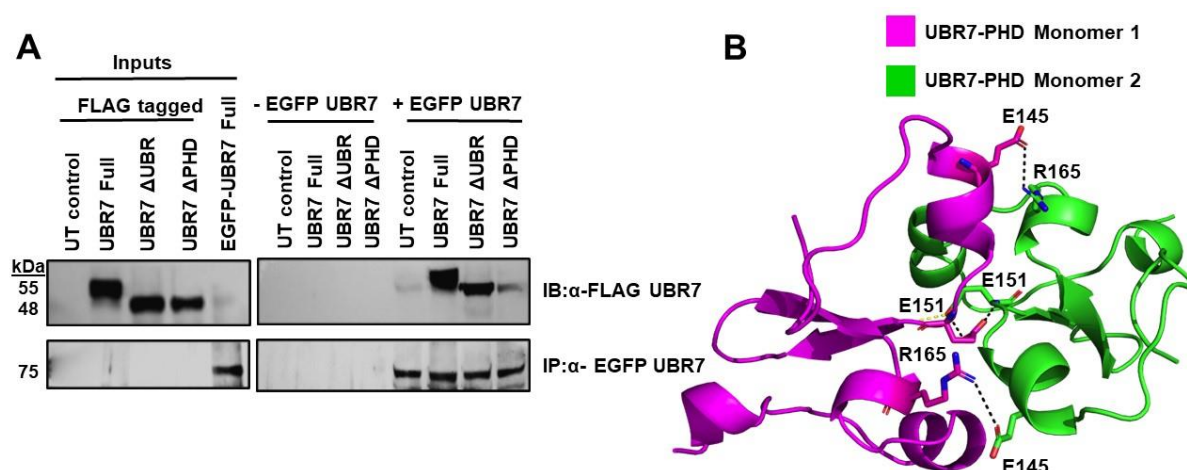


Fig. 4.15 UBR7-PHD finger exists as a dimer in solution. (A) Co-transfection of FLAG-UBR7 or FLAG-conjugated deletion constructs with EGFP-UBR7 was performed in HeLa cells. Following immunoprecipitation with α-EGFP antibody, the interaction was assessed using α-FLAG antibody (IB). **(B)** Molecular model of UBR7-PHD constructed using structure of human BPTF PHD finger (PDB ID: 2F6N) as a template. The potential

residues located within the dimerization interface – E145, E151 and R165 are highlighted as sticks.

To explore if UBR7 has any oligomeric association ex vivo, EGFP-UBR7 WT was co-transfected with either FLAG-UBR7 WT or domain deleted constructs (FLAG-UBR7ΔPHD or FLAG-UBR7ΔUBR) in HeLa cells. Following immunoprecipitation procedure with EGFP antibody, blots were probed to check for association of FLAG-proteins with EGFP-UBR7. It was observed that FLAG-UBR7 WT and FLAG-UBR7 ΔUBR had retained a strong association with EGFP-tagged UBR7. However, the association with EGFP-UBR7 was drastically reduced in the case of FLAG-UBR7ΔPHD (Fig. 4.15A). This indicates that the PHD finger plays a key role in the self-association of UBR7.

In absence of a crystal structure, molecular modelling technique was employed to predict the structure of the UBR7-PHD finger. This was done using Bromodomain PHD finger transcription factor (BPTF)-PHD finger (PDB ID: 2F6N) as a reference template with the PHYRE2 algorithm. Among the many different PHD fingers reported, BPTF-PHD was chosen since it had significant sequence similarity (43%) with UBR7-PHD. The dimeric structure shows that the main chain carbonyl and side-chain hydroxyl group of E151 residues of two monomers form hydrogen bonds with each other, while prominent salt bridge interaction connected E145 of one monomer with the R165 of the other (Fig. 4.15B).

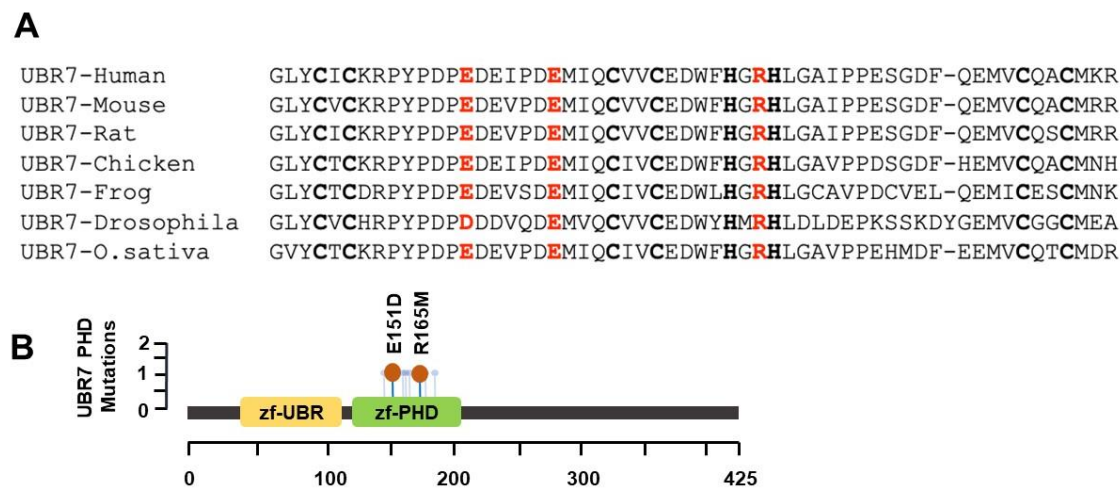


Fig. 4.16 UBR7-PHD finger remains conserved among higher eukaryotes. (A) Multiple sequence alignment of UBR7-PHD across different species. The highly conserved residues

involved in dimerization – E145, E151 and R165 are highlighted in red. The zinc coordinating residues are indicated in bold. Sequences of UBR7-PHD from different species were obtained from the UniProt database.

UniProt accession numbers: UBR7- Human (Q8N806), UBR7-Mouse (Q8BU04), UBR7-Rat (Q642A8), UBR7-Chicken (Q5ZMN4) and UBR7-Frog (Q569T8). (B) UBR7 missense somatic mutations within the PHD finger in different cancers identified from CBioPortal database.

The three residues, E145, E151 and R165, which are located at the dimerization interface are found conserved across higher eukaryotes (Fig. 4.16A). It is also to be noted that the cBioPortal database, which lists the mutations of different proteins reported in various cancers, mentions that the two of the three residues encompassing the dimerization interface are reported to be target to mutations (E151D and R165M) which are implicated in uterine endometrial carcinoma and papillary renal cell carcinoma, respectively (Fig. 4.16B).

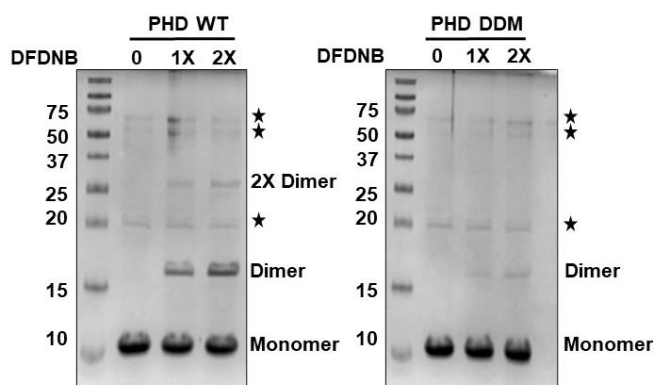


Fig. 4.17 DFDNB mediated cross-linking of UBR7-PHD WT and triple mutant (DDM). Star (★) denotes non-specific contamination bands.

With the purpose of investigating the role of these residues in the dimerization of UBR7, I next mutated these interface residues. The disability of the triple mutant E145K/E151A/R165A to form dimers was confirmed using a 1,5-difluoro-2,4-dinitrobenzene (DFDNB)- mediated cross-linking assay (Fig 4.17), where the triple mutant showed a significant loss of dimerization upon cross-linking when compared to UBR7-PHD WT. I have referred to the UBR7 triple mutant as dimer deficient mutant (DDM) in the figure and legend section of this thesis from here onwards.

Next, I determined the molecular mass of UBR7-PHD WT and its corresponding mutant DDM using size exclusion chromatography coupled with multiple-angle light scattering (SEC-MALS) (Fig. 4.18A) under low salt buffer conditions. Interestingly, UBR7-PHD WT yielded a molecular mass of 18.3 kDa (the expected monomer mass was 8.4 kDa), which was consistent with a dimer. However, the obtained mass in the case of UBR7-PHD DDM was 6.6 kDa (expected monomer mass was 8.2 kDa) (Fig. 4.18B), thereby confirming its dimer deficiency.

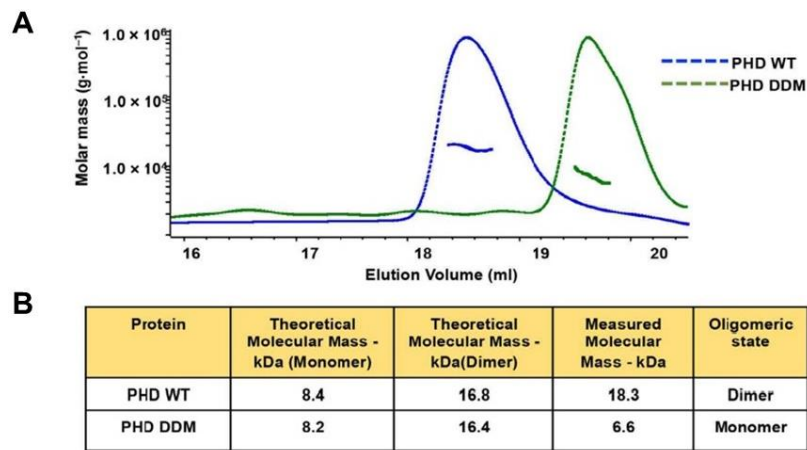


Fig. 4.18 SEC-MALS analysis of untagged UBR7-PHD WT and DDM mutant. WT is denoted by a dashed blue line and dimer-deficient triple mutant (DDM) denoted by dashed green line, respectively. (A) Dashed lines represent the refractive index trace for WT and mutant proteins and the horizontal line under the peak represents the average molar mass (y-axis) distribution across the peak as determined by MALS. (B) Above molar mass measurements by SEC-MALS are summarized in a table.

Once the triple mutant was confirmed to be dimer deficient, I tried to investigate the role of those amino acid residues encompassing the dimer interface on E3 ligase activity of UBR7-PHD. This was achieved by performing an *in vitro* ubiquitination assay involving WT UBR7-PHD and DDM and their ability to mediate H2BK120Ub formation was compared.

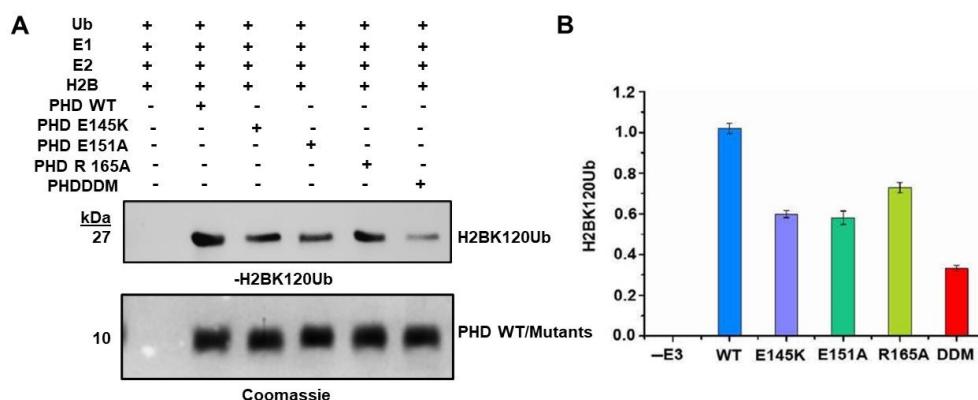


Fig. 4.19 UBR7-PHD dimerization is essential for its E3 ligase activity. (A) *In vitro* ubiquitination assay with untagged UBR7-PHD WT, single mutants (E145K, E151A, R165A) and triple mutant DDM. Ubiquitinated H2B was detected with α -H2BK120Ub antibody. PHD WT and its mutant proteins were visualized by Coomassie Brilliant Blue staining. (B) Quantification of the obtained H2BK120Ub level in the above assay using IMAGEJ with $n=3$ repeats; error bars represent the SD.

With respect to UBR7-PHD WT, the single mutants involving mutations of the three residues that constituted the dimerization interface (E145K, E151A, R165A) was found to have a modest reduction in H2BK120Ub formation. Moreover, the dimer-deficient triple mutant (DDM) showed a significant reduction in its ubiquitinating ability towards histone H2BK120 (Fig. 4.19A). The extent of substrate ubiquitination in case of the mutants with respect to WT was quantified using IMAGEJ software (Fig. 4.19B). This suggests that the PHD finger mediated dimerization of UBR7 plays a key role in maintaining its E3 ligase activity. Often, dimerization of RING E3 ligases plays an important role in their E2 or substrate-binding function. In order to gain more insights into the role of UBR7-PHD dimerization on its E2 and substrate interaction, I performed a Ni-NTA pull-down assay with His-tagged UbcH6c and untagged PHD WT or DDM. The results suggested that upon dimer deficiency, UBR7 binding to UbcH6 is severely compromised (Fig. 4.20A). However, in the case of the substrate, there was a moderate reduction in interaction between PHD DDM and histone H2A-H2B dimer when compared to WT in a Ni-NTA pulldown assay (Fig. 4.20B). These results, taken together, indicates that dimerization of UBR7 plays an essential role in both UbcH6c binding and to an extent, substrate H2B binding.

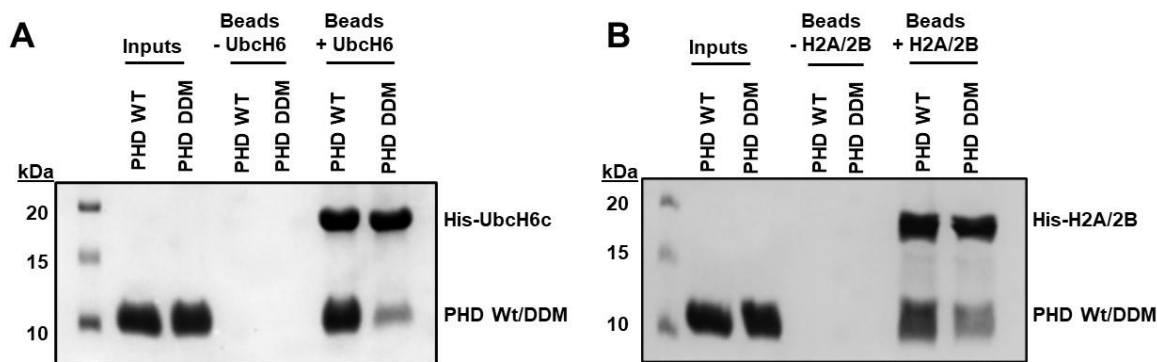


Fig. 4.20 Dimerization of UBR7 regulates its E2 and substrate binding ability. (A) Interaction of His-UbcH6c with PHD WT or DDM mutant was analyzed by a nickel-nitrilotriacetic acid pull-down assay run on SDS/PAGE. **(B)** SDS/PAGE profile of Ni-NTA pulldown assay involving UBR7 PHD WT and DDM with H2A-H2B dimer.

4.5 Comparison of the Mechanism of UBR7 Mediated Ubiquitin Transfer with other H2BK120 Monoubiquitin Ligases

Among the different RING E3 ligases that are reported in the catalysis of histone ubiquitinations, the RNF20/40 complex, which mediates H2BK120 monoubiquitination in humans is very well known. Though it exists as a heterodimeric state within the cell, RING fingers from both RNF20 and RNF40 have been reported to be able to homo-dimerize individually *in vitro* and are capable of targeting E2~Ub thioester for hydrolysis (Foglizzo *et al*, 2016). Since RNF20 is an active dimeric RING E3 ligase by itself and also uses the same E2 enzyme as UBR7, UbcH6 to target the H2BK120 site, hence I became interested in finding out how these two E3 ligases, UBR7 and RNF20 differ with each other in terms of mechanism of ubiquitin transfer.

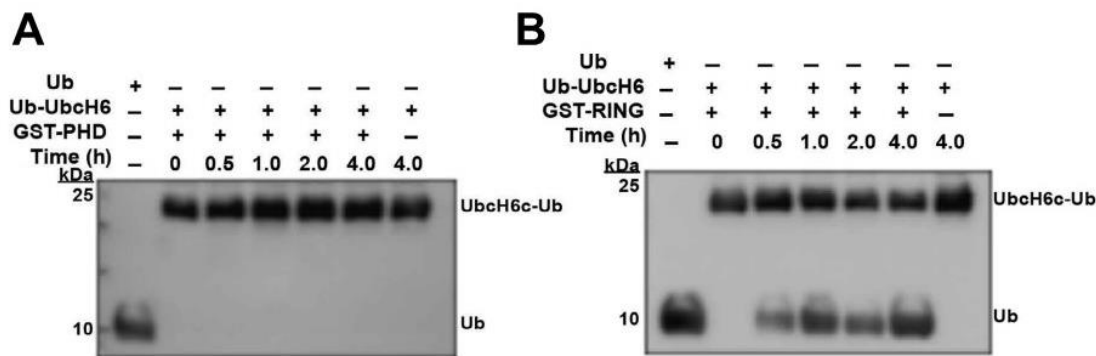


Fig. 4.21 E2~Ub thioester targeting of UBR7-PHD and RNF20-RING in absence of substrate. (A, B) UbcH6c~Ub thioester conjugate was incubated with UBR7-PHD (A) or with RNF20-RING (B) in absence of histone H2B substrate and the ubiquitin transfer was measured by western blotting with anti-Ub antibody.

To investigate how ubiquitin transfer is facilitated from UbcH6 to H2B, by the E3 ligases UBR7 or RNF20, I performed a ubiquitin discharge assay using UbcH6~Ub thioester. In this assay, I incubated purified UbcH6~Ub conjugate with either UBR7-PHD finger or RNF20-RING with increasing time points. The hydrolysis of UbcH6~Ub was measured by detecting the release of free ubiquitin. Interestingly, UBR7-PHD could not hydrolyze the thioester bond of UbcH6~Ub in absence of substrate H2B, whereas the RNF20 RING finger alone could do so (Fig. 4.21).

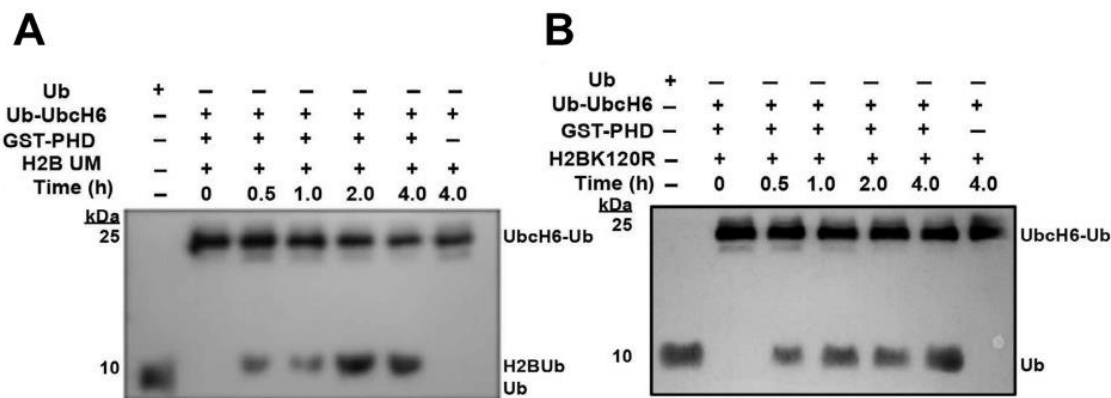


Fig. 4.22 Substrate induced E2~Ub thioester hydrolysis of UBR7-PHD. Single-turnover discharge assay with UBR7-PHD (A) in the presence of substrate C-terminal peptide of histone H2B (114-125) (B) or its mutant (H2B K120R).

I next tried to determine whether UBR7 needed to interact with its substrate to initiate ubiquitin release. To address this, I repeated the ubiquitin discharge assay with UBR7-PHD, only this time in presence of H2B C-terminal peptide (114–125). Successful transfer of ubiquitin to H2B peptide was indeed observed (Fig. 4.22A). This suggested that substrate binding is essential for UBR7 to target and hydrolyze E2~Ub thioester. It was worth investigating if the presence of lysine 120 residue of histone H2B is involved in ubiquitin release or if the substrate-binding was sufficient. I performed ubiquitin discharge assay with UBR-PHD and H2B K120R peptide and was able to detect ubiquitin release in a time-dependent manner. This further confirmed that the substrate binding is essential and sufficient for the E3 ligase activity of UBR7 (Fig. 4.22B).

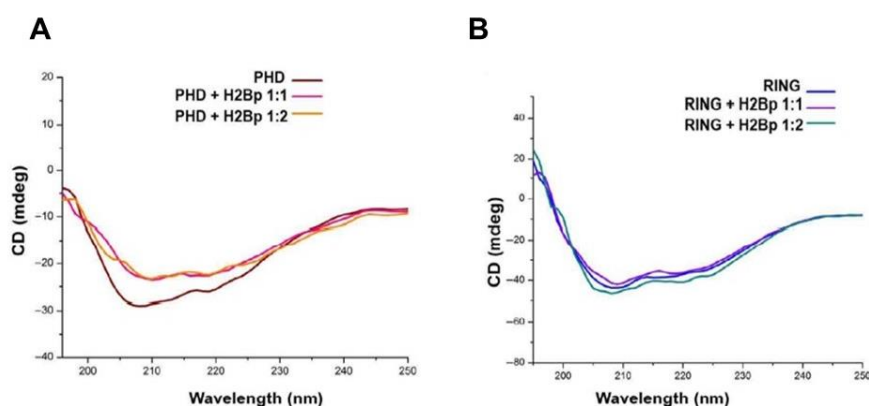


Fig. 4.23 Substrate association induces conformational change in UBR7-PHD finger. Far-UV CD spectra of untagged UBR7-PHD (A) and RNF20 RING (B) with and without peptide.

To investigate if substrate-binding played a key role in the activation of UBR7-PHD and subsequent ubiquitin discharge from E2~Ub, I employed CD spectroscopy to study any possible structural changes in UBR7-PHD that might occur in presence of substrate H2B. Upon analysis of far UV CD spectra, it was found that UBR7-PHD underwent a significant conformational change after substrate H2B binding (Fig.4.23A) but no such change was observed in the case of RNF20 RING (Fig. 4.23B). These results clearly establish that the ubiquitin transfer mechanism of the E3 ubiquitin ligases, UBR7 and RNF20, which ubiquitinates the same histone H2B substrate with the E2 conjugating enzyme UbcH6, is distinctly different.

Chapter 5

Discussion

Plant Homeodomain (PHD) finger proteins are well established as versatile chromatin readers which are associated with recognising and binding methylated and unmodified histone H3K4 mark (both H3K4Me3 and H3K4Me0) or acetylated H3 or H4 (Musselman & Kutateladze, 2011; Sanchez & Zhou, 2011). However, the catalytic function of this domain has not been explored as of yet. Different members of the UBR protein family have been associated with E3 ubiquitin ligase function due to the presence of canonical RING, HECT or F-Box domains (Tasaki *et al*, 2009). UBR7, the smallest member of the UBR protein family was earlier reported to carry a PHD domain, but not much was known about its function. It was only recently that UBR7 was established as a histone E3 ligase which targeted H2BK120 for transfer of a single ubiquitin (Adhikary *et al*, 2019). Even then, the mechanism of E3 ligase function of UBR7 was not well understood. The current thesis work involves an attempt to understand how UBR7 interacts with its E2 partner UbcH6 and mediates ubiquitin transfer to histone H2B. These results, for the first time, establish the role of the UBR7-PHD finger as a ‘writer’ of an epigenetic mark which is H2BK120 monoubiquitination. Since it was already known that UBR7 utilizes UbcH6 as E2 for ubiquitin conjugation, we started by looking for the E2 interacting region of E3 ligase UBR7 in the context of its binding to UbcH6. Of the two different functional domains and a largely unstructured C-terminal tail, PHD finger of UBR7 was found to specifically interact with UbcH6 as well as substrate H2B. Next, I also identified the critical amino acid stretch of UbcH6 that is involved in UBR7 interaction. Several E2-E3 structures reveal that the E3 binding site of most E2 enzymes, including UbcH6 is comprised of helix H1 and loops L1 and L2 which are generally located distant to catalytic cysteine residue (Stewart *et al*, 2016). However, in the case of UBR7, amino acids encompassing loop L4 and L7 of UbcH6 was found to play a crucial role in catalysis. Upon their deletion, although the E2-Ub thioester forming ability stayed intact, there was a reduction in the overall H2BK120Ub levels as the interaction of UbcH6 with E3 UBR7 was severely compromised.

Histone H2B was previously established as a substrate of UBR7, and it was known that UBR7 targets H2B for monoubiquitination in both free dimeric and nucleosomal states (Adhikary *et al*, 2019). In the current work, I further established that the C-terminal tail of histone H2B (114-125) was sufficient for its interaction with UBR7 and for its ability to get ubiquitinated. mediating this interaction. Point mutation of H2B targeting its substrate lysine residue (K120R) was found to abolish its capacity to be ubiquitinated. However, it was interesting to note that the lysine residue at the 120th position of H2B was not critically involved in its E3 binding as

the H2B K120R mutant showed similar binding to UBR7-PHD finger as compared to H2B WT peptide. Thus, this might indicate that some critical residues other than lysine 120 is involved in E3 binding activity of substrate. However, the UBR7 binding residues for histone H2B are yet to be identified and validated. There have been several reports where RING fingers being part of larger macromolecular complexes where there are separate modules of E2 and substrate binding. This is evident in case of multi-subunit RING E3 of Cullin RING ligase (CRL) superfamily, such as the well-studied SCF which hosts a small RING finger protein (Rbx1), a cullin protein (Cul1) and an adaptor protein (Skp1) which is involved in binding of substrate recognition elements (Metzger *et al* , 2014). There have also been reports of E2 binding involving the N terminal of RING finger and not with the RING finger directly. I, however, found that the UBR7 PHD finger alone is capable of mediating H2BK120Ub formation, similar to its full-length counterpart. This result suggests that UBR7 PHD can accommodate both E2 and substrate binding all on its own. Through a mass-spectrometric experiment, I also ensured that the H2B monoubiquitination that was mediated by UBR7 was specifically onto the K120 residue, and not K116 or K125.

I also found out that UBR7-PHD finger exists a dimeric E3 ligase module *in vitro*. UBR7 dimerization was then experimentally validated in cells whereupon deletion of PHD finger, the self-association seems to be compromised. This was similar in line with reports of many of the RING finger proteins, involved in catalysis of various substrates both histone and non-histone, having a dimeric association which is crucial for their E3 ligase activity. The critical residues present at the dimerization interface of higher eukaryotes were found to be highly conserved. The importance of these residues was further confirmed in The Cancer Genome Atlas (TCGA) database, which lists the different mutations reported in cancer. Two of the three residues found at the dimerization interface were found mutated in uterine endometrial carcinoma and papillary renal cell carcinoma, among other cancers.

In an *in vitro* system, mutation of all three of these residues resulted in dimer deficiency of UBR7-PHD. Dimer deficiency of UBR7-PHD had significantly compromised the E3 ligase function of UBR7. It was also seen to reduce substrate binding and especially E2 association of UBR7. This might be due to instability of the PHD finger due to dimer deficiency or might be an indication that the E2 binding surface is located adjacent to the dimerization interface and is disrupted when the dimerization is compromised. Since the role of dimerization of UBR7-PHD was not previously explored, the current study establishes UBR7 as a dimer which

is crucial for eliciting the enzymatic catalytic activity. However, the role of these key residues involved in dimerization needs to be studied in a cellular context to understand their contribution towards tumorigenicity.

There have been several previous reports focussing on RNF20 and its E3 ligase activity towards the catalysis of H2BK120 monoubiquitination. To address how the two E3 ligases, using the same E2 enzyme while targeting the same substrate, might operate differently, I compared the mechanism of ubiquitin transfer between UBR7 and RNF20. The requirement of substrate was found critical for efficient hydrolysis of UbcH6- Ub thioester bond in case of UBR7. This was, however, dispensable for RNF20. This result suggests that the choice of E3 ligase to be employed, whether UBR7 or RNF20, is made by the UbcH6-mediated ubiquitination machinery based on the availability and relative abundance of substrate histone H2B. The substrate binding to PHD finger is sufficient to bring about activation of UBR7 PHD finger, as the presence of the lysine residue was not deemed essential for targeting and effective hydrolysis of UbcH6-Ub thioester. Upon further investigation, I found that the substrate association induces a significant conformational change within the UBR7 E3 ligase module which then becomes primed for initiating the ubiquitin discharge from UbcH6. This sort of conformational change was not observed in RNF20.

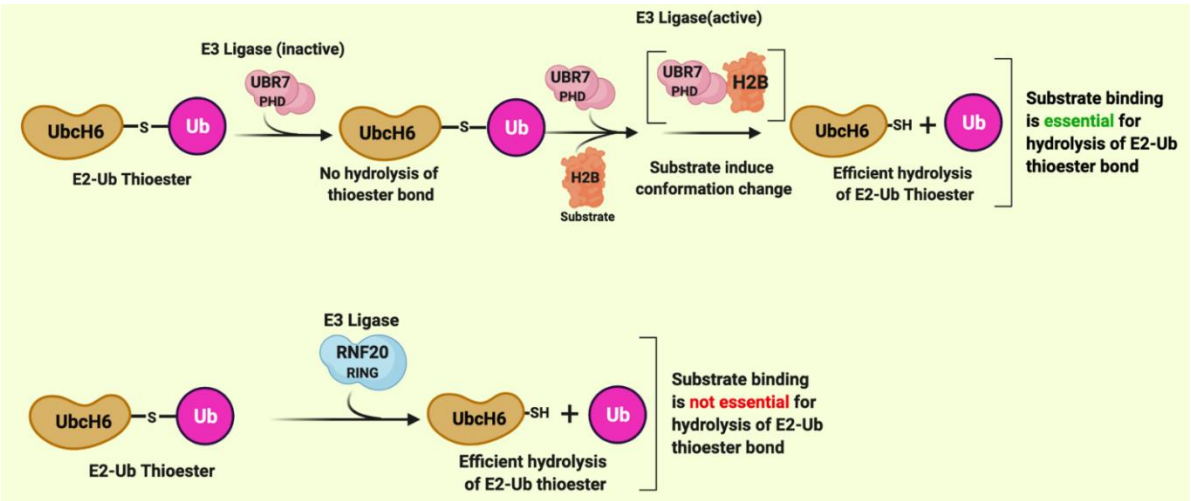


Fig. 5.1 Working model of ubiquitin transfer mechanism of UBR7. PHD finger dimerization and substrate H2B association are essential for the E3 ligase function of UBR7 unlike RNF20.

The diving deep into the differentiating the mechanism of function of both E3 ligases, the present study has clearly established a mechanism distinct from the canonical RNF20-mediated transfer to histone H2BK120.

Monoubiquitination mediated activation of signalling cascade has been previously implicated in transcriptional regulation and DNA repair. The role of human E3 ligases in regulating H2BK120Ub levels within the cell and thereby acting as tumour suppressor against uncontrolled cell proliferation, invasion and metastasis are of vital importance. It is only by fully understanding the molecular details of how the cell machinery chooses one enzyme over the other and differentiates between their mechanism on the account of substrate and co-substrate availability, that we may make leaping therapeutic progress towards our fight against cancer.

Bibliography

1. Olins DE, Olins AL (2003) Chromatin history: our view from the bridge. *Nat Rev Mol Cell Biol* **4**(10): 809–814.
2. Jansen A, Verstrepen KJ (2011) Nucleosome Positioning in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **75**(2): 301–20.
3. Watson JD, Crick FH (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**(4356): 737–8.
4. Gall JG (1963) Kinetics of deoxyribonuclease action on chromosomes. *Nature* **198**: 36–38.
5. Zubay G, Doty P (1959) The Isolation and Properties of Deoxyribonucleoprotein Particles Containing Single Nucleic Acid Molecules. *J Mol Biol* **1**: 1–20.
6. Allfrey VG, Faulkner R, Mirsky AE (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc Natl Acad Sci USA* **51**(5): 786–794.
7. Johns EW (1969) The histones, their interactions with DNA, and some aspects of gene control. *Ciba Foundation Symposium-Homeostatic Regulators* Chichester, UK: John Wiley & Sons, Ltd.
8. Clark RJ, Felsenfeld G (1971) Structure of chromatin. *Nature New Biology* **229**(4): 101–106.
9. Itzhaki RF (1971) Studies on the accessibility of deoxyribonucleic acid in deoxyribonucleoprotein to cationic molecules. *Biochem J* **122**(4): 583–592.
10. Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). *Science* **183**(4122): 330–332.
11. D'Anna JAJ, Isenberg I (1974) A histone cross-complexing pattern. *Biochemistry* **13**(24): 4992–4997.
12. Roark DE, Geoghegan TE, Keller GH (1974) A two-subunit histone complex from calf thymus. *Biochem Biophys Res Comm* **59**(2): 542–547.
13. Kornberg R (1974) Chromatin structure: a repeating unit of histones and DNA. *Science* **184**(4139): 868–871.
14. Kornberg RD, Thomas JO (1974) Chromatin structure; oligomers of the histones. *Science* **184**(4139): 865–868.
15. Oudet P, Gross-Bellard M, Chambon P (1975) Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell* **4**(4): 281–300.

16. Lorch Y, LaPointe JW, Kornberg RD (1987) Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* **49**(2): 203-10.
17. Han M, Grunstein M (1988) Nucleosome loss activates yeast downstream promoters in vivo. *Cell* **55**(6): 1137-45.
18. Richmond TJ, Finch JT, Rushton B, Rhodes D, Klug A (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**(5986): 532-7.
19. Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ (1998) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**(6648): 251-60.
20. Arents G, Burlingame RW, Wang BC, Love WE, Moudrianakis EN (1991) The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci USA* **88**(22): 10148-10152.
21. Simpson RT (1978) Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones. *Biochemistry* **17**(25): 5524-31.
22. McGinty RK & Tan S (2015) Nucleosome structure and function. *Chem Rev* **115**(6): 2255-73.
23. du Preez LL, Patterton HG (2013) Secondary structures of the core histone N-terminal tails: their role in regulating chromatin structure. *Subcell Biochem* **61**: 37-55.
24. Zheng C, Hayes JJ (2003) Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. *J Biol Chem* **278**(26): 24217-24.
25. Barbera AJ, Chodaparambil JV, Kelley-Clarke B, Joukov V, Walter JC, Luger K, Kaye KM (2006) The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA. *Science* **311**(5762): 856-61.
26. Makde RD, England JR, Yennawar HP, Tan S (2010) Structure of RCC1 chromatin factor bound to the nucleosome core particle. *Nature* **467**(7315): 562-6.
27. Armache KJ, Garlick JD, Canzio D, Narlikar GJ, Kingston RE (2011) Structural basis of silencing: Sir3 BAH domain in complex with a nucleosome at 3.0 Å resolution. *Science* **334**(6058): 977-82.
28. Kato H, Jiang J, Zhou BR, Rozendaal M, Feng H, Ghirlando R, Xiao TS, Straight AF, Bai Y (2013) A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. *Science* **340**(6136): 1110-3.

29. McGinty RK, Henrici RC, Tan S (2014) Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. *Nature* **514**(7524): 591-6.
30. Morgan MT, Haj-Yahya M, Ringel AE, Bandi P, Brik A, Wolberger C (2016) Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**(6274): 725-8.
31. Wilson MD, Benlekhir S, Fradet-Turcotte A, Sherker A, Julien JP, McEwan A, Noordermeer SM, Sicheri F, Rubinstein JL, Durocher D (2016) The structural basis of modified nucleosome recognition by 53BP1. *Nature* **536**(7614): 100-3.
32. Mattioli F, Uckelmann M, Sahtoe DD, van Dijk WJ, Sixma TK (2014) The nucleosome acidic patch plays a critical role in RNF168-dependent ubiquitination of histone H2A. *Nat Commun* **5**: 3291.
33. White CL, Suto RK, Luger K (2001) Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *EMBO J* **20**(18): 5207-18.
34. Clapier CR, Chakravarthy S, Petosa C, Fernández-Tornero C, Luger K, Müller CW (2008) Structure of the Drosophila nucleosome core particle highlights evolutionary constraints on the H2A-H2B histone dimer. *Proteins* **71**(1): 1-7.
35. Tsunaka Y, Kajimura N, Tate S, Morikawa K (2005) Alteration of the nucleosomal DNA path in the crystal structure of a human nucleosome core particle. *Nucleic Acids Res* **33**(10): 3424-34.
36. Kurat CF, Recht J, Radovani E, Durbic T, Andrews B, Fillingham J (2014) Regulation of histone gene transcription in yeast. *Cell Mol Life Sci* **71**(4): 599-613.
37. Talbert PB, Henikoff S (2010) Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**(4): 264-75.
38. Venkatesh S, Workman JL (2015) Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* **16**(3): 178-89.
39. Buschbeck M, Hake SB (2017) Variants of core histones and their roles in cell fate decisions, development and cancer. *Nat Rev Mol Cell Biol* **18**(5): 299-314.
40. McKinley KL, Cheeseman IM (2016) The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol* **17**(1): 16-29.

41. Schneiderman JI, Orsi GA, Hughes KT, Loppin B, Ahmad K (2012) Nucleosome-depleted chromatin gaps recruit assembly factors for the H3.3 histone variant. *Proc Natl Acad Sci USA* **109**(48): 19721-6.
42. Ray-Gallet D, Woolfe A, Vassias I, Pellentz C, Lacoste N, Puri A, Schultz DC, Pchelintsev NA, Adams PD, Jansen LE, Almouzni G (2011) Dynamics of histone H3 deposition in vivo reveal a nucleosome gap-filling mechanism for H3.3 to maintain chromatin integrity. *Mol Cell* **44**(6): 928-41.
43. Podhorecka M, Skladanowski A, Bozko P (2010) H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *J Nucleic Acids* 2010: 920161.
44. Henikoff S, Smith MM (2015) Histone variants and epigenetics. *Cold Spring Harb Perspect Biol* **7**(1): a019364.
45. Tachiwana H, Kurumizaka H (2011) Structure of the CENP-A nucleosome and its implications for centromeric chromatin architecture. *Genes Genet Syst* **86**(6): 357-64.
46. Dechassa ML, Wyns K, Li M, Hall MA, Wang MD, Luger K (2011) Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat Commun* **2**: 313.
47. Tachiwana H, Kagawa W, Osakabe A, Kawaguchi K, Shiga T, Hayashi-Takanaka Y, Kimura H, Kurumizaka H (2010) Structural basis of instability of the nucleosome containing a testis-specific histone variant, human H3T. *Proc Natl Acad Sci USA* **107**(23): 10454-9
48. Starich MR, Sandman K, Reeve JN, Summers MF (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. *J Mol Biol* **255**(1): 187-203.
49. Decanniere K, Babu AM, Sandman K, Reeve JN, Heinemann U (2000) Crystal structures of recombinant histones HMfA and HMfB from the hyperthermophilic archaeon *Methanothermus fervidus*. *J Mol Biol* **303**(1): 35-47.
50. Bowman GD, Poirier MG (2015) Post-translational modifications of histones that influence nucleosome dynamics. *Chem Rev* **115**(6): 2274-95.
51. Zhao Z, Shilatifard A (2019) Epigenetic modifications of histones in cancer. *Genome Biol* **20**: 245.

52. Akhtar A, Becker PB (2000) Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol Cell* **5**(2): 367-75.
53. Shogren-Knaak M, Peterson CL (2006) Switching on chromatin: mechanistic role of histone H4-K16 acetylation. *Cell Cycle* **5**(13): 1361-5.
54. Lu X, Simon MD, Chodaparambil JV, Hansen JC, Shokat KM, Luger K (2008) The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. *Nat Struct Mol Biol* **15**(10): 1122-4.
55. Margueron R, Trojer P, Reinberg D (2005) The key to development: interpreting the histone code? *Curr Opin Genet Dev* **15**(2): 163-76.
56. Lawrence M, Daujat S, Schneider R (2016) Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet* **32**(1): 42-56.
57. Ausio J, Dong F, van Holde KE (1989) Use of selectively trypsinized nucleosome core particles to analyze the role of the histone "tails" in the stabilization of the nucleosome. *J Mol Biol* **206**(3): 451-63.
58. Tollervey JR, Lunyak VV (2012). Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics* **7**(8): 823-40.
59. Tropberger P, Schneider R (2013) Scratching the (lateral) surface of chromatin regulation by histone modifications. *Nat Struct Mol Biol* **20**(6): 657-61.
60. Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**(6796): 593-9.
61. Jørgensen S, Schotta G, Sørensen CS (2013) Histone H4 lysine 20 methylation: key player in epigenetic regulation of genomic integrity. *Nucleic Acids Res* **41**(5): 2797-806.
62. Di Lorenzo A, Bedford MT (2011) Histone arginine methylation. *FEBS Lett* **585**(13): 2024-2031.
63. Shi Y, Lan F, Matson C, Mulligan P, Whetstine JR, Cole PA, Casero RA, Shi Y (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**(7): 941-53.

64. Whetstine JR, Nottke A, Lan F, Huarte M, Smolikov S, Chen Z, Spooner E, Li E, Zhang G, Colaiacovo M, Shi Y (2006) Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**(3), 467-81.
65. Mosammaparast N, Shi Y (2010) Reversal of histone methylation: biochemical and molecular mechanisms of histone demethylases. *Annu Rev Biochem* **79**: 155-179
66. Parthun MR (2007) Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **26**(37): 5319-5328.
67. Hodawadekar SC, Marmorstein R (2007) Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. *Oncogene* **26**(37): 5528-40.
68. Yang XJ, Seto E (2007) HATs and HDACs: from structure, function and regulation to novel strategies for therapy and prevention. *Oncogene* **26**(37): 5310-8.
69. Tjeertes JV, Miller KM, Jackson SP (2009) Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells. *EMBO J* **28**(13): 1878-89.
70. Grant PA, Duggan L, Côté J, Roberts SM, Brownell JE, Candau R, Ohba R, Owen-Hughes T, Allis CD, Winston F, Berger SL, Workman JL (1997) Yeast Gcn5 functions in two multi-subunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes Dev* **11**(13): 1640-50.
71. Yang XJ, Seto E (2008) The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol* **9**(3): 206-18.
72. Oki M, Aihara H, Ito T (2007) Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Subcell Biochem* **41**: 319-36.
73. Hu S, Xie Z, Onishi A, Yu X, Jiang L, Lin J, Rho HS, Woodard C, Wang H, Jeong JS, Long S, He X, Wade H, Blackshaw S, Qian J, Zhu H (2009) Profiling the human protein-DNA interactome reveals ERK2 as a transcriptional repressor of interferon signaling. *Cell* **139**(3): 610-22.
74. Alaskhar Alhamwe B, Khalaila R, Wolf J, von Bülow V, Harb H, Alhamdan F, Hii CS, Prescott SL, Ferrante A, Renz H, Garn H, Potaczek DP (2018) Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy Asthma Clin Immunol* **14**:39.

75. Dawson MA, Bannister AJ, Göttgens B, Foster SD, Bartke T, Green AR, Kouzarides T (2009) JAK2 phosphorylates histone H3Y41 and excludes HP1alpha from chromatin. *Nature* **461**(7265): 819-22.
76. Sugiyama K, Sugiura K, Hara T, Sugimoto K, Shima H, Honda K, Furukawa K, Yamashita S, Urano T (2002) Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene* **21**(20): 3103-11.
77. Goto H, Yasui Y, Nigg EA, Inagaki M (2002) Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. *Genes Cells* **7**(1): 11-7.
78. Hassa PO, Haenni SS, Elser M, Hottiger MO (2006) Nuclear ADP-ribosylation reactions in mammalian cells: where are we today and where are we going? *Microbiol Mol Biol Rev* **70**(3): 789-829.
79. Cohen-Armon M, Visochek L, Rozensal D, Kalal A, Geistrikh I, Klein R, Bendetz-Nezer S, Yao Z, Seger R (2007) DNA-independent PARP-1 activation by phosphorylated ERK2 increases Elk1 activity: a link to histone acetylation. *Mol Cell* **25**(2): 297-308.
80. Krishnakumar R, Kraus WL (2010) PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol Cell* **39**(5): 736-49.
81. Hershko A, Ciechanover A (1998) The ubiquitin system. *Annu Rev Biochem* **67**:425-79.
82. Lee JS, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaumik SR, Shilatifard A (2007) Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. *Cell* **131**(6): 1084-96.
83. Kim J, Guermah M, McGinty RK, Lee JS, Tang Z, Milne TA, Shilatifard A, Muir TW, Roeder RG(2009) RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**(3):459-71.
84. Weake VM, Workman JL(2008) Histone ubiquitination: triggering gene activity. *Mol Cell* **29**(6):653-63. doi: 10.1016/j.molcel.2008.02.014.
85. Gatti M, Pinato S, Maspero E, Soffientini P, Polo S, Penengo L (2012) A novel ubiquitin mark at the N-terminal tail of histone H2As targeted by RNF168 ubiquitin ligase. *Cell Cycle* **11**(13): 2538-2544. doi:10.4161/cc.20919.

86. Wang Z, Zhang H, Liu J, Cheruiyot A, Lee JH, Ordog T, Lou Z, You Z, Zhang Z (2016) USP51 deubiquitylates H2AK13,15ub and regulates DNA damage response. *Genes Dev* **30**(8): 946-59.
87. Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, Zhang Y (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**(7010): 873-8.
88. Uckelmann M, Sixma TK (2017) Histone ubiquitination in the DNA damage response. *DNA Repair (Amst)* **56**:92-101.
89. Tamburri S, Lavarone E, Fernández-Pérez D, Conway E, Zanotti M, Manganaro D, Pasini D (2020) Histone H2AK119 Mono-Ubiquitination Is Essential for Polycomb-Mediated Transcriptional Repression. *Mol Cell* **77**(4): 840-856.e5.
90. Daou S, Hammond-Martel I, Mashtalir N, Barbour H, Gagnon J, Iannantuono NV, Nkwe NS, Motorina A, Pak H, Yu H, Wurtele H, Milot E, Mallette FA, Carbone M, Affar el B (2015) The BAP1/ASXL2 Histone H2A Deubiquitinase Complex Regulates Cell Proliferation and Is Disrupted in Cancer. *J Biol Chem* **290**(48): 28643-63.
91. Kalb R, Mallery DL, Larkin C, Huang JT, Hiom K (2014) BRCA1 is a histone-H2A-specific ubiquitin ligase. *Cell Rep* **8**(4): 999-1005.
92. Densham RM, Garvin AJ, Stone HR, Strachan J, Baldock RA, Daza-Martin M, Fletcher A, Blair-Reid S, Beesley J, Johal B, Pearl LH, Neely R, Keep NH, Watts FZ, Morris JR (2016) Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. *Nat Struct Mol Biol* **23**(7): 647-55.
93. Saredi G, Huang H, Hammond CM, Alabert C, Bekker-Jensen S, Forne I, Reverón-Gómez N, Foster BM, Mlejnkova L, Bartke T, Cejka P, Mailand N, Imhof A, Patel DJ, Groth A (2016) H4K20me0 marks post-replicative chromatin and recruits the TONSL–MMS22L DNA repair complex. *Nature* **534**(7609): 714-718.
94. Nakamura K, Kato A, Kobayashi J, Yanagihara H, Sakamoto S, Oliveira DV, Shimada M, Tauchi H, Suzuki H, Tashiro S, Zou L, Komatsu K (2011) Regulation of homologous recombination by RNF20-dependent H2B ubiquitination. *Mol Cell* **41**(5): 515-28.
95. Zheng S, Li D, Lu Z, Liu G, Wang M, Xing P, Wang M, Dong Y, Wang X, Li J, Zhang S, Peng H, Ira G, Li G, Chen X (2018) Bre1-dependent H2B ubiquitination

- promotes homologous recombination by stimulating histone eviction at DNA breaks. *Nucleic Acids Res* **46**(21): 11326-11339.
96. Chernikova SB, Razorenova OV, Higgins JP, Sishc BJ, Nicolau M, Dorth JA, Chernikova DA, Kwok S, Brooks JD, Bailey SM, Game JC, Brown JM (2012) Deficiency in mammalian histone H2B ubiquitin ligase Bre1 (Rnf20/Rnf40) leads to replication stress and chromosomal instability. *Cancer Res* **72**(8): 2111-9.
 97. Northam MR, Trujillo KM (2016) Histone H2B mono-ubiquitylation maintains genomic integrity at stalled replication forks. *Nucleic Acids Res* **44**(19):9245-9255.
 98. Atanassov BS, Mohan RD, Lan X, Kuang X, Lu Y, Lin K, McIvor E, Li W, Zhang Y, Florens L, Byrum SD, Mackintosh SG, Calhoun-Davis T, Koutelou E, Wang L, Tang DG, Tackett AJ, Washburn MP, Workman JL, Dent SY (2016) ATXN7L3 and ENY2 coordinate activity of multiple H2B deubiquitinases important for cellular proliferation and tumor growth. *Mol Cell* **62**(4):558-71.
 99. Ai H, Guo Y, Sun D, Liu S, Qi Y, Guo J, Qu Q, Gong Q, Zhao S, Li J, Liu L (2019) Examination of the deubiquitylation site selectivity of USP51 by using chemically synthesized ubiquitylated histones. *Chembiochem* **20**(2): 221-229.
 100. Qin W, Wolf P, Liu N, Link S, Smets M, La Mastra F, Forné I, Pichler G, Hörl D, Fellingner K, Spada F, Bonapace IM, Imhof A, Harz H, Leonhardt H (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res* **25**(8):911-29.
 101. Li T, Wang L, Du Y, Xie S, Yang X, Lian F, Zhou Z, Qian C (2018) Structural and mechanistic insights into UHRF1-mediated DNMT1 activation in the maintenance DNA methylation. *Nucleic Acids Res* **46**(6):3218-3231.
 102. Oya E, Nakagawa R, Yoshimura Y, Tanaka M, Nishibuchi G, Machida S, Shirai A, Ekwall K, Kurumizaka H, Tagami H, Nakayama JI (2019) H3K14 ubiquitylation promotes H3K9 methylation for heterochromatin assembly. *EMBO Rep* **20**(10): e48111.
 103. Zhang X, Li B, Rezaeian AH, Xu X, Chou PC, Jin G, Han F, Pan BS, Wang CY, Long J, Zhang A, Huang CY, Tsai FJ, Tsai CH, Logothetis C, Lin HK (2017) H3 ubiquitination by NEDD4 regulates H3 acetylation and tumorigenesis. *Nat Commun* **8**: 14799.

104. Han J, Zhang H, Zhang H, Wang Z, Zhou H, Zhang Z (2013) A Cul4 E3 ubiquitin ligase regulates histone hand-off during nucleosome assembly. *Cell* **155**(4):817-29.
105. Yan Q, Dutt S, Xu R, Graves K, Juszczynski P, Manis JP, Shipp MA (2009) BBAP monoubiquitylates histone H4 at lysine 91 and selectively modulates the DNA damage response. *Mol Cell* **36**(1):110-20.
106. Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, Glass CK, Rosenfeld MG (2008) Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **29**(1):69-80.
107. Mattioli F, Vissers JH, van Dijk WJ, Ikpa P, Citterio E, Vermeulen W, Marteijn JA, Sixma TK (2012) RNF168 ubiquitinates K13-15 on H2A/H2AX to drive DNA damage signaling. *Cell* **150**(6): 1182-95.
108. Horn V, Uckelmann M, Zhang H, Eerland J, Aarsman I, le Paige UB, Davidovich C, Sixma TK, van Ingen H (2019) Structural basis of specific H2A K13/K15 ubiquitination by RNF168. *Nat Commun* **10**(1):1751.
109. Fang J, Chen T, Chadwick B, Li E, Zhang Y (2004) Ring1b-mediated H2A ubiquitination associates with inactive X chromosomes and is involved in initiation of X inactivation. *J Biol Chem* **279**(51): 52812-5.
110. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129**(4): 823-37.
111. Hu Q, Botuyan MV, Cui G, Zhao D, Mer G (2017) Mechanisms of Ubiquitin-Nucleosome Recognition and Regulation of 53BP1 Chromatin Recruitment by RNF168/169 and RAD18. *Mol Cell* **66**(4): 473-487.e9.
112. Witus SR, Burrell AL, Farrell DP, Kang J, Wang M, Hansen JM, Pravat A, Tuttle LM, Stewart MD, Brzovic PS, Chatterjee C, Zhao W, DiMaio F, Kollman JM, Klevit RE (2021) BRCA1/BARD1 site-specific ubiquitylation of nucleosomal H2A is directed by BARD1. *Nat Struct Mol Biol* **28**(3): 268-277.
113. West MH, Bonner WM (1980) Histone 2B can be modified by the attachment of ubiquitin. *Nucleic Acids Res* **8**(20): 4671-80.
114. Robzyk K, Recht J, Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**(5452): 501-4.

115. Wood A, Krogan NJ, Dover J, Schneider J, Heidt J, Boateng MA, Dean K, Golshani A, Zhang Y, Greenblatt JF, Johnston M, Shilatifard A (2003) Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell* **11**(1): 267-74.
116. Hwang WW, Venkatasubrahmanyam S, Ianculescu AG, Tong A, Boone C, Madhani HD (2003) A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell* **11**(1): 261-6.
117. Zhu B, Zheng Y, Pham AD, Mandal SS, Erdjument-Bromage H, Tempst P, Reinberg D (2005) Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol Cell* **20**: 601–611.
118. Tweedie-Cullen RY, Reck JM, Mansuy IM (2009) Comprehensive mapping of post-translational modifications on synaptic, nuclear, and histone proteins in the adult mouse brain. *J Proteome Res* **8**: 4966–4982.
119. Geng F, Tansey WP (2008) Polyubiquitylation of histone H2B. *Mol Biol Cell* **19**: 3616–3624.
120. Chatterjee C, McGinty RK, Fierz B, Muir TW (2010) Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat Chem Biol* **6**: 267–269.
121. Fierz B, Chatterjee C, McGinty RK, Bar-Dagan M, Raleigh DP, Muir TW (2011) Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat Chem Biol* **7**: 113–119.
122. Fierz B, Kilic S, Hieb AR, Luger K, Muir TW (2012) Stability of nucleosomes containing homogenously ubiquitylated H2A and H2B prepared using semisynthesis. *J Am Chem Soc* **134**: 19548– 19551.
123. Chandrasekharan MB, Huang F, Sun ZW (2009) Ubiquitination of histone H2B regulates chromatin dynamics by enhancing nucleosome stability. *Proc Natl Acad Sci USA* **106**: 16686– 16691.
124. Fleming AB, Kao CF, Hillyer C, Pikaart M, Osley MA (2008) H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell* **31**: 57–66.
125. Worden EJ, Hoffmann NA, Hicks CW, Wolberger C (2019) Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. *Cell* **176**(6): 1490-1501.e12.

126. Schneider J, Wood A, Lee JS, Schuster R, Dueker J, Maguire C, Swanson SK, Florens L, Washburn MP, Shilatifard A (2005) Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* **19**: 849–856.
127. Shahbazian MD, Zhang K, Grunstein M (2005) Histone H2B ubiquitylation controls processive methylation but not monomethylation by Dot1 and Set1. *Mol Cell* **19**: 271–277.
128. Sun ZW, Allis CD (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature* **418**: 104–108.
129. Anderson CJ, Baird MR, Hsu A, Barbour EH, Koyama Y, Borgnia MJ, McGinty RK (2019) Structural Basis for Recognition of Ubiquitylated Nucleosome by Dot1L Methyltransferase. *Cell Rep* **26**(7): 1681-1690.e5.
130. Valencia-Sánchez MI, De Ioannes P, Wang M, Vasilyev N, Chen R, Nudler E, Armache JP, Armache KJ (2019) Structural Basis of Dot1L Stimulation by Histone H2B Lysine 120 Ubiquitination. *Mol Cell* **74**(5): 1010-1019.e6.
131. Jang S, Kang C, Yang HS, Jung T, Hebert H, Chung KY, Kim SJ, Hohng S, Song JJ (2019) Structural basis of recognition and destabilization of the histone H2B ubiquitinated nucleosome by the DOT1L histone H3 Lys79 methyltransferase. *Genes Dev* **33**(11-12): 620-625.
132. Hsu PL, Shi H, Leonen C, Kang J, Chatterjee C, Zheng N (2019) Structural Basis of H2B Ubiquitination-Dependent H3K4 Methylation by COMPASS. *Mol Cell* **76**(5): 712-723.e4.
133. Worden EJ, Zhang X, Wolberger C (2020) Structural basis for COMPASS recognition of an H2B-ubiquitinated nucleosome. *Elife* **9**: e53199.
134. Schulze JM, Hentrich T, Nakanishi S, Gupta A, Emberly E, Shilatifard A, Kobor MS (2011) Splitting the task: Ubp8 and Ubp10 deubiquitinate different cellular pools of H2BK123. *Genes Dev* **25**: 2242–2247.
135. Giannattasio M, Lazzaro F, Plevani P, Muzi-Falconi M (2005) The DNA damage checkpoint response requires histone H2B ubiquitination by Rad6-Bre1 and H3 methylation by Dot1. *J Biol Chem* **280**: 9879–9886.
136. Nakamura K, Kato A, Kobayashi J, Yanagihara H, Sakamoto S, Oliveira DV, Shimada M, Tauchi H, Suzuki H, Tashiro S, Zou L, Komatsu K (2011) Regulation of

- homologous recombination by RNF20- dependent H2B ubiquitination. *Mol Cell* **41**: 515–528.
137. Moyal L, Lerenthal Y, Gana-Weisz M, Mass G, So S, Wang SY, Eppink B, Chung YM, Shalev G, Shema E, Shkedy D, Smorodinsky NI, van Vliet N, Kuster B, Mann M, Ciechanover A, Dahm-Daphi J, Kanaar R, Hu MC, Chen DJ, Oren M, Shiloh Y (2011) Requirement of ATM-dependent monoubiquitylation of histone H2B for timely repair of DNA double-strand breaks. *Mol Cell* **41**: 529–542.
 138. Pavri R, Zhu B, Li G, Trojer P, Mandal S, Shilatifard A, Reinberg D (2006) Histone H2B monoubiquitination functions cooperatively with FACT to regulate elongation by RNA polymerase II. *Cell* **125**: 703–717.
 139. Mao P, Meas R, Dorgan KM, Smerdon MJ (2014) UV damage-induced RNA polymerase II stalling stimulates H2B deubiquitylation. *Proc Natl Acad Sci USA* **111**: 12811–12816.
 140. Morgan MT, Haj-Yahya M, Ringel AE, Bandi P, Brik A, Wolberger C (2016) Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **351**(6274): 725-8.
 141. Wang H, Dienemann C, Stützer A, Urlaub H, Cheung ACM, Cramer P (2020) Structure of the transcription coactivator SAGA. *Nature* **577** (7792): 717-720.
 142. Henry KW, Wyce A, Lo WS, Duggan LJ, Emre NC, Kao CF, Pillus L, Shilatifard A, Osley MA, Berger SL (2003) Transcriptional activation via sequential histone H2B ubiquitylation and deubiquitylation, mediated by SAGA-associated Ubp8. *Genes Dev* **17**(21): 2648-63.
 143. Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* **194**: 531–544.
 144. Hershko A, Heller H, Elias S, Ciechanover A (1983) Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown. *J Biol Chem* **258**: 8206– 8214.
 145. Kwon YT, Ciechanover A (2017) The Ubiquitin Code in the Ubiquitin-Proteasome System and Autophagy. *Trends Biochem Sci* **42**: 873–886.
 146. Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch, S (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**: 135–141.

147. Watanabe K, Tateishi S, Kawasuji M, Tsurimoto T, Inoue H, Yamaizumi M (2004) Rad18 guides pol-eta to replication stalling sites through physical interaction and PCNA monoubiquitination. *EMBO J* **23**: 3886–3896.
148. Yang WL, Zhang X, Lin HK (2010) Emerging role of Lys-63 ubiquitination in protein kinase and phosphatase activation and cancer development. *Oncogene* **29**: 4493–4503.
149. Komander D, Rape M (2012) The ubiquitin code. *Annu Rev Biochem* **81**: 203–29.
150. Kliza K, Husnjak K (2020) Resolving the Complexity of Ubiquitin Networks. *Front Mol Biosci* **7**: 21.
151. Ciechanover A, Heller H, Katz-Etzion R, Hershko A (1981) Activation of the heat-stable polypeptide of the ATP-dependent proteolytic system. *Proc Natl Acad Sci USA* **78**: 761–765.
152. Handley PM, Mueckler M, Siegel NR, Ciechanover A, Schwartz A (1991). Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proc Natl Acad Sci USA* **88**: 258–262.
153. McGrath JP, Jentsch S, Varshavsky A (1991) UBA1: an essential yeast gene encoding ubiquitin activating enzyme. *EMBO J* **10**: 227–236.
154. Huang DT, Hunt HW, Zhuang M, Ohi MD, Holton JM, Schulman BA (2007) Basis for a ubiquitin-like protein thioester switch toggling E1-E2 affinity. *Nature* **445**: 394–398.
155. Lois LM, Lima CD (2005) Structures of the SUMO E1 provide mechanistic insights into SUMO activation and E2 recruitment to E1. *EMBO J* **24**: 439–451.
156. Lee I, Schindelin H (2008) Structural insights into E1-catalyzed ubiquitin activation and transfer to conjugating enzymes. *Cell* **134**: 268–278.
157. Haas AL, Rose IA (1982) The mechanism of ubiquitin activating enzyme. A kinetic and equilibrium analysis. *J Biol Chem* **257**: 10329–10337.
158. Pickart CM, Kasperek EM, Beal R, Kim A (1994) Substrate properties of site-specific mutant ubiquitin protein (G76A) reveal unexpected mechanistic features of ubiquitin-activating enzyme (E1). *J Biol Chem* **269**, 7115–7123.
159. Olsen SK, Lima CD (2013) Structure of a ubiquitin E1-E2 complex: insights to E1-E2 thioester transfer. *Mol Cell* **49**: 884–896.

160. Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS (2016) E2 enzymes: more than just middle men. *Cell Res* **26**(4):423-40.
161. Wu PY, Hanlon M, Eddins M, Tsui C, Rogers RS, Jensen JP, Matunis MJ, Weissman AM, Wolberger C, Pickart CM (2003) A conserved catalytic residue in the ubiquitin-conjugating enzyme family. *EMBO J* **22**: 5241–5250.
162. Deshaies RJ, Joazeiro CA (2009) RING domain E3 ubiquitin ligases. *Annu Rev Biochem* **78**: 399-434.
163. Huibregtse JM, Scheffner M, Beaudenon S, Howley PM (1995) A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase. *Proc Natl Acad Sci USA* **92**, 2563– 2567.
164. Rotin D, Kumar S (2009) Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* **10**, 398–409.
165. Huang A, Hibbert RG, de Jong RN, Das D, Sixma TK, Boelens R (2011). Symmetry and asymmetry of the RING-RING dimer of Rad18. *J Mol Biol* **410**, 424–435.
166. Verdecia MA, Joazeiro CA, Wells NJ, Ferrer JL, Bowman ME, Hunter T, Noel JP (2003). Conformational flexibility underlies ubiquitin ligation mediated by the WWP1 HECT domain E3 ligase. *Mol Cell* **11**, 249–259.
167. Tang D, Xiang Y, De Renzis S, Rink J, Zheng G, Zerial M, Wang Y (2011) The ubiquitin ligase HACE1 regulates Golgi membrane dynamics during the cell cycle. *Nat Commun* **2**: 501.
168. Hao Z, Duncan GS, Su YW, Li WY, Silvester J, Hong C, You H, Brenner D, Gorrini C, Haight J, Wakeham A, You-Ten A, McCracken S, Elia A, Li Q, Detmar J, Jurisicova A, Hobeika E, Reth M, Sheng Y, Lang PA, Ohashi PS, Zhong Q, Wang X, Mak TW (2012) The E3 ubiquitin ligase Mule acts through the ATM-p53 axis to maintain B lymphocyte homeostasis. *J Exp Med* **209**: 173–186.
169. Kamadurai HB, Qiu Y, Deng A, Harrison JS, Macdonald C, Actis M, Rodrigues P, Miller DJ, Souphron J, Lewis SM, Kurinov I, Fujii N, Hammel M, Piper R, Kuhlman B, Schulman BA (2013) Mechanism of ubiquitin ligation and lysine prioritization by a HECT E3. *Elife* **2**: e00828.
170. Freemont PS (1993) The RING finger. A novel protein sequence motif related to the zinc finger. *Ann NY Acad Sci* **684**: 174–192.

171. Joazeiro CA, Wing SS, Huang H, Leverson JD, Hunter T, Liu YC (1999). The tyrosine kinase negative regulator c-Cbl is a RING-type, E2-dependent ubiquitin-protein ligase. *Science* **286**: 309– 312.
172. Kamura T, Conrad MN, Yan Q, Conaway RC, Conaway JW (1999) The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2. *Genes Dev* **13**: 2928– 2933.
173. Lorick KL, Jensen JP, Fang S, Ong AM, Hatakeyama S, Weissman AM (1999) RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. *Proc Natl Acad Sci USA* **96**: 11364–69.
174. Ohta T, Michel JJ, Schottelius AJ, Xiong Y (1999) ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol Cell* **3**: 535–541.
175. Christensen DE, Brzovic PS, Klevit RE (2007) E2-BRCA1 RING interactions dictate synthesis of mono- or specific polyubiquitin chain linkages. *Nat Struct Mol Biol* **14**: 941–48.
176. Plechanovova A, Jaffray EG, Tatham MH, Naismith JH, Hay RT (2012) Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**: 115-120.
177. Plechanovova A, Jaffray EG, McMahon SA, Johnson KA, Navratilova I, Naismith JH, Hay RT (2011) Mechanism of ubiquitylation by dimeric RING ligase RNF4. *Nat Struct Mol Biol* **18**: 1052-1059.
178. Galanty Y, Belotserkovskaya R, Coates J, Jackson SP (2012) RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev* **26**: 1179-1195.
179. Koegl M, Hoppe T, Schlenker S, Ulrich HD, Mayer TU, Jentsch S (1999) A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* **96**: 635– 644.
180. Hatakeyama S, Yada M, Matsumoto M, Ishida N, Nakayama KI (2001) U box proteins as a new family of ubiquitin-protein ligases. *J Biol Chem* **276**: 33111–33120.
181. Beasley SA, Hristova VA, Shaw GS (2007) Structure of the Parkin in-between-ring domain provides insights for E3-ligase dysfunction in autosomal recessive Parkinson's disease. *Proc Natl Acad Sci USA* **104**(9): 3095-100.

182. Duda DM, Olszewski JL, Schuermann JP, Kurinov I, Miller DJ, Nourse A, Alpi AF, Schulman BA (2013) Structure of HHARI, a RING-IBR-RING ubiquitin ligase: autoinhibition of an Ariadne-family E3 and insights into ligation mechanism. *Structure* **21**(6): 1030-41.
183. Kumar A, Aguirre JD, Condos TE, Martinez-Torres RJ, Chaugule VK, Toth R, Sundaramoorthy R, Mercier P, Knebel A, Spratt DE, Barber KR, Shaw GS, Walden H (2015) Disruption of the autoinhibited state primes the E3 ligase parkin for activation and catalysis. *EMBO J* **34**(20): 2506-21.
184. Lechtenberg BC, Rajput A, Sanishvili R, Dobaczewska MK, Ware CF, Mace PD, Riedl SJ (2016) Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature* **529**(7587): 546-50.
185. Tokunaga F, Sakata S, Saeki Y, Satomi Y, Kirisako T, Kamei K, Nakagawa T, Kato M, Murata S, Yamaoka S, Yamamoto M, Akira S, Takao T, Tanaka K, Iwai K (2009) Involvement of linear polyubiquitylation of NEMO in NF-kappaB activation. *Nat Cell Biol* **11**(2): 123-32.
186. Clark IE, Dodson MW, Jiang C, Cao JH, Huh JR, Seol JH, Yoo SJ, Hay BA, Guo M (2006) Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. *Nature* **441**(7097): 1162-6.
187. Tan NG, Ardley HC, Scott GB, Rose SA, Markham AF, Robinson PA (2003) Human homologue of ariadne promotes the ubiquitylation of translation initiation factor 4E homologous protein, 4EHP. *FEBS Lett* **554**(3): 501-4.
188. Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**: 273-304.
189. Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**(6765):41-5.
190. Kim J, Roeder RG (2009) Direct Bre1-Paf1 complex interactions and RING finger-independent Bre1-Rad6 interactions mediate histone H2B ubiquitylation in yeast. *J Biol Chem* **284**(31): 20582-92.
191. Kim J, Hake SB & Roeder RG (2005) The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol Cell* **20**: 759–770.

192. Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N, Raver-Shapira N, Minsky N, Pirngruber J, Tarcic G, Hublarova P, Moyal L, Gana-Weisz M, Shiloh Y, Yarden Y, Johnsen SA, Vojtesek B, Berger SL, Oren M (2008) The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* **22**(19): 2664-76.
193. Foglizzo M, Middleton AJ, Day CL (2016) Structure and Function of the RING Domains of RNF20 and RNF40, Dimeric E3 Ligases that Monoubiquitylate Histone H2B. *J Mol Biol* **428**(20): 4073-4086.
194. Schindler U, Beckmann H, Cashmore AR (1993) HAT3.1, a novel Arabidopsis homeodomain protein containing a conserved cysteine-rich region. *Plant J* **4**: 137 – 150.
195. Musselman CA, Kutateladze TG (201) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* **39**(21): 9061-71.
196. Li H, Ilin S, Wang W, Duncan EM, Wysocka J, Allis CD, Patel DJ (2006) Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature* **442**(7098): 91-5.
197. Peña PV, Davrazou F, Shi X, Walter KL, Verkhusha VV, Gozani O, Zhao R, Kutateladze TG (2006) Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **442**(7098): 100-3.
198. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, Carney D, Peña P, Lan F, Kaadige MR, Lacoste N, Cayrou C, Davrazou F, Saha A, Cairns BR, Ayer DE, Kutateladze TG, Shi Y, Côté J, Chua KF, Gozani O (2006) ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**(7098): 96-9.
199. Wysocka J, Swigut T, Xiao H, Milne TA, Kwon SY, Landry J, Kauer M, Tackett AJ, Chait BT, Badenhorst P, Wu C, Allis CD (2006) A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**(7098): 86-90.
200. Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**(1): 58-69.
201. Matthews AG, Kuo AJ, Ramón-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN, Walter KL, Utz PJ, Shi Y,

- Kutateladze TG, Yang W, Gozani O, Oettinger MA (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**(7172): 1106-10.
202. Ramón-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, Yang W (2007) The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proc Natl Acad Sci USA* **104**(48): 18993-8.
 203. Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S (2007) A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**(4): 561-71.
 204. Fiedler M, Sánchez-Barrena MJ, Nekrasov M, Mieszczanek J, Rybin V, Müller J, Evans P, Bienz M (2008) Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Mol Cell* **30**(4): 507-18.
 205. Champagne KS, Saksouk N, Peña PV, Johnson K, Ullah M, Yang XJ, Côté J, Kutateladze TG (2008) The crystal structure of the ING5 PHD finger in complex with an H3K4me3 histone peptide. *Proteins* **72**(4): 1371-6.
 206. Palacios A, Muñoz IG, Pantoja-Uceda D, Marcaida MJ, Torres D, Martín-García JM, Luque I, Montoya G, Blanco FJ (2008) Molecular basis of histone H3K4me3 recognition by ING4. *J Biol Chem* **283**(23): 15956-64.
 207. Peña PV, Hom RA, Hung T, Lin H, Kuo AJ, Wong RP, Subach OM, Champagne KS, Zhao R, Verkhusha VV, Li G, Gozani O, Kutateladze TG (2008) Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *J Mol Biol* **380**: 303–312.
 208. van Ingen H, van Schaik FM, Wienk H, Ballering J, Rehmann H, Dechesne AC, Kruijzer JA, Liskamp RM, Timmers HT, Boelens R (2008) Structural insight into the recognition of the H3K4me3 mark by the TFIID subunit TAF3. *Structure* **16**: 1245–1256.
 209. Hung T, Binda O, Champagne KS, Kuo AJ, Johnson K, Chang HY, Simon MD, Kutateladze TG, Gozani O (2009) ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Mol Cell* **33**: 248–256.

210. Wang GG, Song J, Wang Z, Dormann HL, Casadio F, Li H, Luo JL, Patel DJ, Allis CD (2009) Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**: 847–851.
211. Wen H, Li J, Song T, Lu M, Kan PY, Lee MG, Sha B, Shi X (2010) Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J Biol Chem* **285**: 9322–9326.
212. Horton JR, Upadhyay AK, Qi HH, Zhang X, Shi Y, Cheng X (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol* **17**: 38–43.
213. Wang Z, Song J, Milne TA, Wang GG, Li H, Allis CD, Patel DJ (2010) Pro isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CYP33 and HDAC-mediated repression. *Cell* **141**: 1183–1194.
214. Chang PY, Hom RA, Musselman CA, Zhu L, Kuo A, Gozani O, Kutateladze TG, Cleary ML (2010) Binding of the MLL PHD3 finger to histone H3K4me3 is required for MLL-dependent gene transcription. *J Mol Biol* **400**: 137–144.
215. Park S, Osmers U, Raman G, Schwantes RH, Diaz MO, Bushweller JH (2010) The PHD3 domain of MLL acts as a CYP33-regulated switch between MLL-mediated activation and repression. *Biochemistry* **49**: 6576–6586.
216. Yang Y, Hu L, Wang P, Hou H, Lin Y, Liu Y, Li Z, Gong R, Feng X, Zhou L, Zhang W, Dong Y, Yang H, Lin H, Wang Y, Chen CD, Xu Y (2010) Structural insights into a dual-specificity histone demethylase ceKDM7A from *Caenorhabditis elegans*. *Cell Res* **20**: 886–898.
217. Miller TC, Rutherford J, Johnson CM, Fiedler M, Bienz M (2010) Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor. *J Mol Biol* **401**: 969–984.
218. Lan F, Collins RE, De Cegli R, Alpatov R, Horton JR, Shi X, Gozani O, Cheng X, Shi Y (2007) Recognition of unmethylated histone H3 lysine 4 links BHC80 to LSD1-mediated gene repression. *Nature* **448**: 718–722.
219. Org T, Chignola F, Hetenyi C, Gaetani M, Rebane A, Liiv I, Maran U, Mollica L, Bottomley MJ, Musco G, Peterson P (2008) The autoimmune regulator PHD finger binds to non-methylated histone H3K4 to activate gene expression. *EMBO Rep* **9**: 370–376.

220. Koh AS, Kuo AJ, Park SY, Cheung P, Abramson J, Bua D, Carney D, Shoelson SE, Gozani O, Kingston RE, Benoist C, Mathis D (2008) Aire employs a histone-binding module to mediate immunological tolerance, linking chromatin regulation with organ-specific autoimmunity. *Proc Natl Acad Sci USA* **105**: 15878–15883.
221. Lange M, Kaynak B, Forster UB, Tonjes M, Fischer JJ, Grimm C, Schlesinger J, Just S, Dunkel I, Krueger T, Mebus S, Lehrach H, Lurz R, Gobom J, Rottbauer W, Abdelilah-Seyfried S, Sperling S (2008) Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes Dev* **22**: 2370–2384.
222. Chignola F, Gaetani M, Rebane A, Org T, Mollica L, Zucchelli C, Spitaleri A, Mannella V, Peterson P, Musco G (2009) The solution structure of the first PHD finger of autoimmune regulator in complex with non-modified histone H3 tail reveals the antagonistic role of H3R2 methylation. *Nucleic Acids Res* **37**: 2951–2961.
223. Chakravarty S, Zeng L, Zhou MM (2009) Structure and site-specific recognition of histone H3 by the PHD finger of human autoimmune regulator. *Structure* **17**:670–679.
224. Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M (2009) Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* **10**: 1235–1241.
225. Musselman CA, Mansfield RE, Garske AL, Davrazou F, Kwan AH, Oliver SS, O'Leary H, Denu JM, Mackay JP, Kutateladze TG (2009) Binding of the CHD4 PHD2 finger to histone H3 is modulated by covalent modifications. *Biochem J* **423**: 179–187.
226. Saksouk N, Avvakumov N, Champagne KS, Hung T, Doyon Y, Cayrou C, Paquet E, Ullah M, Landry AJ, Cote V, Yang XJ, Gozani O, Kutateladze TG, Côté J (2009) HBO1 HAT complexes target chromatin throughout gene coding regions via multiple PHD finger interactions with histone H3 tail. *Mol Cell* **33**: 257–265.
227. Zeng L, Zhang Q, Li S, Plotnikov AN, Walsh MJ, Zhou MM (2010) Mechanism and regulation of acetylated histone binding by the tandem PHD finger of DPF3b. *Nature* **466**: 258–262
228. Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, Tsai CY, Shi X, Schwarzer D, Plunkett W, Aronow B, Gozani O, Fischle W, Hung MC, Patel DJ,

- Barton MC (2010) TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**: 927–932.
229. Mansfield RE, Musselman CA, Kwan AH, Oliver SS, Garske AL, Davrazou F, Denu JM, Kutateladze TG, Mackay JP (2011) Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem* **286**: 11779–11791.
230. Iwase S, Xiang B, Ghosh S, Ren T, Lewis PW, Cochrane JC, Allis CD, Picketts DJ, Patel DJ, Li H, Shi Y (2011) ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* **18**: 769–776.
231. Iwase S, Lan F, Bayliss P, de la Torre-Ubieta L, Huarte M, Qi HH, Whetstine JR, Bonni A, Roberts TM, Shi Y (2007) The X-linked mental retardation gene SMCX/JARID1C defines a family of histone H3 lysine 4 demethylases. *Cell* **128**: 1077–1088.
232. Li F, Huarte M, Zaratiegui M, Vaughn MW, Shi Y, Martienssen R, Cande WZ (2008) Lid2 is required for coordinating H3K4 and H3K9 methylation of heterochromatin and euchromatin. *Cell* **135**: 272–283.
233. Shi X, Kachirskaia I, Walter KL, Kuo JH, Lake A, Davrazou F, Chan SM, Martin DG, Fingerman IM, Briggs SD, Howe L, Utz PJ, Kutateladze TG, Lugovskoy AA, Bedford MT, Gozani O (2007) Proteome-wide analysis in *Saccharomyces cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem* **282**: 2450–2455.
234. Ruthenburg AJ, Allis CD, Wysocka J (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell* **25**: 15 – 30.
235. Zhang Y (2006) It takes a PHD to interpret histone methylation. *Nat Struct Mol Biol* **13**(7):572-4.
236. Kirmizis A, Santos-Rosa H, Penkett CJ, Singer MA, Vermeulen M, Mann M, Bähler J, Green RD, Kouzarides T (2007) Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449**(7164): 928-32.
237. Wang ZX, Patel DJ (2011) Combinatorial readout of dual histone modifications by paired chromatin-associated modules. *J Biol Chem* **286**: 18363– 18368.

238. Ruthenburg AJ, Li H, Milne TA, Dewell S, McGinty RK, Yuen M, Ueberheide B, Dou Y, Muir TW, Patel DJ, Allis CD (2011) Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. *Cell* **145**(5): 692-706.
239. Zeng L, Yap KL, Ivanov AV, Wang X, Mujtaba S, Plotnikova O, Rauscher FJ 3rd, Zhou MM (2008) Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* **15**(6):626-33.
240. Prenzel T, Begus-Nahrmann Y, Kramer F, Hennion M, Hsu C, Gorsler T, Hintermair C, Eick D, Kremmer E, Simons M, Beissbarth T, Johnsen SA (2011) Estrogen-dependent gene transcription in human breast cancer cells relies upon proteasome-dependent monoubiquitination of histone H2B. *Cancer Res* **71**(17):5739-53.
241. Urasaki Y, Heath L, Xu CW (2012) Coupling of glucose deprivation with impaired histone H2B monoubiquitination in tumors. *PLoS ONE* **7**: e36775.
242. Hahn MA, Dickson KA, Jackson S, Clarkson A, Gill AJ & Marsh DJ (2012) The tumor suppressor CDC73 interacts with the ring finger proteins RNF20 and RNF40 and is required for the maintenance of histone 2B monoubiquitination. *Human Molecular Genetics* **21**: 559–568.
243. Shema E, Tirosh I, Aylon Y, Huang J, Ye C, Moskovits N, Raver-Shapira N, Minsky N, Pirngruber J, Tarcic G, Hublarova P, Moyal L, Gana-Weisz M, Shiloh Y, Yarden Y, Johnsen SA, Vojtesek B, Berger SL, Oren M (2008) The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev* **22**(19): 2664-76.
244. Shema E, Kim J, Roeder RG, Oren M (2011) RNF20 inhibits TFIIS facilitated transcriptional elongation to suppress pro-oncogenic gene expression. *Molecular Cell* **42**: 477–488.
245. Varambally S, Yu J, Laxman B, Rhodes DR, Mehra R, Tomlins SA, Shah RB, Chandran U, Monzon FA, Becich MJ, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM (2005) Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**(5): 393-406.
246. Glinsky GV, Berezovska O, Glinskii AB (2005) Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. *J Clin Invest* **115**: 1503–1521.

247. Li J, Olson LM, Zhang Z, Li L, Bidder M, Nguyen L, Pfeifer J, Rader JS (2008) Differential display identifies overexpression of the USP36 gene, encoding a deubiquitinating enzyme, in ovarian cancer. *Int J Med Sci* **5**(3): 133-42.
248. Hochstrasser M, Varshavsky A (1990) In vivo degradation of a transcriptional regulator: The yeast alpha 2 repressor. *Cell* **61**: 697–708.
249. Tasaki T, Mulder LC, Iwamatsu A, Lee MJ, Davydov IV, Varshavsky A, Muesing M, Kwon YT (2005) A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**(16): 7120-36.
250. Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS, Kwon YT (2009) The substrate recognition domains of the N-end rule pathway. *J Biol Chem* **284**(3): 1884-95.
251. Xie Y, Varshavsky A (1999) The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. *EMBO J* **18**(23): 6832-44.
252. Pan M, Zheng Q, Wang T, Liang L, Mao J, Zuo C, Ding R, Ai H, Xie Y, Si D, Yu Y, Liu L, Zhao M (2021) Structural insights into Ubr1-mediated N-degron polyubiquitination. *Nature* **600**(7888): 334-338.
253. An JY, Kim EA, Jiang Y, Zakrzewska A, Kim DE, Lee MJ, Mook-Jung I, Zhang Y, Kwon YT (2010) UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. *Proc Natl Acad Sci USA* **107**(5): 1912-7.
254. Qian H, Zhang Y, Wu B, Wu S, You S, Zhang N, Sun Y (2020) Structure and Function of HECT E3 Ubiquitin Ligases and their Role in Oxidative Stress. *J Transl Int Med* **8**(2): 71-79.
255. Henderson MJ, Russell AJ, Hird S, Munoz M, Clancy JL, Lehrbach GM, Calanni ST, Jans DA, Sutherland RL, Watts CK (2002) EDD, the human hyperplastic discs protein, has a role in progesterone receptor coactivation and potential involvement in DNA damage response. *J Biol Chem* **277**: 26468–78.
256. Shearer RF, Iconomou M, Watts CK, Saunders DN (2015) Functional roles of the E3 ubiquitin ligase UBR5 in cancer. *Mol Cancer Res* **13**: 1523–32.
257. Liao L, Song M, Li X, Tang L, Zhang T, Zhang L, Pan Y, Chouchane L, Ma X (2017) E3 Ubiquitin Ligase UBR5 Drives the Growth and Metastasis of Triple-Negative Breast Cancer. *Cancer Res* **77**(8): 2090-2101.

258. Matsuura K, Huang NJ, Cocce K, Zhang L, Kornbluth S (2017) Downregulation of the proapoptotic protein MOAP-1 by the UBR5 ubiquitin ligase and its role in ovarian cancer resistance to cisplatin. *Oncogene* **36**: 1698–706.
259. Zhang T, Cronshaw J, Kanu N, Snijders AP, Behrens A (2014) UBR5-mediated ubiquitination of ATMIN is required for ionizing radiation-induced ATM signaling and function. *Proc Natl Acad Sci USA* **111**: 12091–6.
260. Yang M, Jiang N, Cao QW, Ma MQ, Sun Q (2016) The E3 ligase UBR5 regulates gastric cancer cell growth by destabilizing the tumor suppressor GKN1. *Biochem Biophys Res Commun* **478**: 1624–9.
261. Wang J, Zhao X, Jin L, Wu G, Yang Y (2017) UBR5 contributes to colorectal cancer progression by destabilizing the tumor suppressor ECRG4. *Dig Dis Sci* **62**: 2781–9.
262. Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, Chu C, Koepp DM, Elledge SJ, Pagano M, Conaway RC, Conaway JW, Harper JW, Pavletich NP (2002) Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* **416**(6882): 703-9.
263. Bai C, Sen P, Hofmann K, Ma L, Goebel M, Harper JW, Elledge SJ (1996) SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. *Cell* **86**(2): 263-74.
264. Adhikary S, Chakravarti D, Terranova C, Sengupta I, Maitituoheti M, Dasgupta A, Srivastava DK, Ma J, Raman AT, Tarco E, Sahin AA, Bassett R, Yang F, Tapia C, Roy S, Rai K, Das C (2019) Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor. *Nat Commun* **10**(1): 1398.
265. Campos EI, Smits AH, Kang YH, Landry S, Escobar TM, Nayak S, Ueberheide BM, Durocher D, Vermeulen M, Hurwitz J, Reinberg D (2015) Analysis of the Histone H3.1 Interactome: A Suitable Chaperone for the Right Event. *Mol Cell* **60**(4): 697-709.
266. Hogan AK, Sathyan KM, Willis AB, Khurana S, Srivastava S, Zasadzińska E, Lee AS, Bailey AO, Gaynes MN, Huang J, Bodner J, Rosencrance CD, Wong KA, Morgan MA, Eagen KP, Shilatifard A, Foltz DR (2021) UBR7 acts as a histone chaperone for post-nucleosomal histone H3. *EMBO J* **40**(24): e108307.
267. Srivastava S, Sahu U, Zhou Y, Hogan AK, Sathyan KM, Bodner J, Huang J, Wong KA, Khalatyan N, Savas JN, Ntziachristos P, Ben-Sahra I, Foltz DR (2021)

- NOTCH1-driven UBR7 stimulates nucleotide biosynthesis to promote T cell acute lymphoblastic leukemia. *Sci Adv* **7**(5): eabc9781.
268. Isobe Y, Okumura M, McGregor LM, Brittain SM, Jones MD, Liang X, White R, Forrester W, McKenna JM, Tallarico JA, Schirle M, Maimone TJ, Nomura DK (2020) Manumycin polyketides act as molecular glues between UBR7 and P53. *Nat Chem Biol* **16**(11): 1189-1198.
 269. Wheaton K, Sarkari F, Johns BS, Davarinejad H, Egorova O, Kaustov L, Raught B, Saridakis V & Sheng Y (2017) UBE2E1/UBCH6 is a critical in vivo E2 for the PRC1-catalyzed ubiquitination of H2A at Lys-119. *J Biol Chem* **292**, 2893–2902.
 270. Onitilo AA, Engel JM, Greenlee RT, Mukesh BN (2009) Breast cancer subtypes based on ER/PR and HER2 expression: comparison of clinicopathologic features and survival. *Clin Med Res* **7**: 4–13.
 271. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, Lickley LA, Rawlinson E, Sun P, Narod SA (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* **13**: 4429–4434.
 272. Haffty BG, Yang Q, Reiss M, Kearney T, Higgins SA, Weidhaas J, Harris L, Hait W, Toppmeyer D (2006) Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J Clin Oncol* **24**: 5652–5657.
 273. Vagia E, Mahalingam D, Cristofanilli M (2020) The Landscape of Targeted Therapies in TNBC. *Cancers (Basel)* **12**(4): 916.
 274. Dyer PN, Edayathumangalam RS, White CL, Bao Y, Chakravarthy S, Muthurajan UM, Luger K (2004) Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol* **375**: 23-44.
 275. Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F & Debyser Z (2005) The interaction of LEDGF/p75 with integrase is lentivirus specific and promotes DNA binding. *J Biol Chem* **280**: 17841–17847.
 276. Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G, Kim BJ, Li C, Chen R, Li W, Wang Y, O'Malley BW, Qin J (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145**: 787–799.
 277. Wy J (2006) Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* **40**: 339–343.

278. Adhikary S, Sanyal S, Basu M, Sengupta I, Sen S, Srivastava DK, Roy S & Das C (2016) Selective recognition of H3.1K36 dimethylation/H4K16 acetylation facilitates the regulation of all-trans-retinoic acid (ATRA)-responsive genes by putative chromatin reader ZMYND8. *J Biol Chem* **291**: 2664–2681.
279. Chakrabarti S, Bhattacharyya D & Dasgupta D (2000) Structural basis of DNA recognition by anticancer antibiotics, chromomycin A3, and mithramycin: Roles of minor groove width and ligand flexibility. *Biopolymers* **56**: 85–95.
280. Mondal P, Sen S, Klein BJ, Tiwary N, Gadad SS, Kutateladze TG, Roy S & Das C (2020) TCF19 promotes cell proliferation through binding to the histone H3K4me3 mark. *Biochemistry* **59**: 389–399.
281. Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**: 845–858.
282. Ray A, Lindahl E & Wallner B (2012) Improved model quality assessment using ProQ2. *BMC Bioinform* **13**: 1–12.
283. Anandapadamanaban M, Kyriakidis NC, Csizmók V, Wallenhammar A, Espinosa AC, Ahlner A, Round AR, Trewhella J, Moche M, Wahren-Herlenius M, Sunnerhagen M (2019) E3 ubiquitin-protein ligase TRIM21-mediated lysine capture by UBE2E1 reveals substrate-targeting mode of a ubiquitin-conjugating E2. *J Biol Chem* **294**(30): 11404-11419.
284. Sanchez R & Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**: 364–372.
285. Metzger MB, Pruneda JN, Klevit RE, Weissman AM (2014) RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta* **1843**(1): 47-60.

List of Publications

1. **Dasgupta A**, Mondal P, Dalui S, Das C and Roy S (2022) Molecular characterization of substrate-induced ubiquitin transfer by UBR7-PHD finger, a newly identified histone H2BK120 ubiquitin ligase' *FEBS J* **289**(7): 1842-1857. * **Editor's choice article**
2. Adhikary S, Chakravarti D, Terranova C, Sengupta I, Maitituoheti M, **Dasgupta A**, Srivastava DK, Ma J, Raman AT, Tarco E, Sahin AA, Bassett R, Yang F, Tapia C, Roy S, Rai K, Das C (2019) Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor. *Nat Commun* **10**(1):1398.

List of Conferences Attended

1. Poster presentation at Dutch Chromatin Meeting, Nijmegen, Virtual Meeting, November 2021.
2. Poster presentation at CSHL ‘Ubiquitin, Autophagy, & Diseases’ Virtual Meeting, April 2021.
3. Oral presentation at Annual Meeting of Society of Biological Chemists Kolkata Chapter, Shankarpur, India, March 2021.
4. Poster presentation at Saha Institute of Nuclear Physics International Cancer Meeting, Kolkata, September 2018.

Reprints



Dutch Chromatin Meeting 2021, Nijmegen

Certificate

Anirban Dasgupta
participated in the Dutch Chromatin Meeting
on the 12th of November 2021
and presented a poster.

On behalf of the organisers

Nelleke Spruijt



Radboud Universiteit Nijmegen





CERTIFICATE OF ATTENDANCE

Anirban Dasgupta

Was a participant at the Cold Spring Harbor Laboratory Conference

Ubiquitin, Autophagy & Disease

April 27 – 30, 2021 (Virtual)

They presented their abstract “Substrate recognition activates dimeric UBR7-PHD finger for unique mode of ubiquitin transfer” as a virtual poster at the above meeting.

Catie Carr, Conference Coordinator



SOCIETY OF BIOLOGICAL CHEMISTS (I), KOLKATA CHAPTER

Organizes Symposium on

"Current Trends in Biology for Human Diseases and Medicine"

From 19-21st March 2021

&

Present this Certificate to

ANIRBAN DASGUPTA

**in appreciation of Oral presentation/Co-Chair/Participation
in the conference at Hotel "NEST" Shankarpur, Digha**

Shoyeb

President

Tanya Das.

Secretary



Certificate of Participation

This is to certify that

Anirban Dasgupta

has successfully participated in

SINP International Cancer Meeting 2018

Cancer Biology - Still A Challenge in 21st Century

**&
SINP School on Epigenetics**

September 26-28, 2018

Saha Institute of Nuclear Physics



Kolkata, India

Subrata Banerjee
Subrata Banerjee
Chairman

Debashis Mukhopadhyay
Debashis Mukhopadhyay
Joint Convener

Chandrima Das
Chandrima Das
Joint Secretary

Molecular characterization of substrate-induced ubiquitin transfer by UBR7-PHD finger, a newly identified histone H2BK120 ubiquitin ligase

Anirban Dasgupta¹, Payel Mondal^{2,3}, Sambit Dalui¹, Chandrima Das^{2,3}  and Siddhartha Roy¹ 

¹ Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research (CSIR) – Indian Institute of Chemical Biology, Kolkata, India

² Biophysics and Structural Genomics Division, Saha Institute of Nuclear Physics, Kolkata, India

³ Homi Bhabha National Institute, Mumbai, India

Keywords

chromatin; E3 ligase; monoubiquitination; oligomerization; PHD finger

Correspondence

S. Roy, Structural Biology and Bioinformatics Division, Council of Scientific and Industrial Research (CSIR) – Indian Institute of Chemical Biology, 4, Raja S.C. Mullick Road, Kolkata – 700032, India
 Tel: +91 3324995733
 E-mail: roysiddhartha@iicb.res.in

(Received 5 March 2021, revised 21 October 2021, accepted 3 November 2021)

doi:10.1111/febs.16262

Monoubiquitination of histone H2B at lysine 120 plays a vital role in active transcription and DNA damage response pathways. Ubiquitin protein ligase E3 component N-recognin 7 (UBR7) has been recently identified as an H2BK120 monoubiquitin ligase. However, the molecular details of its ubiquitin transfer mechanism are not well understood. Here, we report that the plant homeodomain (PHD) finger of UBR7 is essential for its association with E2 UbcH6 and consequent ubiquitin transfer to its substrate histone H2B. We also identified the critical region of UbcH6 involved in this function and shown that the residues stretching from 114 to 125 of histone H2B C-terminal tail are sufficient for UBR7/UbcH6-mediated ubiquitin transfer. We also employed antibody-independent mass spectrometry to confirm UBR7-mediated ubiquitination of the H2B C-terminal tail. We demonstrated that the PHD finger of UBR7 forms a dimer and this dimerization is essential for ubiquitination of histone H2B. We mapped the critical residues involved in the dimerization and mutation of these residues that abrogate E3 ligase activity and are associated with cancer. Furthermore, we compared the mode of ubiquitin discharge from UbcH6 mediated by UBR7 and RING finger protein 20 (RNF20) through a thioester hydrolysis assay. Interestingly, binding of substrate H2B to UBR7 induces a conformational change in the PHD finger, which triggers ubiquitin transfer from UbcH6. However, the RNF20 RING finger alone is sufficient to promote the release of ubiquitin from UbcH6. Overall, the mechanism of ubiquitin transfer by the newly identified E3 ubiquitin ligase UBR7 is markedly different from that of RNF20.

Introduction

The complex and elaborately packed state of DNA and histone proteins, folded into a 30-nm fibre within the eukaryotic nucleus, is referred to as chromatin. Structural studies of the chromatin fibre have revealed

multiple tandem arrays of constituent units called nucleosomes [1], which are composed of 146-bp long DNA wrapped 1.65 times around two copies of H2A–H2B dimer and a single copy of H3–H4 tetramer

Abbreviations

BARD1, BRCA1-associated RING domain protein; BPTF, bromodomain PHD finger transcription factor; BRCA1, breast cancer susceptibility gene 1; DDM, dimer-deficient triple mutant; DFDNB, 1,5-difluoro-2,4-dinitrobenzene; FAM, fluorescein amidites; PDB, Protein Data Bank; PHD, plant homeodomain; RNF20, RING finger protein 20; RNF40, RING finger protein 40; SEC-MALS, size exclusion chromatography coupled with multiple-angle light scattering; UBR7, ubiquitin protein ligase E3 component N-recognin 7; WT, wild-type.

assembled together into the core octamer through interactions of their highly conserved histone fold motifs [2,3]. Core histone tails, especially of histones H2A and H2B, are involved in folding of the chromatin fibre by securing the electrostatic charge of DNA internally and bridging successive nucleosomes [3–5]. Histones have been reported as being enzymatically modified [6,7] to alter chromatin structure and recruit histone modifiers such that replication, transcriptional regulation and several other processes are facilitated [8,9]. Among the different post-translational modifications, histone ubiquitination is one of the less understood modifications implicated in transcription [10–12] and the DNA damage response [13].

Histones H2A and H2B are predominantly monoubiquitinated, which causes an alteration in nucleosomal dynamics and initiates cross-talk with other histone modifications such as H3K4 and H3K79 trimethylation [14]. H2A lysine 119 (K119) has been previously described as a major ubiquitination site [15,16], acted upon by E3 ligases Ring1b-Bmi1 (part of polycomb repressive complex 1) [17,18], breast cancer susceptibility gene 1 (BRCA1)/BRCA1-associated RING domain protein (BARD) [19] and 2A-HUB [20] in different contexts, and it is known to bring about chromatin compaction and gene repression [10,20]. Monoubiquitination of histone H2B at lysine 123 (K123) in yeast is mediated by Rad6-Bre1 [21,22] and lysine 120 (K120) in human is mediated by RING finger protein 20 (RNF20)/finger protein 40 (RNF40) complex [23]. By contrast to H2A, monoubiquitination of H2B is associated with transcribed regions of highly expressed genes and correlated with transcriptional memory and elongation [11,24]. The H2B C-terminal tail, when ubiquitinated, significantly enhances the surface area of the nucleosome [25], thereby enabling H2BK120Ub marker to establish cross-talk with other modifications within the nucleosome core, consequently stimulating downstream processes [26]. Disruptions of the machinery regulating H2BK120Ub and its downstream signaling have been implicated in oncogenesis, with a reduction of the global level of H2BK120Ub reported in a variety of cancers [27–29]. Other E3 ligases, such as RING finger protein 8, BRG1- or BRM-associated factor 250b, mouse double minute 2 and BRCA1-BARD1, have also been associated with H2B K120 monoubiquitination [28–30], although the RNF20/40 complex is by far the most well-studied ubiquitination enzyme for H2BK120Ub [23,31]. Among the other closely related ubiquitin-conjugating enzymes (E2), UbcH6 (also called E2E1) has been reported as the common E2 enzyme in both H2A and H2B monoubiquitination [23,32].

The RING finger of Bre1 in yeast, responsible for its E3 ligase activity, displays an overall evolutionary conservation for its mouse homologue RNF20. However, there are specific divergences reported in the form of coiled-coil motifs, which are longer and more complex in mouse and *Drosophila* than their yeast counterpart [33]. Apart from the canonical interaction between E2-Rad6 and E3-Bre1 (via RING domain), a non-RING Rad6-binding domain of Bre1 has also been reported to interact with the opposite region of the Rad6 active site, ensuring monoubiquitination and preventing polyubiquitination of H2BK123 [34]. RNF20 forms a heterodimeric complex in humans, with related RING finger protein 40 (RNF40) and loss of either one results in the global reduction of the H2BK120Ub level. RING fingers of both RNF20 and RNF40 are active E3 ligases and exist individually as homodimers in solution *in vitro*, although the reduced stability of RNF20 RING dimer bound to Ube2B-Ub conjugate suggests the importance of sequences outside the RING domain that play a role in stabilizing the complex inside the cell [35].

Ubiquitin protein ligase E3 component N-recognin 7 (UBR7) has been recently found to monoubiquitinate H2B at K120 position and function as a breast tumor suppressor, antagonizing epithelial–mesenchymal transition through activation of cadherin 4 [36]. UBR7 belongs to the UBR protein family of mammalian E3 ligases, comprising seven members (UBR1–7), which share a common 70-residue zinc finger UBR box that, according to previous studies, recognizes and binds to the N-terminal residues of short-lived proteins, thereby leading to destabilization (N-degrons), and is involved in the N-end rule pathway [37]. However, although UBR3, UBR6 and UBR7 harbour an UBR box, they do not associate with N-degrons. The UBR family members exhibits neither size, nor sequence similarity, but present signature domains associated specifically with E3 ligase activity via RING in UBR1/UBR2/UBR3, HECT in UBR5 or F-box in UBR6 [38]. UBR7 harbours a plant homeodomain (PHD), which is absent in any other proteins of the UBR family and is a known chromatin binding module commonly reported to be a reader of unmodified, acetylated and methylated histones [39,40]. Although it is known that UBR7 utilizes UbcH6 as its cognate E2 partner for mediating ubiquitin transfer via its PHD finger motif [36], the detailed catalytic mechanism is not yet well understood. In the present study, we show that the UBR7-PHD finger in partnership with E2 UbcH6 is involved in a novel mode of ubiquitin transfer to substrate histone H2B. We characterized the interaction of the UBR7-PHD finger with E2 UbcH6 and

substrate H2B and investigated the effect of dimerization on UBR7-PHD E3 ligase activity. By comparative biochemical assays, we probed the molecular mechanism of UBR7-PHD finger-mediated ubiquitin transfer and how it differs from the mechanism of the already well-studied RNF20 RING finger. Overall, the present study aimed to understand the molecular details of the UBR7/UbcH6 complex with respect to transferring ubiquitin to substrate H2B through a distinct mechanism compared to RNF20/UbcH6 complex.

Results

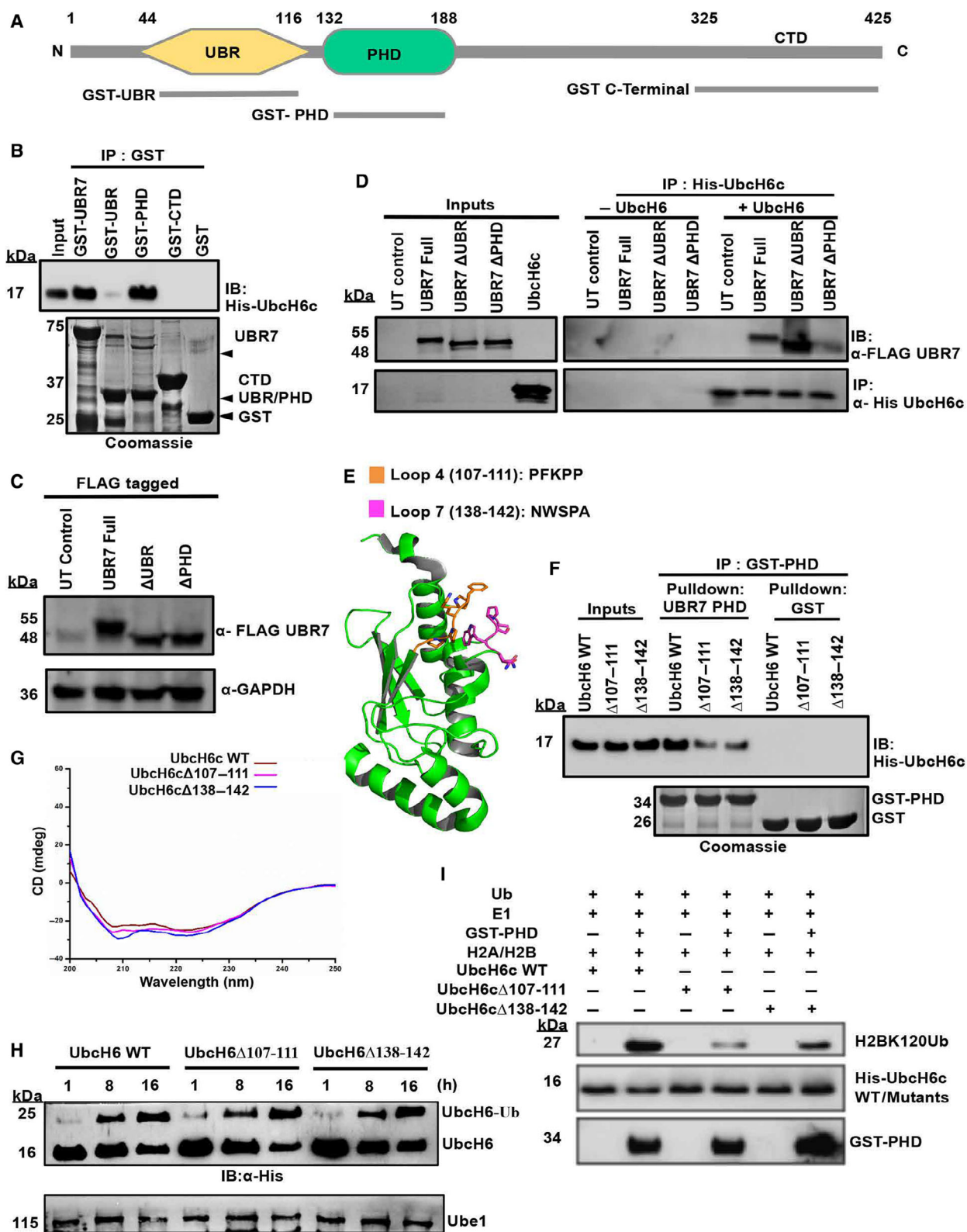
Mapping the interaction between UBR7 and its E2 interacting partner, UbcH6

Among the different members of the UBR family in humans sharing the conserved UBR box, UBR7 alone contains a PHD finger that is absent from the rest of the family members [37]. We have recently identified its role as a H2BK120 monoubiquitinating enzyme with an important role as a suppressor of breast cancer metastasis [36]. UBR7 contains two functional domains, UBR box and PHD finger, present at the N-terminal part of the protein, whereas the C-terminal domain is unstructured (Fig. 1A). We first mapped the domains of UBR7 involved in interaction with its E2 conjugate partner, UbcH6. GST-tagged UBR7-full length, UBR box, PHD finger and C-terminal domain were purified and assessed for interaction with His-UbcH6 core domain (UbcH6c). The interaction was found to be robust with wild-type (WT) and the PHD finger of UBR7, but not with the UBR box or the C-terminal region of UBR7 (Fig. 1B). We subsequently expressed FLAG-conjugated constructs of full-length or truncated UBR7, with either UBR box (44–116) or PHD finger (132–188) deletion in HeLa cells and found that their expression levels were comparable (Fig. 1C). Upon incubation with bacterially expressed His-UbcH6c and subsequent pull-down with nickel-nitrilotriacetic acid beads, we observed that full-length UBR7 showed a robust interaction. However, deletion

of PHD finger showed a significant reduction in its interaction with UbcH6c, whereas the deletion of UBR box showed no such change (Fig. 1D). This again confirmed the UBR7 interacts with UbcH6 through PHD finger.

Because the X-ray crystal structure of UbcH6 core domain encompassing residues 37–193 [Protein Data Bank (PDB) ID: 5LBN] has been previously reported, we attempted to identify the critical residues of UbcH6 involved in interaction with UBR7. Interestingly, by mapping the surface charge of residues around the catalytic cysteine C131, we observed that loop 4 (residues 107–111) and loop 7 (residues 138–142) of UbcH6 are possibly involved in binding with the UBR7-PHD finger (Fig. 1E). We further generated His-tagged constructs of loop 4 and loop 7 deleted UbcH6c mutants and compared their ability to interact with GST-UBR7-PHD finger as opposed to His-UbcH6c WT protein. Both the loop deleted mutants of UbcH6 showed reduced ability to interact with UBR7-PHD finger (Fig. 1F). Thus, our results indicated that PHD finger of UBR7 critically interacted with loop 4 and loop 7 of UbcH6. We next aimed to investigate whether loop deletion have any effect on UbcH6 conformation. Accordingly, we performed CD spectroscopy with WT and loop deletion E2 mutants (Fig. 1G). We did not observe any significant structural change upon the loop deletion. We also performed a time-dependent E2-Ub charging experiment with the WT and loop deletion E2 mutants using E1 enzyme. We did not observe any alteration in the UbcH6-Ub level with the WT and loop deletion E2 mutants at any particular time point (Fig. 1H). These results suggest that loop deletion of UbcH6 neither changes the conformation of Ubc fold, nor reduces E1–E2 interaction with the loop deleted mutants that can impair E2-Ub charging. We subsequently performed an *in vitro* ubiquitination assay to further confirm the role of loop 4 and loop 7 of UbcH6 in UBR7-mediated histone H2BK120 ubiquitination. We observed that, compared to WT UbcH6c, loop 4 and loop 7 deleted UbcH6c shows decreased H2BK120

Fig. 1. Mapping the interacting domains of UBR7 and UbcH6. (A) Domain architecture of human UBR7 highlighting the N-terminal UBR box (yellow) and PHD finger (green). The grey bars below represent the GST-tagged constructs used in the study. (B) *In vitro* GST pull-down assay of UBR7 and its individual domains (UBR, PHD, C-terminal domain) with UbcH6 core domain (37–193), blots probed with α -His and α -GST antibodies. (C) Expression of FLAG-tagged UBR7-full length and its domain deletion constructs in HeLa cells. (D) Nickel-nitrilotriacetic acid pull-down assay of His-tagged UbcH6c with overexpressed UBR7 and its deletion constructs, blots probed with α -FLAG and α -His antibodies. (E) Crystal structure of UbcH6 (PDB ID: 5LBN) highlighting loop 4 (orange) and loop 7 (magenta), adjacent to catalytic cysteine residue C131. Generated using PYMOL, version 2.1.1. (F) GST pull-down assay of UBR7-PHD with His-tagged UbcH6c WT and its loop deletion constructs. (G) Far-UV CD spectra of UbcH6c WT and its loop deletion constructs. (H) Time-course experiment of ubiquitin charging with UbcH6c WT and loop deletion mutants. UbcH6c-Ub species, UbcH6c and Ube1 proteins were detected by α -His antibody. (I) *In vitro* ubiquitination assay of UBR7 with UbcH6 WT and loop deletion constructs. H2BK120Ub was detected by α -H2BK120Ub antibody.



ubiquitination levels (Fig. 1I). These results indicate that PHD finger of UBR7 is involved in interaction with loop 4 and loop 7 of UbcH6 and this interaction is crucial for UBR7-mediated histone H2BK120 monoubiquitination.

H2B C-terminal peptide is sufficient for UBR7-mediated ubiquitination

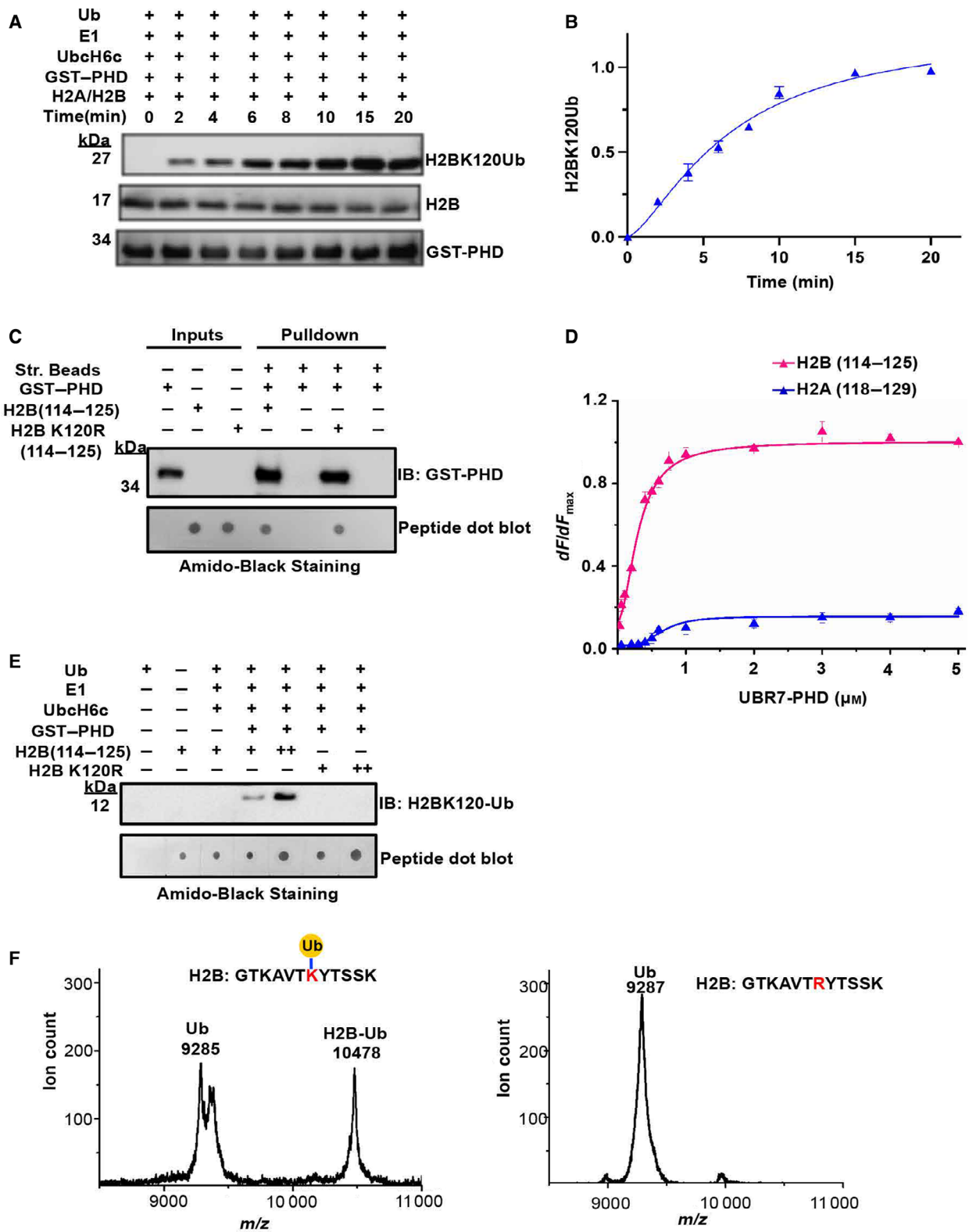
To understand the role of UBR7-PHD finger in catalyzing histone H2BK120Ub, we performed a time course experiment where we measured formation of the product H2BK120Ub from the substrate histone H2B over 0–20 min. During the assay, the enzyme and substrate concentrations were kept constant. We observed a time-dependent increase in H2BK120 ubiquitination up to 15 min (Fig. 2A). We performed the single-turnover substrate ubiquitination assay three times. The average of the values for H2BK120Ub formation of each time point were plotted in a product versus time graph using the Hill equation (Fig. 2B). We subsequently aimed to identify the minimum stretch of amino acid sequence of histone H2B substrate that was required for ubiquitin transfer. Because UBR7 can selectively ubiquitinate H2B either free or in different levels of association into H2A/H2B dimer, core histones and nucleosomes [36], we next examined whether full-length H2B containing histone fold domain or the C-terminal tail containing lysine 120 is sufficient for H2BK120 monoubiquitination. To address this, we generated C-terminal H2B peptide spanning amino acid residues 114–125 and H2B K120R mutant peptide of the same stretch, and first performed an *in vitro* interaction assay. We observed that both peptides showed significant interaction with UBR7-PHD finger (Fig. 2C). We employed fluorescence spectroscopy to characterize the association of UBR7-PHD finger with fluorescein amidites (FAM)-conjugated H2B peptide and found the binding affinity (K_d) to be around $0.3 \pm 0.02 \mu\text{M}$. No such association with UBR7-PHD was observed in the case of FAM-

conjugated H2A peptide (118–129) (Fig. 2D). Subsequently, an *in vitro* ubiquitination assay was performed to identify the ability of UBR7-PHD finger to ubiquitinate H2B WT or H2B K120R peptides. We observed that only the H2B WT peptide was ubiquitinated as detected by anti-H2BK120Ub antibody. The corresponding K120R mutant peptide showed no ubiquitination band even though it interacts with the UBR7-PHD finger (Fig. 2E). Interestingly, the C-terminal tail of histone H2B contains three lysine residues at positions 116, 120 and 125, which have potential to become ubiquitinated. To confirm the exact site of ubiquitination, we used mass spectrometry in which we obtained a distinct peak corresponding to 10.478 kDa for the combined mass of H2B WT peptide (~1.2 kDa) and monoubiquitin (~9.2 kDa), whereas no such peak was seen in the case of H2B K120R mutant peptide (Fig. 2F). The control run without UBR7-PHD enzyme also failed to detect any ubiquitinated H2B peptide (data not shown). Our results indicate that the C-terminal tail of histone H2B is necessary and sufficient for UBR7-mediated ubiquitination.

Dimerization of UBR7 critically regulates its E3 ubiquitin ligase activity

Several RING fingers of E3 ligases form homo- or heterodimers in solution. We intended to understand the oligomeric status of UBR7 and the role of the same with respect to its E3 ligase activity. To determine the oligomeric status of the protein *ex vivo*, HeLa cells were co-transfected with EGFP-UBR7 in the presence of either FLAG-UBR7 WT or FLAG-UBR7-PHD/UBR domain deleted constructs (FLAG-UBR7 Δ PHD or FLAG-UBR7 Δ UBR). Following EGFP immunopull-down, we probed for the association of FLAG-proteins with EGFP-UBR7. We observed that, unlike FLAG-UBR7 WT or FLAG-UBR7 Δ UBR, the association of FLAG-UBR7 Δ PHD with EGFP-UBR7 was the lowest (Fig. 3A). This indicates that the PHD finger of UBR7 is essential for dimerization. To fine-map

Fig. 2. UBR7/UbcH6 enzyme complex transfers ubiquitin to histone H2B C terminus. (A) UBR7-PHD mediated single-turnover substrate ubiquitination assay using E2 UbcH6c and substrate H2A/H2B dimer involving different reaction time points (0–30 min). Each experiment was repeated three times. (B) Quantification of substrate ubiquitination shown corresponding to each time point plotted in a product versus time graph. Each triangle represents average H2BK120Ub intensity ($n = 3$ repeats) at the time point of concern, and the data were fitted using the Hill equation. Error bars show the SD. (C) Streptavidin agarose pull-down assay using biotinylated H2B C-terminal peptide (114–125) and its K120R mutant with UBR7-PHD. Interaction was probed with α -GST antibody. (D) Binding isotherm for the interaction of FAM-conjugated H2B C-terminal peptide (GTKAVTKYTSSK) in pink and H2A C-terminal peptide (KKTESHHKAKGK) in blue with UBR7-PHD obtained via fluorescence spectroscopy ($n = 3$ repeats; error bars show the SD). (E) *In vitro* ubiquitination assay with C-terminal WT and K120R mutant peptides of histone H2B as substrates. (F) MALDI-TOF mass spectra of UBR7-PHD-UbcH6-mediated monoubiquitination of histone H2B C-terminal peptides. Ubiquitinated H2B peptide (114–125) corresponding to mass 10.478 kDa and free ubiquitin (Ub) of mass 9.285 kDa are labelled.



the residues of UBR7 involved in dimerization, molecular modelling of UBR7-PHD finger was performed by employing bromodomain PHD finger transcription factor (BPTF)-PHD finger (PDB ID: 2F6N) as a reference template on account of significant sequence similarity (43%) using the PHYRE2 algorithm. The dimeric structure shows a prominent salt bridge interaction between E145 and R165 and a hydrogen-bonding interaction between the main chain carbonyl and side chain hydroxyl group of E151 residues (Fig. 3B). These dimerization interface residues are conserved across different eukaryotes (Fig. 3C). Interestingly, from the cBioPortal database [41], we observed that the critical residues involved in dimerization are mutated to E151D and R165M in uterine endometrial carcinoma and papillary renal cell carcinoma, respectively (Fig. 3D). Subsequently, we mutated these interface residues to investigate their role in the dimerization of UBR7. Dimerization deficiency of the triple mutant E145K/E151A/R165A was validated with a 1,5-difluoro-2,4-dinitrobenzene (DFDNB)-mediated cross-linking assay (Fig. 4A), where we found that, compared to WT, the triple mutant showed a significant loss of dimerization upon cross-linking. To determine the oligomeric status of UBR7-PHD in solution, we used size exclusion chromatography coupled with multiple-angle light scattering (SEC-MALS) to measure the molecular mass of UBR7-PHD WT and dimer-deficient triple mutant (DDM) (Fig. 4B) under low salt buffer conditions. The molecular mass of UBR7-PHD WT was measured at 18.3 kDa (theoretical monomer mass was 8.4 kDa), which corresponded to a dimer. However, UBR7-PHD DDM was 6.6 kDa (theoretical monomer mass was 8.2 kDa) (Fig. 4C), which did not exhibit any dimeric association. The effect of dimerization deficiency on E3 ligase activity of UBR7-PHD was further tested with an *in vitro* ubiquitination assay. Compared to WT, the dimer-deficient single mutants (E145K, E151A, R165A) showed a moderate reduction in H2BK120Ub formation. However, on the other hand, the triple mutant (DDM) showed a significant reduction in its ability to ubiquitinate histone H2BK120, indicating that the PHD finger mediated dimerization of UBR7 is essential for its E3 ligase activity (Fig. 4D). We also quantified the substrate ubiquitination intensity of the mutants with respect to WT using IMAGEJ (<https://imagej.nih.gov/ij/>) (Fig. 4E). To gain insight into the role of dimer-deficient mutant of UBR7 (DDM) on its E2 binding, we performed a nickel-nitrilotriacetic acid pull-down assay with His-tagged UbcH6c and untagged PHD WT or DDM. We observed that UbcH6c binding to DDM is significantly reduced compared to the PHD finger, indicating that

dimerization of UBR7 plays essential role in UbcH6c binding (Fig. 4F).

Distinct catalytic mechanism of UBR7-PHD finger as compared to RNF20 RING finger towards H2BK120 ubiquitination

The H2BK120 site is ubiquitinated by RNF20 as well as UBR7. To determine how ubiquitin is transferred from E2 UbcH6 to its substrate histone H2B, by the E3 ligase UBR7 or RNF20, we performed a thioester hydrolysis assay, where we incubated E2-ubiquitin conjugate with either GST-PHD finger of UBR7 or RING finger of RNF20 with an increasing time point and measured the release of free ubiquitin by western blotting. Interestingly, we observed that the PHD finger of UBR7 could not hydrolyze thioester bond of E2-ubiquitin, whereas the RING finger RNF20 alone could do so (Fig. 5A,B). Thereafter, we aimed to determine whether ubiquitin release requires the association of the substrate with UBR7. To address this, we took WT H2B C-terminal peptide (114–125) and performed the thioester hydrolysis assay and we observed an efficient transfer of ubiquitin to H2B peptide by UBR7 (Fig. 5C). This indicates that the substrate binding to UBR7 is crucial for the ubiquitin transfer mechanism to operate. Next, we considered whether the lysine 120 residue of histone H2B is essential for ubiquitin release or whether the substrate binding is sufficient. To address this, we used the H2B K120R peptide and observed that ubiquitin release could still be detected in a time-dependent manner, indicating that the substrate binding is sufficient for the E3 ligase activity of UBR7 (Fig. 5D). To examine the role of substrate binding in triggering the activation of UBR7-PHD, we employed CD spectroscopy to identify any structural changes in UBR7-PHD that might be induced because of an association with substrate H2B. Upon analysis of far UV CD spectra, we found a significant conformational change occurring in the case of UBR7-PHD after substrate H2B binding (Fig. 5E) but no such change was noted in the case of RNF20 RING (Fig. 5F). These results clearly indicate that the ubiquitin transfer mechanism of the E3 ubiquitin ligases, UBR7 and RNF20, to histone H2B substrate from E2 conjugating enzyme UbcH6, is distinctly different.

Discussion

PHD finger proteins are versatile chromatin readers involved in the recognition of H3K4Me3, H3K4Me0 or acetylated H3/H4 [39,40]. However, to date, no

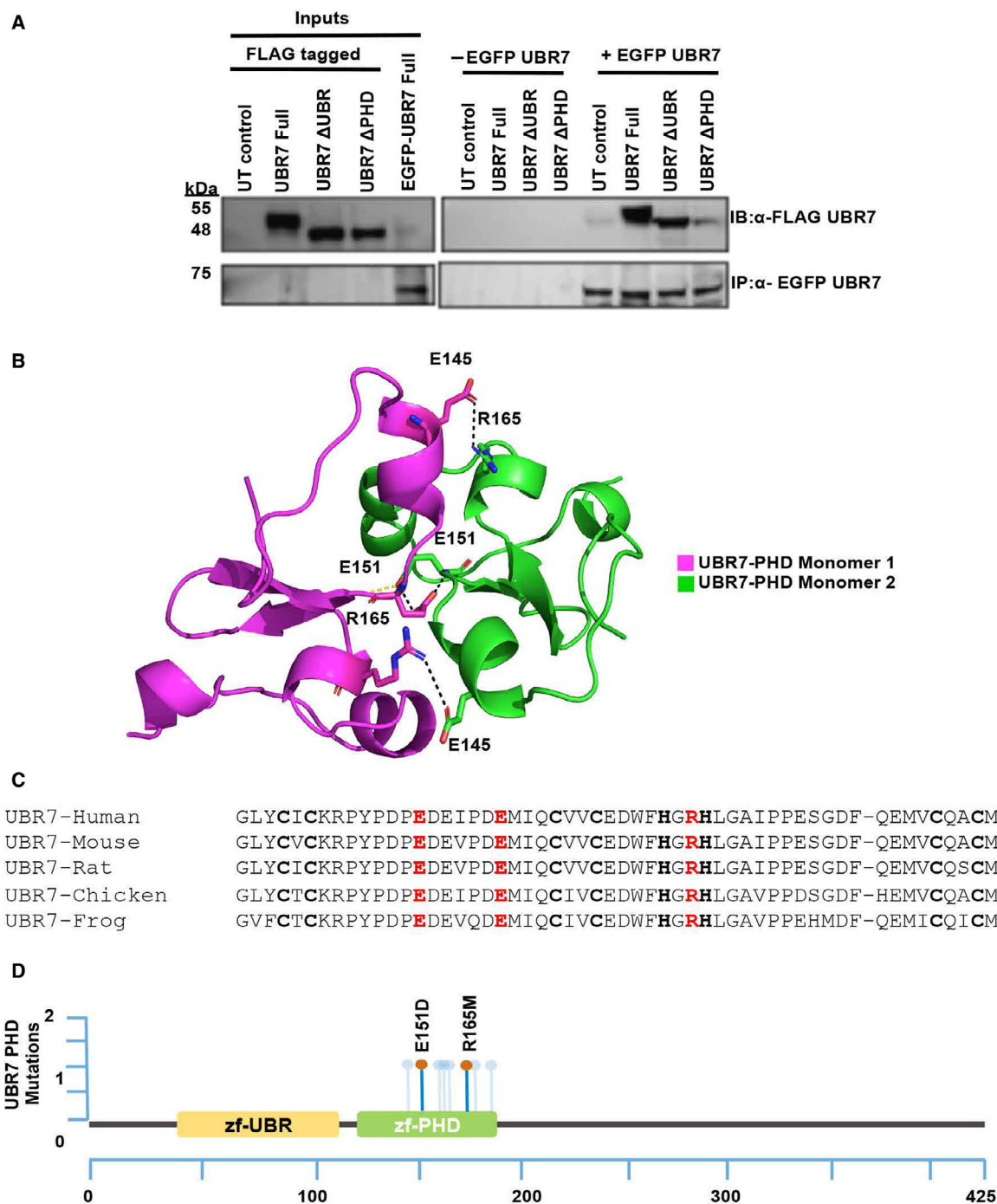


Fig. 3. UBR7-PHD finger exists as a dimer. (A) Transient transfection was performed with EGFP-UBR7 and FLAG-UBR7 or FLAG-conjugated deletion constructs in HeLa cells. Subsequently, immunoprecipitation was performed employing α -EGFP antibody (IP) and probed with α -FLAG antibody (IB). (B) Molecular modelling of UBR7-PHD using the crystal structure of human BPTF (PDB ID: 2F6N) as template. Potential residues involved in dimerization – E145, E151 and R165 are highlighted as sticks. Generated using PYMOL, version 2.1.1. (C) Multiple sequence alignment of UBR7-PHD across different species with highly conserved residues E145, E151 and R165 involved in dimerization are highlighted in red. The zinc coordinating residues within the PHD zinc finger is indicated in bold. Sequences were obtained from the UniProt database (<https://www.uniprot.org/>) and aligned using CLUSTAL OMEGA (<https://www.ebi.ac.uk/Tools/msa/clustalo>). UniProt accession numbers: UBR7-Human (Q8N806), UBR7-Mouse (Q8BU04), UBR7-Rat (Q642A8), UBR7-Chicken (Q5ZMN4) and UBR7-Frog (Q569T8). (D) Schematic representation of UBR7 missense somatic mutations within the PHD finger in different cancers identified from TCGA database.

catalytic function of this domain has been established. UBR family members have been reported to have E3 ubiquitin ligase activity through canonical RING, HECT or F-Box [38]. It was recently reported that UBR7, a histone-interacting protein, is involved in the monoubiquitination of H2BK120 [36]. However, no molecular details pertaining to its ability to enzymatically catalyse the ubiquitination of H2B was available. The present study aimed to understand the mechanism of ubiquitin transfer mediated by UBR7 from its E2 interacting partner UbcH6 to histone H2B. For the first time, our results establish the role of the PHD finger of UBR7 in the monoubiquitination of histone H2BK120.

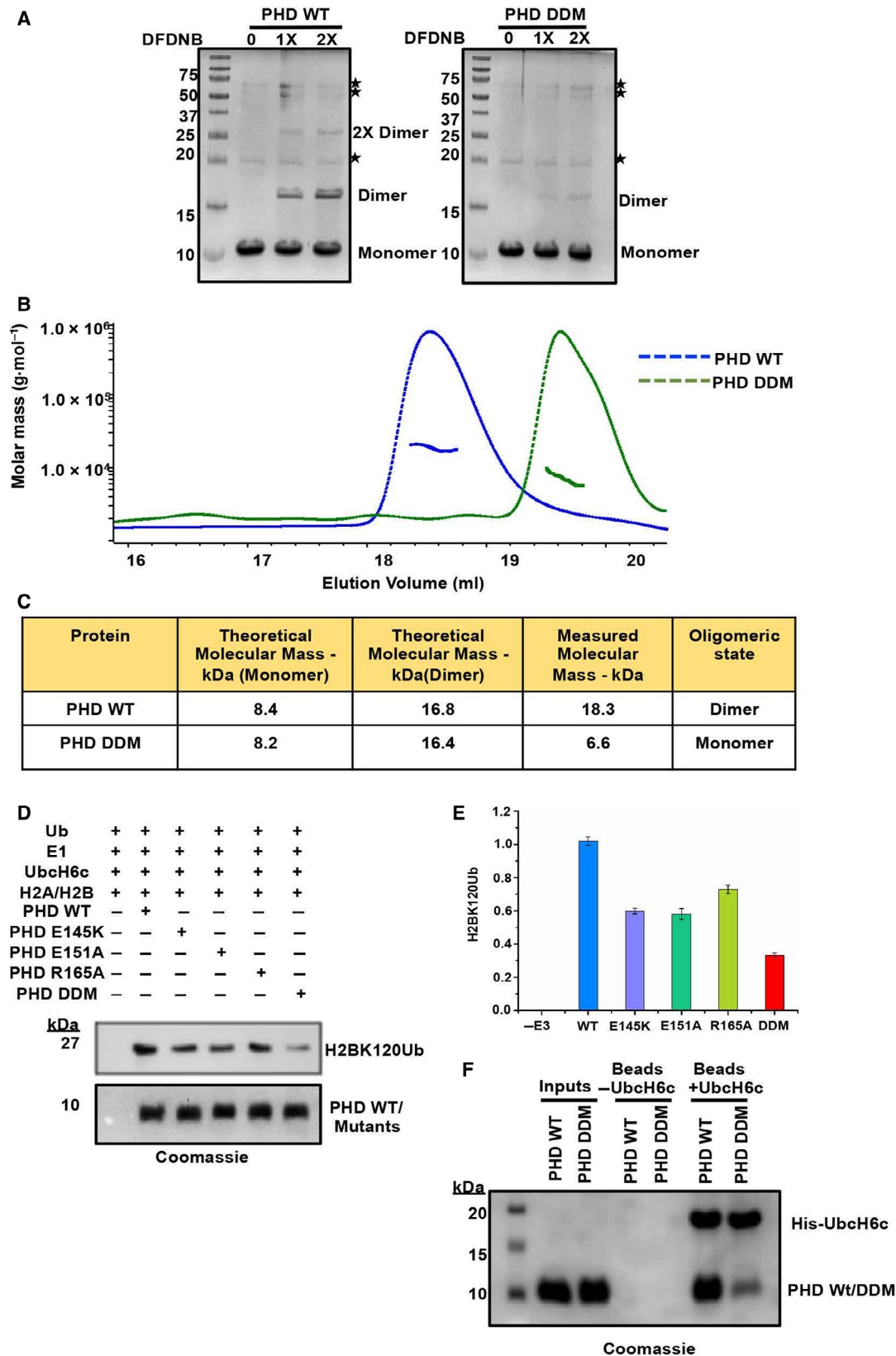
To decipher the molecular mechanism, we first mapped the interacting regions of UBR7 with its E2, UbcH6 and substrate histone H2B. It was observed that the PHD finger of UBR7 specifically interacts with UbcH6 as well as H2B. Next, we focused on identifying the critical amino acid stretch of UbcH6 that is involved in an association with UBR7 and found that amino acids encompassing loop 4 and loop 7 of UbcH6 play a critical role. The other interacting partner of UBR7 was the substrate histone H2B. We established that the C-terminal tail of histone H2B was critical in mediating this interaction. Following the interaction map, we established that, for H2BK120 ubiquitination, the C-terminal histone peptide harbouring K120 residue is sufficient for ubiquitination. Site-specific mutation of H2BK120R abolished its ability to become ubiquitinated. Similarly, loop 4 and loop 7 residues of UbcH6 played a critical role in ubiquitination because deleting these stretches compromised the ubiquitin transfer to histone H2B mediated by UBR7.

We also observed that the UBR7-PHD finger has a dimeric association *in vitro*. We subsequently validated the same finding in cells. The critical residues located in the dimerization interface were found to be highly conserved. This indicated a potential role of these residues in UBR7 function. We assessed the contribution of these residues in the E3 ubiquitin ligase activity of

UBR7. We have observed that mutation of these residues compromised the catalytic function of UBR7 significantly. Because the role of the oligomeric status of this protein has not been previously explored, the present study prominently highlights that UBR7 acts as a dimer for eliciting the enzymatic catalytic activity. The importance of these residues is also reinforced from The Cancer Genome Atlas (TCGA) database (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>), where mutations to these dimerization interface residues were reported in various cancers. Thus, the dimerization-defective mutant of UBR7 with compromised H2BK120 ubiquitinating function has serious consequences in a physiological context, which can be correlated with the propagation of tumorigenicity. However, elucidation of the role of these key residues involved in dimerization needs to be established in a cellular context.

We subsequently compared the mechanism of ubiquitin transfer between UBR7 and a canonical RING Finger H2BK120 ubiquitin ligase, RNF20. The requirement of substrate for the hydrolysis of UbcH6-Ub thioester bond was absolutely essential for UBR7 and dispensable for RNF20. This indicates that UbcH6-mediated ubiquitination machinery makes clear the choice of which E3 ligase, UBR7 or RNF20, to employ for the monoubiquitination of H2BK120 on the basis of the abundance of substrate histone H2B. The substrate binding to PHD finger is sufficient because we observed that, in the presence of H2B C-terminal peptide, be it WT or mutant (K120R), the catalytic hydrolysis of UbcH6-Ub thioester is active. We then attempted to determine whether substrate binding to UBR7 contributed to the ubiquitin discharge process. We found that, unlike RNF20, UBR7 requires histone H2B binding to initiate the ubiquitin discharge from UbcH6, which maybe brought about by a conformational change. Ubiquitination mediated activation of signalling pathways has a direct implication in regulating the cellular transcription program as well as DNA repair. The molecular details that evolved from the present study have clearly established

Fig. 4. Dimerization of UBR7-PHD is essential for its E3 ligase activity. (A) DFDNB-mediated cross-linking assay with untagged UBR7-PHD WT and PHD dimer-deficient triple mutant (DDM) E145K/E151A/R165A. Non-specific contamination bands are marked by a star (★). (B) SEC-MALS analysis of untagged UBR7-PHD WT and dimer-deficient triple mutant (DDM) denoted by a dashed blue line and dashed green line, respectively. Dashed lines denote the refractive index trace for proteins eluted and the horizontal line under the peak corresponds to the average molar mass (y-axis) distribution across the peak as determined by MALS. (C) Table of molar mass measurements by SEC-MALS. (D) *In vitro* ubiquitination assay with untagged UBR7-PHD WT, single mutants (E145K, E151A, R165A) and triple mutant DDM. Western blotting was performed with α -H2BK120Ub antibody. PHD WT and its mutant proteins were shown by Coomassie Brilliant blue staining. (E) Quantification of H2BK120Ub level *in* ubiquitination assay using PHD WT and mutants using IMAGEJ with $n = 3$ repeats; error bars show the SD. (F) Interaction of His-UbcH6c with WT PHD or DDM mutant was analyzed by a nickel-nitrilotriacetic acid pull-down assay run on SDS/PAGE.



a mechanism distinct from the canonical RNF20-mediated ubiquitin transfer to histone H2BK120.

Materials and methods

Constructs and mutagenesis

Expression vectors containing ubiquitin (#10861) and Ube1 (#34965) were purchased from Addgene (Watertown, MA, USA). UBR7-full length (1–425), Δ UBR (44–116) or Δ PHD (132–188) was cloned into mammalian vector pcDNA3.1+/C-(K)DYK. UbcH6 core domain (37–193) and UBR7-PHD (132–203) were subcloned into pET28A (Novagen, Madison, WI, USA) and pGEX6P-1 (GE Healthcare, Chicago, IL, USA) vectors, respectively, using restriction enzymes *NdeI/XhoI* and *BamHI/XhoI* (New England Biolabs, Ipswich, MA, USA). The mutant and deletion clones were generated using a QuikChange site-directed mutagenesis kit (Stratagene, San Diego, CA, USA) following standard protocols [42]. All subclones and mutants were further confirmed by sequencing.

Recombinant protein expression and purification

Clones were transformed into *Escherichia coli* BL21 (DE3), overnight bacterial colonies were grown in LB media until *D* of 0.6 was achieved, followed by induction with 1 mM IPTG for 18 h at 18 °C. In the case of zinc finger proteins, IPTG was supplemented with 50 μ M ZnCl₂. For purification of GST-tagged proteins, cells were resuspended in buffer containing 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 5% glycerol, 2 mM DTT and 0.05% NP-40 and sonicated followed by centrifugation at 38 724 *g* for 45 min at 4 °C. Supernatant was incubated with glutathione sepharose resin (GE Healthcare) and washed with lysis buffer without the detergent. Bound proteins were either eluted with freshly prepared buffer containing 25 mM reduced glutathione and 1 mM PMSF to obtain GST-tagged proteins or cleaved with PreScission protease (GE Healthcare) to obtain untagged proteins. GST-tagged ubiquitin was cleaved with thrombin to obtain untagged ubiquitin which were used for enzymatic assays and a MALDI-TOF experiment. To purify His-tagged proteins, pellets were lysed in buffer containing 20 mM Tris-HCl (pH 8.0), 350 mM NaCl, 25 mM imidazole, 5% glycerol, 2 mM 2-mercaptoethanol and 0.05% NP-40. The soluble fraction was then bound to TALON resin (Clontech, Mountain View, CA, USA) and washed with wash buffer containing 500 mM NaCl. Bound proteins were eluted with 300 mM imidazole-containing buffer (pH 8.5) to obtain His-tagged proteins, which were then passed through Hi-Prep 26/10 Desalting column (GE Healthcare) to remove imidazole. His-Ube1 was further purified by anion exchange chromatography using a 5-mL HiTrap Q HP column (GE Healthcare). Proteins were

concentrated using Amicon Ultra Centrifugal Filter units (Millipore, Burlington, MA, USA).

In vitro ubiquitination assay

In vitro ubiquitination assays were performed as described previously [32,43]. In brief, a 25- μ L reaction volume was constituted by incubating together His-Ube1 (0.3 μ M), His-UbcH6c (15 μ M), ubiquitin (35 μ M), GST-UBR7-PHD (0.15 μ M) WT or mutants and H2A/H2B dimer (5 μ M) in 50 mM Tris-HCl (pH 7.6), 5 mM MgCl₂, 2 mM ATP and 2 mM DTT and this was allowed to proceed at 30 °C for 30 min. In the case of the ubiquitination assay with H2B peptides as substrate, the assay time was increased to 6 h. The reaction was quenched by adding 5 μ L of 5 \times Laemmli buffer with separation on 15% SDS/PAGE gels. Ubiquitinated proteins were analysed by western blotting using specific antibodies for H2BK120Ub (#5546; Cell Signaling Technology, Danvers, MA, USA). Substrate was detected with α -H2B (#1790; Abcam, Cambridge, MA, USA), α -GST (#A190-122A; Bethyl Laboratories, Montgomery, TX, USA) and α -His horseradish peroxidase (#1187; Abcam) antibodies were used for affinity tagged E3 and E2s respectively.

In the case of single-turnover substrate assay, E2~Ub thioester was prepared first by incubating ubiquitin, Ube1 and UbcH6c in the buffer mentioned above in the absence of E3 UBR7-PHD and substrate H2A/H2B. The UbcH6c~Ub species was purified through size exclusion chromatography using a Superdex 75 pg column (GE Healthcare) and stored at –80 °C. For the assay, 10 μ M of UbcH6c~Ub was incubated with 2 μ M GST-PHD and 5 μ M H2A/H2B dimer in ATP-less buffer. The reactions were stopped at specific time points and product formation was analysed by western blotting. The experiments were repeated three times and the average of the values for each time point were plotted. Separate gels were run for blotting H2BK120Ub and GST proteins as a result of overlapping molecular weight. The blots were subsequently quantified using IMAGEJ.

Thioester hydrolysis assay

UbcH6c~Ub thioesters were prepared and purified as described above and the single-turnover discharge assay was performed in accordance with a previously established protocol [35]. Next, 10 μ M UbcH6c~Ub thioester was mixed with GST-tagged UBR7-PHD or RNF20 RING (2 μ M), in the presence or absence of H2B C-terminal peptides as substrate, followed by incubation at room temperature for different time points. Laemmli buffer was added to stop the reactions and loaded onto a 15% SDS/PAGE gel, and the gels were further analysed by western blotting using α -ubiquitin (#43124; Cell Signaling Technology) antibody.

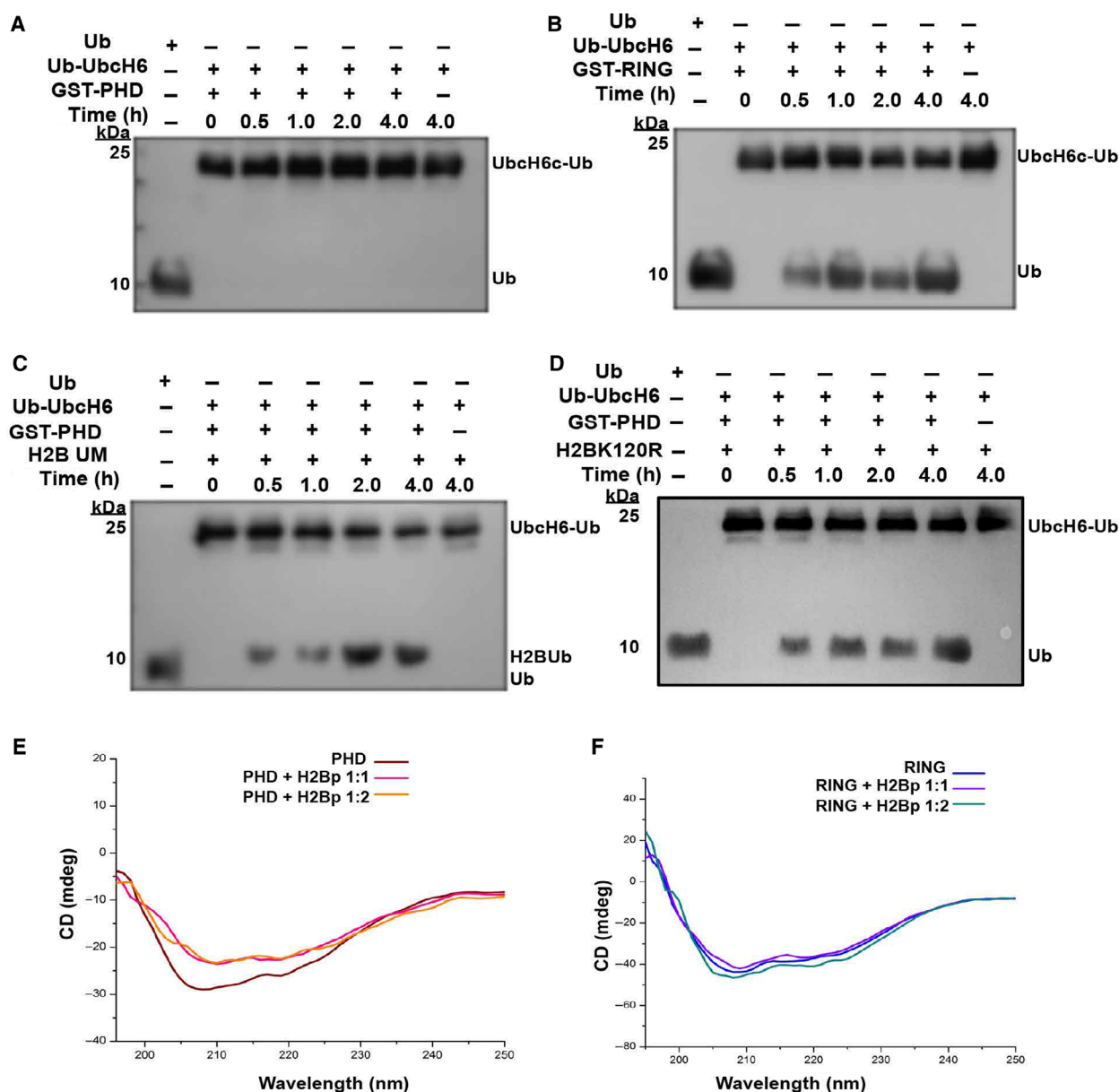


Fig. 5. The role of histone H2B substrate in UBR7-PHD mediated hydrolysis of UbchH6c-Ub thioester bond. (A–D) UbchH6c-Ub thioester conjugate was incubated without (A, B) or with (C, D) histone H2B peptide substrates and the ubiquitin transfer was measured by western blotting with anti-Ub antibody. (A, B) Single-turnover discharge assay with GST-PHD (A) or GST-RING (B) without substrate histone H2B. (C, D) Single-turnover discharge assay with UBR7-PHD in the presence of substrate C-terminal peptide of histone H2B (114–125) (C) or its mutant (H2B K120R) (D). (E, F) Far-UV CD spectra of untagged UBR7-PHD (E) and RNF20 RING (F) with and without peptide.

Affinity pull-down assays

GST and nickel-nitrilotriacetic acid pull-down assays were performed as described in standard protocols [44,45]. In brief, GST-tagged proteins along with the binding proteins in a equal molar ratio were incubated in binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% NP-40, 1 mM DTT) at 4 °C, then pulled down by glutathione Sepharose beads pre-blocked with binding buffer

containing 5% BSA. The beads after binding were washed thrice and run on 15% SDS/PAGE and further analysed by western blot. For nickel-nitrilotriacetic acid pull-down, FLAG-UBR7 WT or deletion constructs were overexpressed in HeLa cells. His-tagged UbchH6-full length was extramurally added to the cell lysate, followed by nickel-nitrilotriacetic acid pull-down. The beads were washed with 25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 20 mM imidazole

and 0.1% NP-40, three times, and subjected to western blotting for analysis. Twenty percent inputs were loaded. The nickel-nitrilotriacetic acid pull-down pull-down assays with UbCH6c and untagged PHD WT and mutant proteins were performed similarly.

Peptide pull-down assay

Biotinylated H2B C-terminal peptides (114–125) WT and K120R were synthesized by GL Biochem (Shanghai) Ltd (Shanghai, China). Peptide pull-down assays were carried out in accordance with previously established protocols [46]. In brief, GST-tagged UBR7-PHD was incubated with biotinylated H2B peptides in assay buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% NP-40, 1 mM DTT) at 4 °C, overnight. The reactions were thereafter pulled by streptavidin sepharose beads (GE Healthcare) and washed with the same assay buffer, and the protein bound to the beads was loaded onto a 12% SDS/PAGE gel and analyzed by western blotting.

Fluorescence spectroscopy

The spectroscopic titrations were performed using a RF-5301 PC spectrofluorimeter (Shimadzu, Kyoto, Japan) at 25 °C. The fluorescence measurements were carried out using 0.3 μM of the histone peptides dissolved in 20 mM Tris-HCl (pH 6.5), 150 mM NaCl and 1 mM β-Me with increasing concentration of UBR7-full length or PHD finger proteins. Samples were excited at 494 nm and emission was recorded at 520 nm with a scan speed of 200 nm·s⁻¹. The data were averaged over three scans. The 5'-FAM-conjugated peptide H2B (114–125) was synthesized by GL Biochem (Shanghai) Ltd. The data were analysed according to the method described previously [47,48].

$$L + P \rightleftharpoons LP$$

From the above equilibrium state of protein-ligand binding, where ligand (*L*) is UBR7-full length or PHD and *P* is the FAM-conjugated H2B peptide, and the apparent dissociation constant (*K_d*) was determined using non-linear curve fitting analysis (Eqn 1) All experimental points of the observed isotherms were fitted by least squares analysis as shown in Eqn (1):

$$C_0(\Delta F/\Delta F_{\max})^2 - (C_0 + C_p + K_d)(\Delta F/\Delta F_{\max}) + C_p = 0 \quad (1)$$

where *F* is the change in fluorescence emission intensity at 520 nm (excitation of FAM fluorophores at 494 nm) for each point of the titration curve, *F_{max}* is the change in emission intensity when the ligand (UBR7-full length or PHD) saturates the binding to histone peptide, *C_p* is the concentration of ligand and *C₀* is the initial concentration

of the histone peptide. *F_{max}* was, in turn, determined from the double-reciprocal plot:

$$1/\Delta F = 1/\Delta F_{\max} + K_d(\Delta F_{\max}(C_p - C_0)) \quad (2)$$

SEC-MALS

SEC-MALS experiments were performed with purified untagged UBR7-PHD WT or mutant proteins using an ÄKTA-MALS system (GE Healthcare). Some 200 μg of protein was injected in a Superdex 200 10/300 GL column (GE Healthcare) pre-equilibrated in 10 mM Tris (pH 7.5) and 200 mM NaCl at a flow rate of 0.25 mL·min⁻¹. Light scattering was monitored using a DAWN system (Wyatt Technology, Santa Barbara, CA, USA) and the concentration was measured using an Optilab differential refractometer (Wyatt Technology). Molar masses of proteins were calculated using ASTRA, version 7.3.2 (Wyatt Technology).

DFDNB-mediated cross-linking assay

Untagged UBR7-PHD WT and mutant proteins were purified in Hepes-Na (pH 7.5) buffer. Cross-linker DFDNB (Thermo Scientific, Waltham, MA, USA) was dissolved in DMSO at a concentration of 500 μM. DFDNB was added to 5 μg of UBR7-PHD WT and mutant protein in cross-linking buffer containing 20 mM Hepes-Na (pH 7.5) and 150 mM NaCl with a final concentration of 25 (1×) and 50 (2×) μM and the reaction mixture was quenched by adding 5 μL of 5 × Laemmli buffer and separated on a 15% SDS/PAGE gel. The bands were visualized by Coomassie Brilliant blue staining.

CD spectroscopy

CD spectra were recorded using a J-815 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier temperature control unit, which was set to 25 °C. Untagged UBR7-PHD or RNF20 RING were pre-incubated on ice with H2B C-terminal (114–125) peptide at the ratio 1 : 1 or 1 : 2. Far-UV CD spectra of the proteins with or without the peptides were recorded from 195 to 250 nm at room temperature using a 1-mm path quartz cuvette, a resolution of 0.2 nm and a scan rate of 50 nm·min⁻¹. Each spectrum was the average of the three consecutive scans. CD spectra of UbCH6c WT and its loop deletion mutants were recorded similarly.

Cell culture

HeLa cells were grown and maintained in Dulbecco's modified Eagle's medium (Gibco) supplemented with 10% FBS (Gibco, Amarillo, TX, USA), 1% antibiotic-antimycotic

(Gibco) and 1% essential amino acid at 37 °C incubator supplied with 5% CO₂.

Gene overexpression

HeLa cells were either singly transfected with EGFP-UBR7 full-length construct or co-transfected with FLAG-UBR7 full-length or the deletion constructs using Lipofectamine 2000 (Invitrogen, Waltham, MA, USA) in accordance with the manufacturer's instruction. Cells were incubated at 37 °C with supply of 5% CO₂. After 24 h of transfection, cells were harvested for further experiments.

Co-immunoprecipitation and western blot analysis

For the co-immunoprecipitation assay, previously established protocols were followed [49]. Briefly, the cells were resuspended in the cell lysis buffer containing 50 mM Hepes (pH 7.5), 150 mM NaCl, 1.5 mM MgCl₂, 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 0.5% sodium deoxycholate, 5% glycerol, 1 mM DTT and complete protease inhibitor cocktail. They were incubated on ice for 1 h followed by intermittent sonication. After centrifugation at 20 199 g for 10 min at 4 °C, the lysates were subjected to pre-clearing with normal sheep serum. The pre-cleared lysate was incubated overnight with the EGFP antibody (#A01704; GenScript, Piscataway, NJ, USA) followed by incubating with Protein A Sepharose beads (GE Healthcare). The beads were washed thrice with the same buffer after binding and samples were run on SDS/PAGE for western blot analysis. Immobilon nitrocellulose membrane (catalogue. no. HATF00010; Millipore) was used for transfer of proteins with 20% methanol containing buffer for 120 min at 4 °C. Membranes were blocked in 5% BSA and incubated with specific primary antibodies overnight at 4 °C. Blots were next washed thrice with 1 × Tris-buffered saline containing 0.1% Tween-20 and incubated in anti-rabbit (dilution 1 : 15 000) or anti-mouse (dilution 1 : 5000) horseradish peroxidase secondary antibody (Bethyl Laboratories) for 1 h at room temperature. Blots were developed after three more washes with 1 × Tris-buffered saline containing 0.1% Tween-20 in a C400 ChemiDoc imager (Azure Biosystems, Dublin, CA, USA) using Super Signal West Pico Plus Chemiluminescent Substrate (catalogue. no. 34580; Thermo Scientific).

Protein structural modelling

The X-ray structure of BPTF-PHD finger of NURF (PDB ID: 2F6N) obtained from the RCSB Protein Data Bank (<https://www.rcsb.org/>) was used as template for structural modelling of the UBR7-PHD finger. The UBR7-PHD amino acid sequence was threaded onto the structure using

the PHYRE2 algorithm [50]. The quality of the model obtained in PHYRE2.0 was evaluated with the PROQ2 tool [51]. The structural figures were created using PYMOL (Schrödinger, LLC, New York, NY, USA). The two monomers constituting the dimer were differentially coloured and the amino acid residues potentially involved in dimerization were highlighted.

MALDI-TOF mass spectrometry analysis

Samples from UBR7-PHD mediated multi-turnover ubiquitination assays using thrombin cleaved ubiquitin with H2B C-terminal WT and K120 mutant peptides were spotted in triplicate on a MTP 384 BC target and mass analysis was performed using a 4700 Proteomics Analyzer with TOF/TOF optics (Applied Biosystems, Waltham, MA, USA) in reflector mode using a saturated solution of CHCA (α -cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 0.1% trifluoroacetic acid. Analysis of the peaks in the range of 8.5–11 kDa was performed using PEAK EXPLORER (Agilent Technologies, Santa Clara, CA, USA). Prominent peaks were annotated and compared to the expected molecular weight of ubiquitin and ubiquitinated peptides for consistency with the observed molecular weight of the proteins obtained.

Acknowledgements

GST-tagged RNF20 RING construct and recombinant His-tagged Human H2A and H2B constructs were kind gifts from Dr Catherine L. Day (University of Otago, Dunedin) and Dr Shin-ichi Tate (Hiroshima University) respectively. We thank Mr. Sandip Chakraborty and Mr Jishu Mondal of Central Instrumentation Facility division, CSIR-IICB for their assistance with the MALDI-TOF mass spectrometry and Circular Dichroism experiments respectively. We thank Ms Purbita Saha and Mr Hari Shankar Sunil of UT Southwestern Medical Center, Texas for their help with graphical image created by BioRender. This work was supported by research grants received from Council of Scientific and Industrial Research (CSIR)-Network Project (UNSEEN) to SR and Swarna Jayanti Fellowship (DST/SJF/LSA-02/2017-18), Department of Science and Technology to CD. ADG and SD are supported by SRF Fellowship through the University Grant Commission (UGC) and CSIR respectively.

Conflict of interest

The authors declare that they have no conflicts of interest.

Author contributions

AD, CD and SR conceived the study, designed experiments, analysed data, and wrote the manuscript. AD, PM and SD performed experiments and analysed the data. All authors reviewed the results and approved the final version of the manuscript submitted for publication.

References

- Kornberg RD & Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294.
- Felsenfeld G & McGhee JD (1986) Structure of the 30 nm chromatin fiber. *Cell* **44**, 375–377.
- Luger K, Mäder AW, Richmond RK, Sargent DF & Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
- Allan J, Harborne N, Rau DC & Gould H (1982) Participation of core histone “tails” in the stabilization of the chromatin solenoid. *J Cell Biol* **93**, 285–297.
- Zheng C & Hayes JJ (2003) Intra- and inter-nucleosomal protein-DNA interactions of the core histone tail domains in a model system. *J Biol Chem* **278**, 24217–24224.
- Strahl BD & Allis CD (2000) The language of covalent histone modifications. *Nature* **403**, 41–45.
- Fischle W, Wang Y & Allis CD (2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol* **15**, 172–183.
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* **128**, 693–705.
- Bannister AJ & Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* **21**, 381–395.
- Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS & Zhang Y (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873–878.
- Minsky N, Shema E, Field Y, Schuster M, Segal E & Oren M (2008) Monoubiquitinated H2B is associated with the transcribed region of highly expressed genes in human cells. *Nat Cell Biol* **10**, 483–488.
- Xiao T, Kao C-F, Krogan NJ, Sun Z-W, Greenblatt JF, Osley MA & Strahl BD (2005) Histone H2B ubiquitylation is associated with elongating RNA polymerase II. *Mol Cell Biol* **25**, 637–651.
- Kim H, Chen J & Yu X (2007) Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* **316**, 1202–1205.
- Weake VM & Workman JL (2008) Histone ubiquitination: triggering gene activity. *Mol Cell* **29**, 653–663.
- Goldknopf IL & Busch H (1977) Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate protein A24. *Proc Natl Acad Sci USA* **74**, 864–868.
- Nickel BE & Davie JR (1989) Structure of polyubiquitinated histone H2A. *Biochemistry* **28**, 964–968.
- Cao R, Tsukada YI & Zhang Y (2005) Role of Bmi-1 and Ring1A in H2A ubiquitylation and hox gene silencing. *Mol Cell* **20**, 845–854.
- Buchwald G, Van Der Stoop P, Weichenrieder O, Perrakis A, Van Lohuizen M & Sixma TK (2006) Structure and E3-ligase activity of the Ring-Ring complex of Polycomb proteins Bmi1 and Ring1b. *EMBO J* **25**, 2465–2474.
- Zhu Q, Pao GM, Huynh AM, Suh H, Tonnu N, Nederlof PM, Gage FH & Verma IM (2011) BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* **477**, 179–184.
- Zhou W, Zhu P, Wang J, Pascual G, Ohgi KA, Lozach J, Glass CK & Rosenfeld MG (2008) Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol Cell* **29**, 69–80.
- Robzyk K, Recht J & Osley MA (2000) Rad6-dependent ubiquitination of histone H2B in yeast. *Science* **287**, 501–504.
- Wood A, Krogan NJ, Dover J, Schneider J, Heidt J, Boateng MA, Dean K, Golshani A, Zhang Y, Greenblatt JF *et al.* (2003) Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell* **11**, 267–274.
- Zhu B, Zheng Y, Pham AD, Mandal SS, Erdjument-Bromage H, Tempst P & Reinberg D (2005) Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol Cell* **20**, 601–611.
- Fierz B, Chatterjee C, McGinty RK, Bar-Dagan M, Raleigh DP & Muir TW (2011) Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction. *Nat Chem Biol* **7**, 113–119.
- Debelouchina GT, Gerecht K & Muir TW (2017) Ubiquitin utilizes an acidic surface patch to alter chromatin structure. *Nat Chem Biol* **13**, 105–110.
- Chandrasekharan MB, Huang F & Sun ZW (2010) Histone H2B ubiquitination and beyond: regulation of nucleosome stability, chromatin dynamics and the trans-histone H3 methylation. *Epigenetics* **5**, 460–468.
- Wang E, Kawaoka S, Yu M, Shi J, Ni T, Yang W, Zhu J, Roeder RG & Vakoc CR (2013) Histone H2B ubiquitin ligase RNF20 is required for MLL-rearranged leukemia. *Proc Natl Acad Sci USA* **110**, 3901–3906.
- Cole AJ, Clifton-Bligh R & Marsh DJ (2015) Histone H2B monoubiquitination: roles to play in human malignancy. *Endocr Relat Cancer* **22**, T19–T33.
- Johnsen SA (2012) The enigmatic role of H2Bub1 in cancer. *FEBS Lett* **586**, 1592–1601.

- 30 Fuchs G & Oren M (2014) Writing and reading H2B monoubiquitylation. *Biochim Biophys Acta* **1839**, 694–701.
- 31 Kim J, Hake SB & Roeder RG (2005) The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol Cell* **20**, 759–770.
- 32 Wheaton K, Sarkari F, Johns BS, Davarinejad H, Egorova O, Kaustov L, Raught B, Saridakis V & Sheng Y (2017) UBE2E1/UBCH6 is a critical in vivo E2 for the PRC1-catalyzed ubiquitination of H2A at Lys-119. *J Biol Chem* **292**, 2893–2902.
- 33 Wang L, Cao C, Wang F, Zhao J & Li W (2017) H2B ubiquitination: Conserved molecular mechanism, diverse physiologic functions of the E3 ligase during meiosis. *Nucleus* **8**, 461–468.
- 34 Turco E, Gallego LD, Schneider M & Köhler A (2015) Monoubiquitination of histone H2B is intrinsic to the Bre1 RING domain-Rad6 interaction and augmented by a second Rad6-binding site on Bre1. *J Biol Chem* **290**, 5298–5310.
- 35 Foglizzo M, Middleton AJ & Day CL (2016) Structure and function of the RING domains of RNF20 and RNF40, Dimeric E3 ligases that monoubiquitylate histone H2B. *J Mol Biol* **428**, 4073–4086.
- 36 Adhikary S, Chakravarti D, Terranova C, Sengupta I, Maitituoheti M, Dasgupta A, Srivastava DK, Ma J, Raman AT, Tarco E *et al.* (2019) Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor. *Nat Commun* **10**, 1398.
- 37 Tasaki T, Mulder LCF, Iwamatsu A, Lee MJ, Davydov IV, Varshavsky A, Muesing M & Kwon YT (2005) A family of mammalian E3 ubiquitin ligases that contain the UBR box motif A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol Cell Biol* **25**, 7120–7136.
- 38 Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS & Kwon YT (2009) The substrate recognition domains of the N-end rule pathway. *J Biol Chem* **284**, 1884–1895.
- 39 Musselman CA & Kutateladze TG (2011) Handpicking epigenetic marks with PHD fingers. *Nucleic Acids Res* **39**, 9061–9071.
- 40 Sanchez R & Zhou MM (2011) The PHD finger: a versatile epigenome reader. *Trends Biochem Sci* **36**, 364–372.
- 41 Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E *et al.* (2012) The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404.
- 42 Liu H & Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol* **8**, 91.
- 43 Plechanovov A, Jaffray EG, Tatham MH, Naismith JH & Hay RT (2012) Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature* **489**, 115–120.
- 44 Busschots K, Vercammen J, Emiliani S, Benarous R, Engelborghs Y, Christ F & Debyser Z (2005) The interaction of LEDGF/p75 with integrase is lentivirus-specific and promotes DNA binding. *J Biol Chem* **280**, 17841–17847.
- 45 Malovannaya A, Lanz RB, Jung SY, Bulynko Y, Le NT, Chan DW, Ding C, Shi Y, Yucer N, Krenciute G *et al.* (2011) Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787–799.
- 46 Wy J (2006) Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* **40**, 339–343.
- 47 Adhikary S, Sanyal S, Basu M, Sengupta I, Sen S, Srivastava DK, Roy S & Das C (2016) Selective recognition of H3.1K36 dimethylation/H4K16 acetylation facilitates the regulation of all-trans-retinoic acid (ATRA)-responsive genes by putative chromatin reader ZMYND8. *J Biol Chem* **291**, 2664–2681.
- 48 Chakrabarti S, Bhattacharyya D & Dasgupta D (2000) Structural basis of DNA recognition by anticancer antibiotics, chromomycin A3, and mithramycin: Roles of minor groove width and ligand flexibility. *Biopolymers* **56**, 85–95.
- 49 Mondal P, Sen S, Klein BJ, Tiwary N, Gadad SS, Kutateladze TG, Roy S & Das C (2020) TCF19 promotes cell proliferation through binding to the histone H3K4me3 mark. *Biochemistry* **59**, 389–399.
- 50 Kelley LA, Mezulis S, Yates CM, Wass MN & Sternberg MJE (2015) The Pyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**, 845–858.
- 51 Ray A, Lindahl E & Wallner B (2012) Improved model quality assessment using ProQ2. *BMC Bioinform* **13**, 1–12.

ARTICLE

<https://doi.org/10.1038/s41467-019-08986-5>

OPEN

Atypical plant homeodomain of UBR7 functions as an H2BK120Ub ligase and breast tumor suppressor

Santanu Adhikary^{1,2}, Deepavali Chakravarti^{3,4}, Christopher Terranova³, Isha Sengupta¹, Mayinuer Maitituoheti³, Anirban Dasgupta², Dushyant Kumar Srivastava², Junsheng Ma⁵, Ayush T. Raman³, Emily Tarco⁶, Aysegul A. Sahin⁷, Roland Bassett⁵, Fei Yang⁶, Coya Tapia⁷, Siddhartha Roy², Kunal Rai³ & Chandrima Das¹

The roles of Plant Homeodomain (PHD) fingers in catalysis of histone modifications are unknown. We demonstrated that the PHD finger of Ubiquitin Protein Ligase E3 Component N-Recognin7 (UBR7) harbors E3 ubiquitin ligase activity toward monoubiquitination of histone H2B at lysine120 (H2BK120Ub). Purified PHD finger or full-length UBR7 monoubiquitinated H2BK120 in vitro, and loss of UBR7 drastically reduced H2BK120Ub genome-wide binding sites in MCF10A cells. Low UBR7 expression was correlated with occurrence of triple-negative breast cancer and metastatic tumors. Consistently, UBR7 knockdown enhanced the invasiveness, induced epithelial-to-mesenchymal transition and promoted metastasis. Conversely, ectopic expression of UBR7 restored these cellular phenotypes and reduced tumor growth. Mechanistically, UBR7 loss reduced H2BK120Ub levels on cell adhesion genes, including CDH4, and upregulated the Wnt/ β -Catenin signaling pathway. CDH4 overexpression could partially revert UBR7-dependent cellular phenotypes. Collectively, our results established UBR7 as a histone H2B monoubiquitin ligase that suppresses tumorigenesis and metastasis of triple-negative breast cancer.

¹Biophysics and Structural Genomics Division, Saha Institute of Nuclear Physics, 1/AF Bidhan Nagar, Kolkata 700064, India. ²Structural Biology and Bio-Informatics Division, CSIR-Indian Institute of Chemical Biology, 4 Raja S.C. Mullick Road, Kolkata 700032, India. ³Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. ⁴Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. ⁵Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. ⁶Department of Translational Molecular Pathology and Department of Investigational Cancer Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. ⁷Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. These authors contributed equally: Santanu Adhikary, Deepavali Chakravarti, Christopher Terranova. Correspondence and requests for materials should be addressed to S.R. (email: roysiddhartha@iicb.res.in) or to K.R. (email: krai@mdanderson.org) or to C.D. (email: chandrima.das@saha.ac.in)

Breast cancer is the most common cause of cancer mortality in female individuals. The heterogeneity of the disease poses immense challenges in deciphering therapeutic strategies¹. The hormone receptor-negative or triple-negative subtype has the worst prognoses due to the lack of targeted therapies^{2,3}. Although accumulation of genetic defects has been involved in the development of oncogenesis, epigenetic abnormalities play a significant role in the initiation, progression, and metastasis of the disease⁴. Specifically, epithelial-to-mesenchymal transition (EMT), which preludes the onset of metastasis^{5,6}, is thought to be driven by epigenetic alterations^{7,8}. Primarily, histone modifications, which include methylation, acetylation, and ubiquitination, play crucial roles in maintaining homeostasis, failure of which may lead to disease initiation or progression⁹. Importantly, breast cancer cases with worse prognoses have lower levels of H3K18Ac, H4K12Ac, H3K4Me2, H4K20Me3, and H4R3Me2 marks^{9,10}.

Monoubiquitination at lysines on histones H2A and H2B have an antagonistic relationship in oncogenesis^{11–14}. In contrast with polyubiquitination, which marks the protein for its proteasome-mediated degradation, monoubiquitination of histone H2B plays key roles in transcription memory and elongation, DNA damage response, viral infection, stem cell differentiation, and oncogenesis^{15,16}. The E3 ligases for H2B monoubiquitination, RNF20 and RNF40, are reported to act as potent tumor suppressors, regulate DNA double-stranded break repair, and modulate stem cell differentiation^{14,17,18}.

The ubiquitin protein ligase E3 component N-recognin (UBR) family of mammalian E3 ligases containing seven members (UBR1–UBR7) is characterized by a 70-residue zinc finger-type UBR-box domain, which is essential for recognition of the N-degrons^{19–21}. Despite harboring a UBR-box, UBR3, UBR6, and UBR7 do not bind to N-degrons. Although the members of the UBR family of proteins are generally heterogeneous in size and sequence, they harbor specific signatures unique to ubiquitin ligases or a substrate-recognition subunit of the E3 complex like the RING/HECT (really interesting new gene/homologous to the E6AP carboxyl terminus) domain or F-box^{21,22}. A RING domain is present in UBR1, UBR2, and UBR3, a HECT domain is present in UBR5, and an F-box is present in UBR6. Of note, UBR7 has evolved with a plant homeodomain (PHD) finger, which is a putative chromatin-binding module, not present in any other UBR family proteins. Although the PHD finger is well characterized as a reader of methylated, acetylated, or unmodified histones²³, its role in enzymatic catalysis is not known. Furthermore, little is known about the role of UBR7 in carcinogenesis. In the present study, we demonstrated that the UBR7-PHD finger is an H2BK120 monoubiquitin ligase and a tumor suppressor in triple-negative breast cancer cases.

Results

UBR7-PHD monoubiquitinates histone H2B lysine 120 in vitro. UBR7, a protein with an unknown function, contains a UBR-box domain, which is essential for the recognition of N-degrons^{20–22}, and a PHD finger (Fig. 1a), which is highly conserved across species (Supplementary Fig. 1a). Although the sequence alignment of the UBR7-PHD finger (which is stabilized by zinc ion coordination in a cross-braced topology) with other well-characterized H3K4Me3 or H3K4Me0 binders exhibited several conserved residues, it displayed weak interaction with trimethylated histone peptides (Supplementary Fig. 1b–f). Although full-length UBR7 protein interacted with all recombinant histones in vitro, the PHD finger preferentially interacted with recombinant histone H2B (Fig. 1b) and could also immunoprecipitate them from MCF10A cells (Supplementary Fig. 1g). Sequence alignment of the UBR7-PHD with other classical RING

finger-E3 ubiquitin ligases indicated that zinc-coordinating His163 and His166 are unique in contrast to the other RING fingers (Supplementary Fig. 1b). Site-directed mutagenesis of H163S/H166S of the UBR7-PHD did not significantly compromise its association with histone H2B at various levels of organization (Fig. 1c). We observed similar results during immunoprecipitation assays from HEK293T cells (Fig. 1d). However, mutation of lysine 120 to arginine (K120R) in histone H2B abrogated its binding preference for UBR7 as observed through immunoprecipitation assays (Fig. 1e). Based on the zinc coordination fold similarity between the RING and the PHD finger, we hypothesized that UBR-PHD function as an E3 ubiquitin ligase for histone H2B substrate. Purified recombinant UBR7 full-length wild-type protein (UBR7-WT), full-length H163S/H166S catalytic-mutant (UBR7-CM), or individual domains (UBR or PHD) were incubated in the presence of an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme (UbcH6), ATP, inorganic pyrophosphatase, and biotin-tagged ubiquitin. UBR7-WT and the PHD finger alone could monoubiquitinate purified H2B, H2A/H2B dimer, core histone octamers, and purified nucleosomes (Fig. 1f, g and Supplementary Fig. 1h–k), in contrast with the other E3 ubiquitin ligases, which usually act in complex^{24–26}. On the other hand, UBR7-CM failed to promote H2B ubiquitination (Fig. 1f, g and Supplementary Fig. 1j–o). Thus, our results demonstrated the E3 ubiquitin ligase function of UBR7 toward monoubiquitination of histone H2B in vitro.

UBR7 regulates H2BK120Ub levels ex vivo. Next, we sought to determine whether UBR7 regulates H2BK120Ub in mammalian cells. We observed lower UBR7 levels in human and murine breast cancer cells than in their “normal” counterparts. For example, UBR7 messenger RNA (mRNA) transcript and protein levels were lower in MCF7, T47D, MDA-MB-231, and MDA-MB-468 cells than in MCF10A and MCF12A cells (Fig. 2a and Supplementary Fig. 2a). We observed similar patterns in 21PT and 21MT2 compared to 16N (human) and 4T07 compared to 4T1 (murine) cells (Supplementary Fig. 2b–e). Genetic depletion of UBR7 by two different short hairpin RNAs (shRNAs) in MCF10A and MCF12A cells led to a dramatic reduction in global H2BK120Ub levels (Fig. 2b and Supplementary Fig. 2f, g). Importantly, the reduction caused by UBR7-shRNA was rescued by UBR7-WT, but not by UBR7-CM (Fig. 2c and Supplementary Fig. 2h). Consistently, UBR7-WT, but not UBR7-CM, increased H2BK120Ub levels in the MDA-MB-231 and MDA-MB-468 breast cancer cells (Fig. 2d, e and Supplementary Fig. 2h).

To examine changes in H2BK120Ub genome wide, we performed chromatin immunoprecipitation sequencing (ChIP-seq) of H2BK120Ub in control and UBR7-knockdown MCF10A cells. UBR7-knockdown cells had a drastically lower number of H2BK120Ub binding sites (1079) compared to control cells (8401) (Fig. 2f and Supplementary Data 1), which was verified by ChIP-quantitative polymerase chain reaction (qPCR) (Fig. 2g). Similarly, the intensity of H2BK120Ub enrichment was drastically reduced as demonstrated by average density plots and verified by ChIP-qPCR for selected genes (Fig. 2h, i and Supplementary Fig. 2i, j).

To analyze the impact of UBR7-mediated H2BK120Ub on the chromatin landscape, we also performed ChIP-seq for histone modifications H3K79Me2 (transcription), H3K4Me3 (promoters), H3K4Me1 (enhancers), H3K27Ac (active enhancers), H3K27Me3 (polycomb-repressed), and H3K9Me3 (heterochromatin)²⁷. Consistent with prior reports, we noted a loss of H3K79Me2 on H2BK120Ub gene targets upon UBR7 knockdown (Fig. 2j). Identification of chromatin state transitions between control and UBR7-depleted cells using the ChromHMM

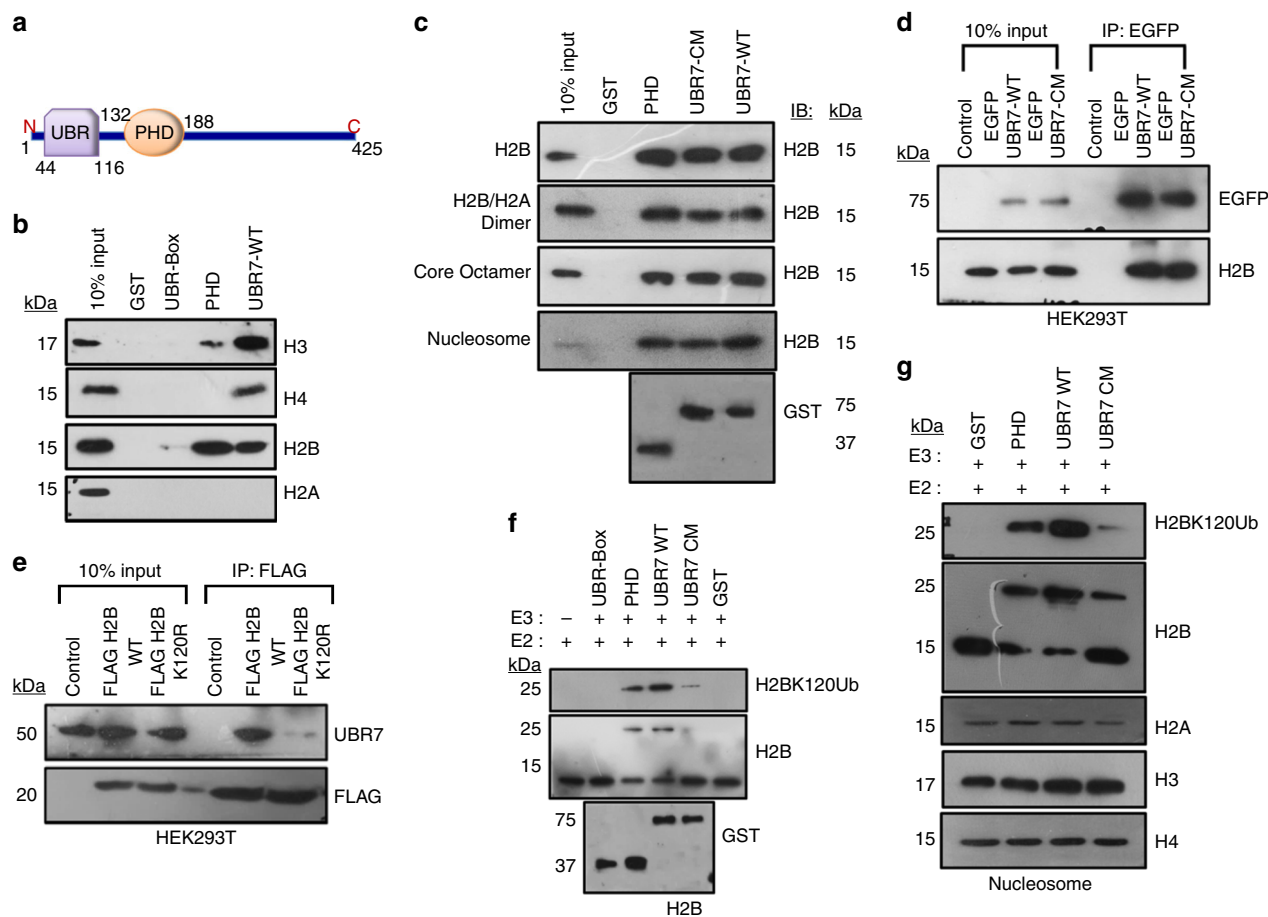


Fig. 1 UBR7 is a histone H2B ubiquitin ligase. **a** Schematic of the domain organization of UBR7. **b** Immunoblots showing the in vitro interaction of purified UBR7 with recombinant histones H3, H4, H2B, and H2A. **c** Interaction of UBR7 full-length wild-type (WT) or catalytic-mutant (CM) with recombinant H2B, H2A/H2B dimer, core octamer, or purified nucleosomes from HeLa cells. **d** Ex vivo interaction of H2B with UBR7 WT or CM in HEK293T cells. **e** Ex vivo interaction of UBR7 with H2B WT or H2B mutant (K120R) in HEK293T cells. **f, g** In vitro ubiquitination assay with recombinant H2B (**f**), or purified nucleosomes from HeLa cells (**g**)

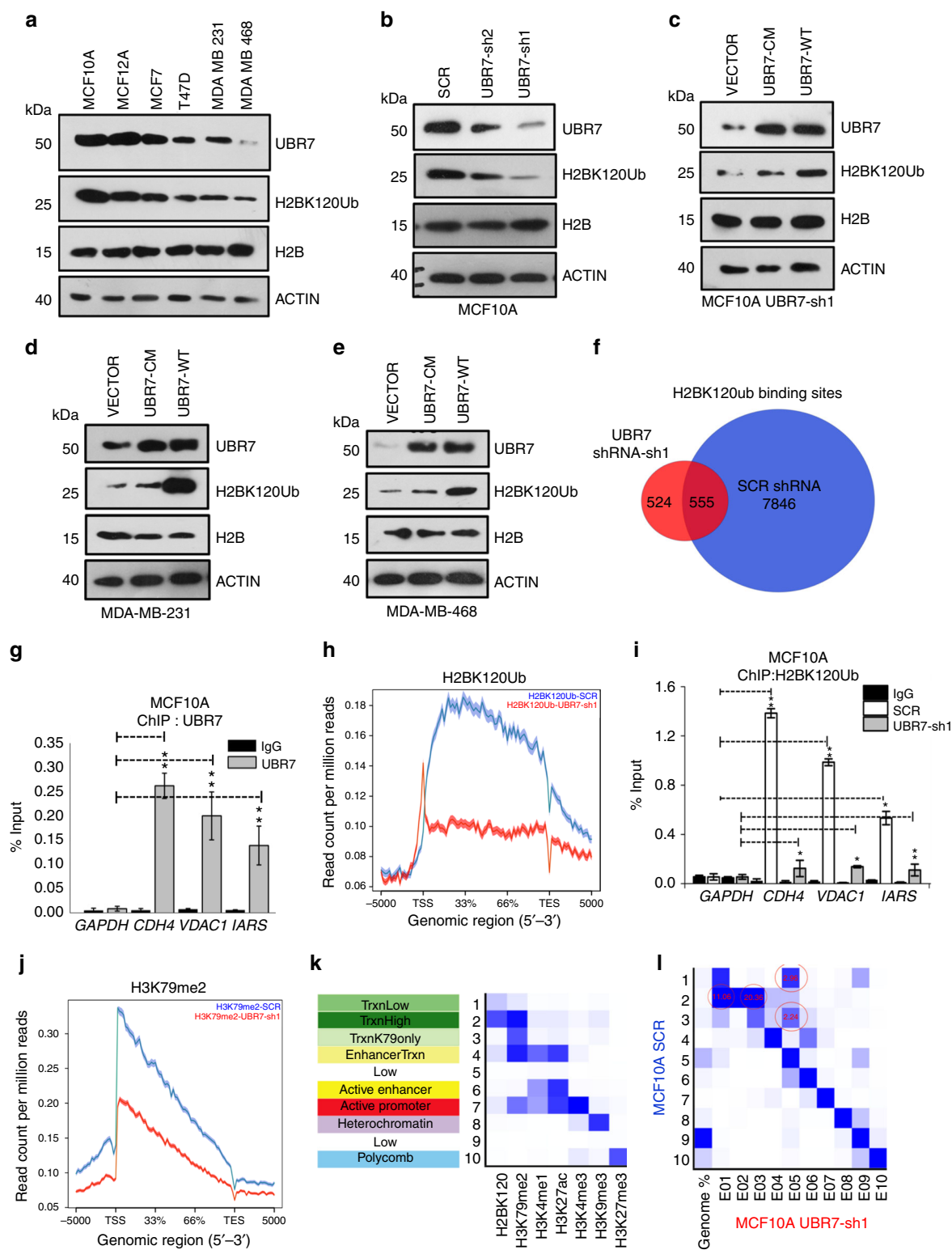
algorithm in a 10-state model (Fig. 2k; see Methods) indicated that H2BK120Ub was primarily present in conjunction with H3K79Me2. Also, the most potent transitions were from highly transcribed states in control cells to low/non-transcribed states in UBR7-depleted cells (States 1 to 5, States 2 to 4 or 1, States 3 to 5, and States 4 to 6) (Fig. 2k, l). Overall, these experiments identified UBR7 as an E3 ubiquitin ligase in vivo and demonstrated the importance of UBR7 in maintaining specific chromatin patterns in cells.

Next, we defined the relative impact of UBR7 and other two known H2BK120Ub E3 ligases, RNF20 and RNF40, on its levels and genomic distribution. Knockdown of all the three E3s, UBR7, RNF20, and RNF40, separately in MCF10A cells led to significant reduction of total H2BK120Ub levels (Supplementary Fig. 3a). Direct comparison between ChIP-seq profiles for H2BK120Ub showed drastic losses at 10,919 sites in UBR7-deficient, 11,005 sites in RNF20-deficient, and 11,069 sites in RNF40-deficient cells (Supplementary Fig. 3b–g). Sites that lose H2BK120Ub upon individual knockdown of each of these enzymes significantly overlapped (Supplementary Fig. 3h–m), suggesting cooperative interaction between UBR7, RNF20, and RNF40 and requirement of all three proteins for maintenance of H2BK120Ub levels in the cell.

Low UBR7 is associated with triple-negative breast cancer. Next, to further explore UBR7's biological function, we examined

its expression in The Cancer Genome Atlas Group (TCGA) breast cancer mRNA expression data²⁸. This analysis revealed reduced UBR7 expression in triple-negative and basal-like breast tumors (Fig. 3a), which was confirmed in an independent cohort of breast cancer tissue microarrays (TMAs) containing 371 breast tumors using a UBR7-specific antibody (Supplementary Fig. 4a and Supplementary Data 2). We found that lower levels of UBR7 correlated significantly with triple-negative status (Fig. 3b) and individual estrogen receptor (ER) status and progesterone receptor (PR) status (Fig. 3c, d and Supplementary Data 2). Also, metastatic intraductal carcinomas harbored considerably lower levels of UBR7 than did primary tumors (Fig. 3e, f). Furthermore, UBR7 was associated with a better metastasis-free survival rate in the aggressive basal-type breast cancer (mesenchymal subtype tumors without endocrine therapy) (Supplementary Fig. 4b).

UBR7 acts as tumor metastasis suppressor gene. A series of loss-of-function and gain-of-function experiments established the role of UBR7 as a tumor and metastasis suppressor gene. Reduction of UBR7 in MCF10A and MCF12A cells increased two-dimensional (2D) proliferation drastically (Fig. 3g and Supplementary Fig. 4c), as well as expression of the proliferation markers *Ki-67*, *PCNA*, and *MCM2*²⁹ (Fig. 3h–j and Supplementary Fig. 4d). Overexpression of UBR7-WT and not UBR7-CM reversed the increased proliferation observed in UBR7-knockdown MCF10A cells (UBR7-sh1) (Fig. 3k) and the



expression of associated proliferation markers (Fig. 3l and Supplementary Fig. 4e, f). Similarly, overexpression of UBR7-WT, but not UBR7-CM, in MDA-MB-231 and MDA-MB-468, two basal-like breast cancer cells that express UBR7 at low levels, dramatically reduced 2D proliferation (Supplementary Fig. 4g, h). Next, we examined three-dimensional (3D) soft agar colony growth of UBR7-depleted MCF10A cells in vitro and found

substantially higher numbers of colonies as well as bigger colonies in UBR7-depleted cells than in control (Fig. 4a, b). Importantly, the impact of wild-type derivatives in the 3D growth assay in MDA-MB-231 cells was abrogated by UBR7-CM (Fig. 4c, d). Overexpression of UBR7-WT and not UBR7-CM abrogated mammary fat pad tumor formation in vivo (Fig. 4e, f). Additionally, immunohistochemical analysis of the proliferation

Fig. 2 UBR7 is downregulated in invasive breast cancer cells. **a** Immunoblots of MCF10A, MCF7, T47D, MDA-MB-231, and MDA-MB-468 cells to monitor expression of UBR7, H2BK120Ub, and H2B. ACTIN was used as a loading control. **b** Immunoblots for UBR7, H2BK120Ub, H2B, and ACTIN (loading control) in MCF10A cells expressing scrambled (SCR), UBR7-sh1, or UBR7-sh2 shRNA. **c–e** Immunoblots for UBR7, H2BK120Ub, H2B, and ACTIN (loading control) in MCF10A UBR7-sh1 (**c**), MDA-MB-231 (**d**), and MDA-MB-468 (**e**) cells expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM) UBR7. ACTIN was used as a loading control. **f** Venn diagram showing overlap of total H2BK120Ub binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **g** Bar plot for quantitative PCR (qPCR) enrichment of UBR7 chromatin immunoprecipitation (ChIP) in MCF10A cells for selected genes. *GAPDH* was used as a negative control. **h** Average genebody density plot for H2BK120Ub binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **i** Bar plot for qPCR enrichment of H2BK120Ub ChIP in MCF10A cells expressing SCR or UBR7-sh1. *GAPDH* was used as a negative control. **j** Average genebody density plot for H3K79me2 binding sites in Control (SCR) and UBR7-sh1-expressing MCF10A cells. **k** Emission parameter for a 10-state chromatin state model called by the default parameters of ChromHMM. States in the left column were annotated based on their closeness to the nearest transcription start sites (TSS) and nature of constituent marks. **l** Overlap enrichment analysis displaying chromatin state transitions between MCF10A control (SCR) cells (Y-axis) and MCF10A UBR7-sh1 cells (X-axis). The most significant state transitions include losses of H2BK120Ub/H3K79me2 low (States 1 to 5), H2BK120Ub/H3K79me2 high (States 2 to 1 or 3) and H3K79me2 only (States 3 to 5), which are highlighted by red circles. In **g**, **i**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. * $P < 0.05$; ** $P < 0.001$

Table 1 Five chosen GO terms from DAVID on differentially expressed genes in UBR7-knockdown MCF10A cells (FDR <0.01; FC >2)

GO TERM	FDR
GO:0006414- translational elongation	2.0E-43
GO:0007049- cell cycle	5.3E-16
GO:0007155- cell adhesion	1.3E-05
GO:0031497- chromatin assembly	5.8E-05
GO:0016126- sterol biosynthesis process	1.5E-03

DAVID Database for Annotation, Visualization and Integrated Discovery, GO gene ontology, FC fold change, FDR false discovery rate

marker Ki-67, tumors derived from mice, confirmed the anti-proliferative role of UBR7-WT, but not UBR7-CM (Fig. 4g). These results supported a tumor-suppressive role for UBR7 in breast cancer.

Consistent with a metastasis-suppressive role for UBR7 and its lower levels in metastatic tumors, UBR7-WT overexpressing MDA-MB-231 cells, but not UBR7-CM-overexpressing MDA-MB-231 cells, were unable to seed to the lung upon intravenous injection, whereas control cells formed overt lung metastases (Fig. 4h). Consistently, UBR7 loss in MCF10A cells enhanced invasion in a Matrigel chamber (Fig. 4i, j) and migration in a scratch assay (Fig. 4k, l and Supplementary Fig. 5a, b). Importantly, these phenotypes were rescued by overexpression of UBR7-WT, but not UBR7-CM, in MCF10A UBR7-sh1 cells (Fig. 4m–p), as well as in MDA-MB-231 and MDA-MB-468 cells (Supplementary Fig. 5c–h). Overall, these results establish the tumor- and metastasis-suppressive functions of UBR7 in triple-negative breast cancer.

UBR7 suppresses EMT. To gain insight into the molecular mechanism of these observations, we performed transcriptomic profiling of MCF10A cells harboring control and UBR7-specific shRNA using RNA-sequencing (RNA-seq) and identified 2348 up-regulated and 2576 downregulated genes (Fig. 5a, Supplementary Fig. 6a, b and Supplementary Data 3). We verified a subset of these genes using individual qPCR experiments (Supplementary Fig. 6c, d). Consistent with the cellular phenotypes, the misregulated genes exhibited enrichment in Cadherin and invasive breast cancer signatures (Fig. 5b, Supplementary Data 4 and Table 1). Importantly, we noted that UBR7-low (shRNA harboring) cells had characteristics of epithelial-to-mesenchymal

transition (EMT), a cellular process typically associated with breast cancer metastasis⁵, as judged by loss of expression of epithelial markers (*CDH1*, *CLDN1*, *CLDN7*, and *CYTK18*) and gain of expression of mesenchymal markers (*CDH2*, *ZEB1*, *SNAIL1*, *SNAIL2*, *TWIST*, *VIM*)^{6,30} in qPCR, western blot, and immunofluorescent analyses (Fig. 5c–f and Supplementary Fig. 7a–c). Overexpression of UBR7-WT in MCF10A UBR7-shRNA harboring cells (UBR7-sh1) rescued the expression of these markers, whereas UBR7-CM did not (Fig. 5g–i). Similarly, overexpression of UBR7-WT, but not UBR7-CM, reduced the expression of mesenchymal markers and induced the expression of epithelial markers in MDA-MB-231 and MDA-MB-468 cells (Fig. 5j, k and Supplementary Fig. 7d–g). Therefore, our results indicated that loss of UBR7 promotes EMT, a phenomenon thought to precede metastasis.

UBR7 suppresses EMT through activation of CDH4. To define the UBR7 gene targets that may drive the observed phenotypes, we overlapped sites that harbored loss of H2BK120Ub in UBR7-knockdown cells using ChIP-seq (7846, $p < 1e-8$) with differentially expressed genes (4924; $p < 0.01$; fold change >1.5) and found that H2BK120Ub targeted 318 downregulated and 117 up-regulated genes (Fig. 6a, Supplementary Fig. 8a, b, and Supplementary Data 5). Several such genes were enriched in cell–cell adhesion processes (Supplementary Data 6) including cadherins such as *CDH4* and *CDH13* (Fig. 6b and Supplementary Fig. 8c–e). *CDH4* (or R-cadherin) is suggested to play important roles in suppressing invasion of basal-type breast cancer³¹. Furthermore, *CDH4* expression exhibited consistent patterns to UBR7 in matched “normal” and “malignant” breast cancer cell lines (Fig. 6c and Supplementary Fig. 8f). Importantly, we detected UBR7 occupancy in *CDH4* locus in control MCF10A cells using ChIP-qPCR as well as rescue of H2BK120Ub levels in MCF10A UBR7-sh1, MDA-MB-231, and MDA-MB-468 cells by overexpression of UBR7-WT, but not UBR7-CM (Figs 2g, 6d–f). Next, we sought to determine whether CDH4 was in part responsible for anti-invasive phenotypes of UBR7. Overexpression of CDH4 in MCF10A and MCF12A UBR7-sh1 cells drastically reduced cellular invasion, migration, proliferation, and suppressing EMT, and established an epistatic relationship between UBR7 and CDH4 (Fig. 6g–m and Supplementary Figs 8g–j, 9a,b). Similarly, CDH4 overexpression drastically reduced the invasive properties of both MDA-MB-231 and MDA-MB-468 cells (Supplementary Fig. 9c–i). Overall, our results indicated that functional loss of UBR7 can be restored by cell–cell adhesion genes like *CDH4*.

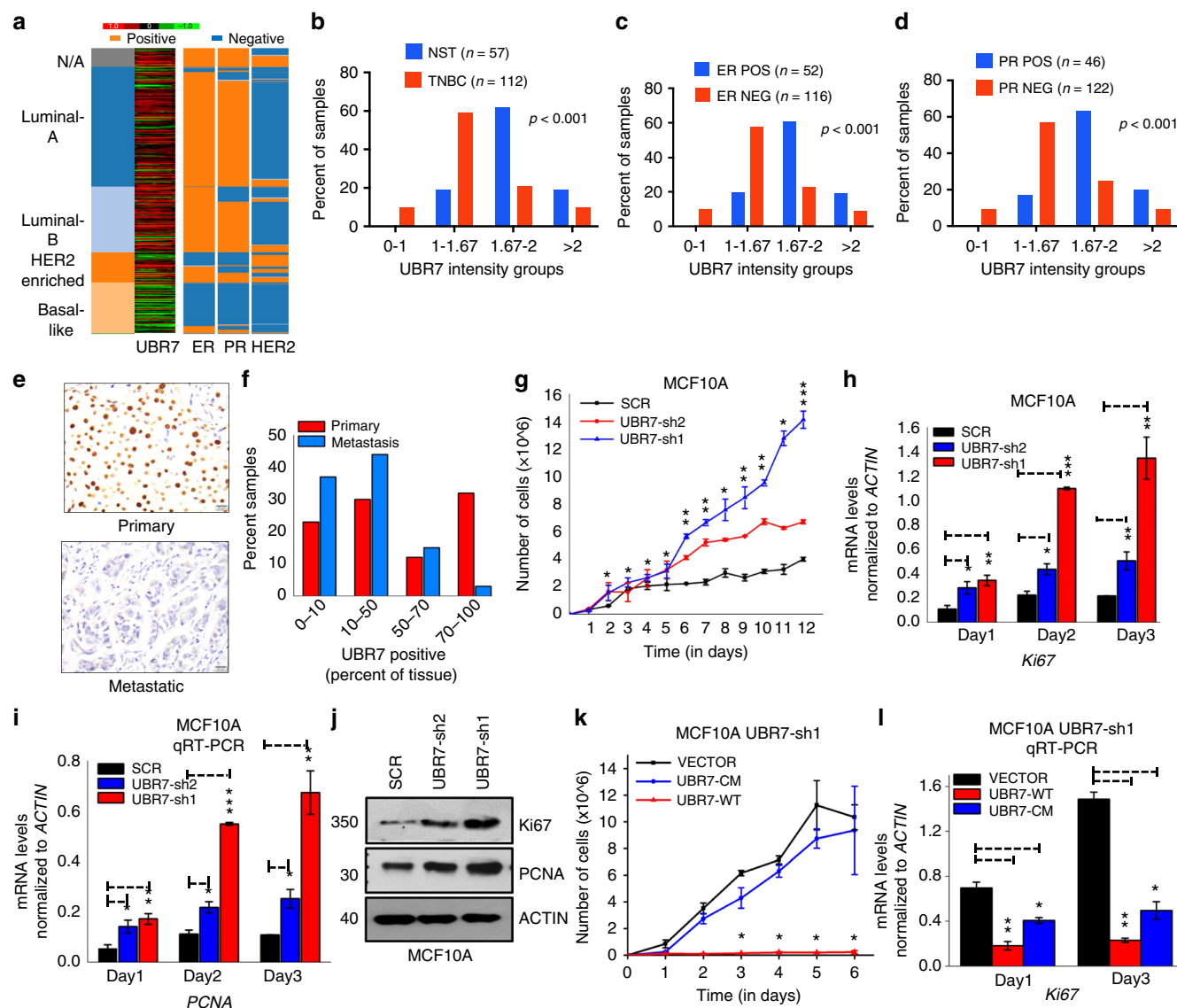
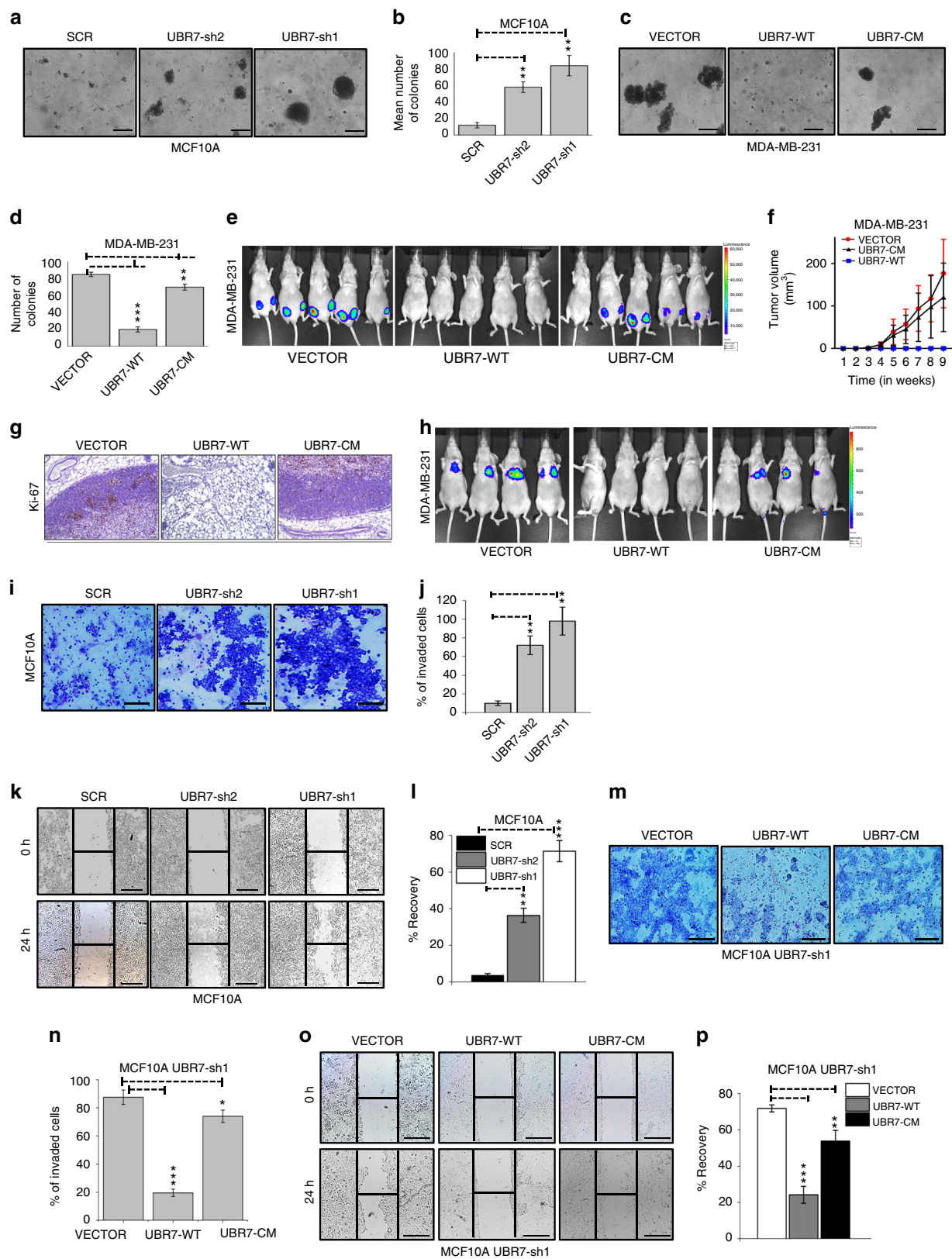


Fig. 3 UBR7 inhibits breast cancer cell proliferation. **a** Heatmap showing that UBR7 expression is downregulated in triple-negative or basal-type breast cancer (analysis of data from TCGA database). Green and red indicates down- and up-regulated, respectively. **b-d** Immunohistochemical analysis representing low UBR7 expression analyzed in triple-negative (**b**), estrogen receptor (ER)-negative (**c**), and progesterone receptor (PR)-negative (**d**) breast tumors. **e, f** Representative images of primary and metastatic invasive breast carcinoma showing UBR7 expression as analyzed immunohistochemically. Scale bar indicates 20 μ m. **g** Proliferation of MCF10A cells expressing scrambled (SCR) or UBR7 (UBR7-sh1/UBR7-sh2) short hairpin RNAs (shRNAs). **h, i** Quantitative real-time PCR (qRT-PCR) analysis of Ki-67 (**h**) and PCNA (**i**) from cultured MCF10A cells expressing SCR or UBR7 (UBR7-sh1/UBR7-sh2) shRNAs in a time-dependent manner. **j** Immunoblots of MCF10A cells expressing SCR or UBR7-shRNAs to monitor expression of Ki-67 and PCNA. ACTIN was used as a loading control. **k** Proliferation of cultured MCF10A UBR7-sh1 cells expressing a vector (VECTOR), WT (UBR7-WT), or catalytic-mutant (UBR7-CM). **l** qRT-PCR analysis of Ki-67 in MCF10A UBR7-sh1 cells expressing vector (VECTOR), wild-type (UBR7-WT), or CM (UBR7-CM) in a time-dependent manner. In all panels, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. $^{*}P < 0.05$; $^{**}P < 0.001$; $^{***}P < 0.0001$

UBR7 silencing activates Wnt/ β -catenin signaling. The Wnt/ β -catenin signaling pathway is deregulated in several cancers, including breast cancer³². It promotes tumor initiation, maintenance and metastasis³². We observed activation of the canonical Wnt/ β -catenin signaling cascade upon loss of UBR7 as evident from the upregulation of key positive regulators including WNT3A, FZD2/3, LRP5/6, ROR2, and DSH2, and down-regulation of negative regulators such as DKK1 (Fig. 7a, b and Supplementary Fig. 9j). Activation of the signaling pathway was further confirmed by nuclear localization of β -catenin upon loss of UBR7 (Fig. 7c). This could serve as a mechanism downstream

of CDH4 as other cadherins, such as CDH1, are known to regulate Wnt signaling^{33,34}. Indeed, restoration of CDH4 altered the nuclear localization of β -catenin to the cytoplasm and down-regulated known β -catenin target genes including AXIN2, CCND1, C-MYC, COX2, and MMP7 (Fig. 7d-f and Supplementary Fig. 9k), thereby maintaining the epithelial state of the cell. Cytoplasmic fraction of β -catenin is phosphorylated by glycogen synthase kinase 3 β (GSK3 β), which marks the protein for β -transducin repeat-containing protein (β -TrCP) mediated proteasomal degradation in the absence of an activating ligand for the signaling pathway. Thus, we examined the association of



β -catenin with this degradation complex in the absence of UBR7. We found a weak association of this degradation complex, which includes GSK3 β and β -TrCP with β -catenin in the absence of UBR7, indicating escape of β -catenin from the degradation

pathway, and entry into the nucleus, thereby activating its target genes (Fig. 7g). Overall, our results demonstrated that UBR7 loss activates the Wnt/ β -catenin signaling pathway, which is inhibited upon CDH4 restoration.

Fig. 4 UBR7 acts as tumor metastasis suppressor gene. **a–d** Soft agar assay with MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs (UBR7-sh1 or -sh2) (**a, b**) or with MDA-MB-231 cells expressing vector (VECTOR), wild-type (UBR7-WT), or catalytic-mutant (UBR7-CM) (**c, d**). **e, f** Tumor formation in mice injected into mammary fat pad with MDA-MB-231 cells expressing an empty vector (VECTOR) or WT (UBR7-WT) or CM (UBR7-CM). $n = 5$ mice per group. **g** Immunohistochemistry (IHC) for proliferation marker Ki-67 in tumors derived from control cells versus UBR7-WT- or UBR7-CM-overexpressing cells. Image was taken at $\times 20$ magnification. Scale bar indicates 50 μm . **h** Tumor metastasis in mice injected into tail vein with MDA-MB-231 cells expressing empty vector (VECTOR) or WT (UBR7-WT) or CM (UBR7-CM). $n = 4$ mice per group. **i, j** Invaded MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs in a Matrigel chamber were photographed and counted. **k, l** Wound healing/migration rate of MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs was monitored. **m, n** Invaded MCF10A UBR7-sh1 cells expressing WT (UBR7-WT) or CM (UBR7-CM) in a Matrigel chamber were photographed and counted. **o, p** Wound healing/migration rate of MCF10A UBR7-sh1 cells expressing WT (UBR7-WT) or CM (UBR7-CM) was monitored. In **a, c, i, k, m, o** scale bar indicates 10 μm . In **b, d, j, l, n**, and **p**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). In **f**, error bars indicate standard deviation (s.d.); $n = 5$ mice per group. P values were calculated using two-tailed t tests. * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

Discussion

PHD fingers are structurally conserved chromatin-binding modules present in proteins that are associated with chromatin and regulate gene transcription³⁵. The versatile function of a PHD finger as an epigenome reader promotes the recruitment of multi-protein complexes to change the chromatin structure, thereby augmenting transcription (initiation, elongation, and termination) or its repression³⁵. A PHD finger fold consists of a C-terminal α -helix and two strands of anti-parallel β -sheet, with a conserved Cys4-His-Cys3 motif in a cross-brace topology, which anchors two zinc atoms³⁶. PHD fingers are well known for reading methylated H3 lysine 4 or unmethylated N-term tail of H3, with a few exceptions. The RING finger, which is well characterized for its E3 ubiquitin ligase activity, and has a Cys3-His-Cys4 motif organization, is reciprocal to the PHD finger³⁷. The UBR7-PHD finger, on the other hand, has a unique motif organization (Cys4-His2-Cys2) different from that of a canonical PHD or RING finger and hence can be considered as an “atypical” PHD finger. We report here an E3 ubiquitin ligase activity by a PHD finger of UBR7 toward histone H2B at lysine 120. Of note, in comparison with other E3 ubiquitin ligases, which act in a complex²⁵, UBR7 can promote enzymatic catalysis in isolation. Indeed, mutating the two His residues, which is instrumental to anchoring the zinc coordination complex, abrogates the catalytic activity. Structural studies will provide more insight into the molecular mechanism of H2BK120Ub catalysis by this atypical domain. Also, a detailed investigation of its reader function could provide more insight into the role of the UBR7-PHD finger in cellular context.

UBR7 loss not only drastically reduced H2BK120Ub but also significantly reduced H3K79Me2 without much effect on H3K4Me3, H3K27Ac, H3K27Me3, and H3K9Me3. This was intriguing because several studies have suggested that H2BK120Ub marked nucleosomes act as templates for DOT1L and SET/COMPASS complexes for making H3K79Me2 and H3K4Me3, respectively^{38–40}. Interestingly, in the chromatin state analyses, cells with UBR7 loss also harbored changes in transcription states that had predominance of H3K79Me2. Since H3K79Me2 mark is linked with transcriptional elongation and observed on genes that are being actively transcribed, we propose that UBR7 may play important roles in transcriptional elongation. Further biochemical studies focusing on isolation of UBR7 protein complexes or identification of interacting partners will be needed to determine its exact function in this process, if any.

Importantly, we found UBR7 loss to be highly correlated with triple-negative and basal-like breast cancer. Although the molecular features that define triple-negative breast cancer (loss of ER, PR, and Her2 expression) are clear, ambiguity exists in the definition of basal-like cancers. UBR7 loss may be a key determining

feature of this aggressive subtype of breast cancer. Notably, we established that UBR7 can suppress breast tumor formation and metastasis in vivo. One possibility of such an event is through UBR7's prevention of a “self-seeding” event, in which local invading cancer cells lead to tumor formation via fusion of propagating colonies⁴¹. Also, it is difficult to determine the relative contribution of proliferative versus invasive role for UBR7 in its pro-metastatic function. Overall, UBR7 loss may be a predictive biomarker and provide specific vulnerabilities to epigenetic inhibitors given its drastic impact on chromatin states. Furthermore, given the substantial impact of UBR7 on maintenance of the epithelial state and inhibition of the plasticity of a cell, determining its role in suppression of other malignancies requires further investigation.

We demonstrated that CDH4/R-cadherin is a major target downstream of UBR7. Cadherins are crucial in the maintenance of cell boundary, tissue morphogenesis, and cell polarity⁴², and aberrant function may lead to severe metastatic neoplasia³³. Whereas UBR7 loss promotes breast tumor metastasis, CDH4 overexpression provides only a partial rescue of such phenotypes. CDH4 was found to be highly expressed in mammary epithelial cells, but severely down-regulated in invasive ductal carcinoma³¹. Moreover, due to heterogeneity of cancer cells, CDH4 was absent in those cells that were poorly differentiated³¹. Although we showed that UBR7 prevents metastatic colonization and represses genes that are crucial for breast cancer metastasis by targeting CDH4, other downstream pathways and target genes may also have important roles. For example, other proteins, such as integrins and cytokines, have already demonstrated roles in breast cancer etiology and metastasis⁴³.

Our results indicate that CDH4 may control β -catenin signaling in a manner similar to E-cadherin-mediated control of β -catenin signaling, which is well documented to play important roles in metastasis. Loss of E-cadherin leads to β -catenin release from the cell surface, resulting in escape from the β -TrCP-mediated degradation pathway, and thereby promoting its nuclear localization and activation of target genes^{32,34,44}. In UBR7-depleted cells, Wnt/ β -catenin signaling was one of the top misregulated pathway, and overexpression of CDH4 downregulates Wnt signaling in UBR7-depleted cells by reinstating β -catenin to the cell membrane and reducing the nuclear pool. In addition, we observed that several other Wnt signaling regulators (such as WNT3A, FZD2/3, LRP5/6, ROR2, DSH2, and DDK1) were transcriptionally controlled by UBR7, strongly suggesting that UBR7 is a key mediator of the Wnt/ β -catenin signaling cascade. Collectively, our results demonstrated that UBR7 is a H2B E3 ubiquitin ligase that suppresses triple-negative subtype of breast cancer by activating CDH4/R-cadherin expression and inhibiting the canonical Wnt/ β -catenin signaling pathway.

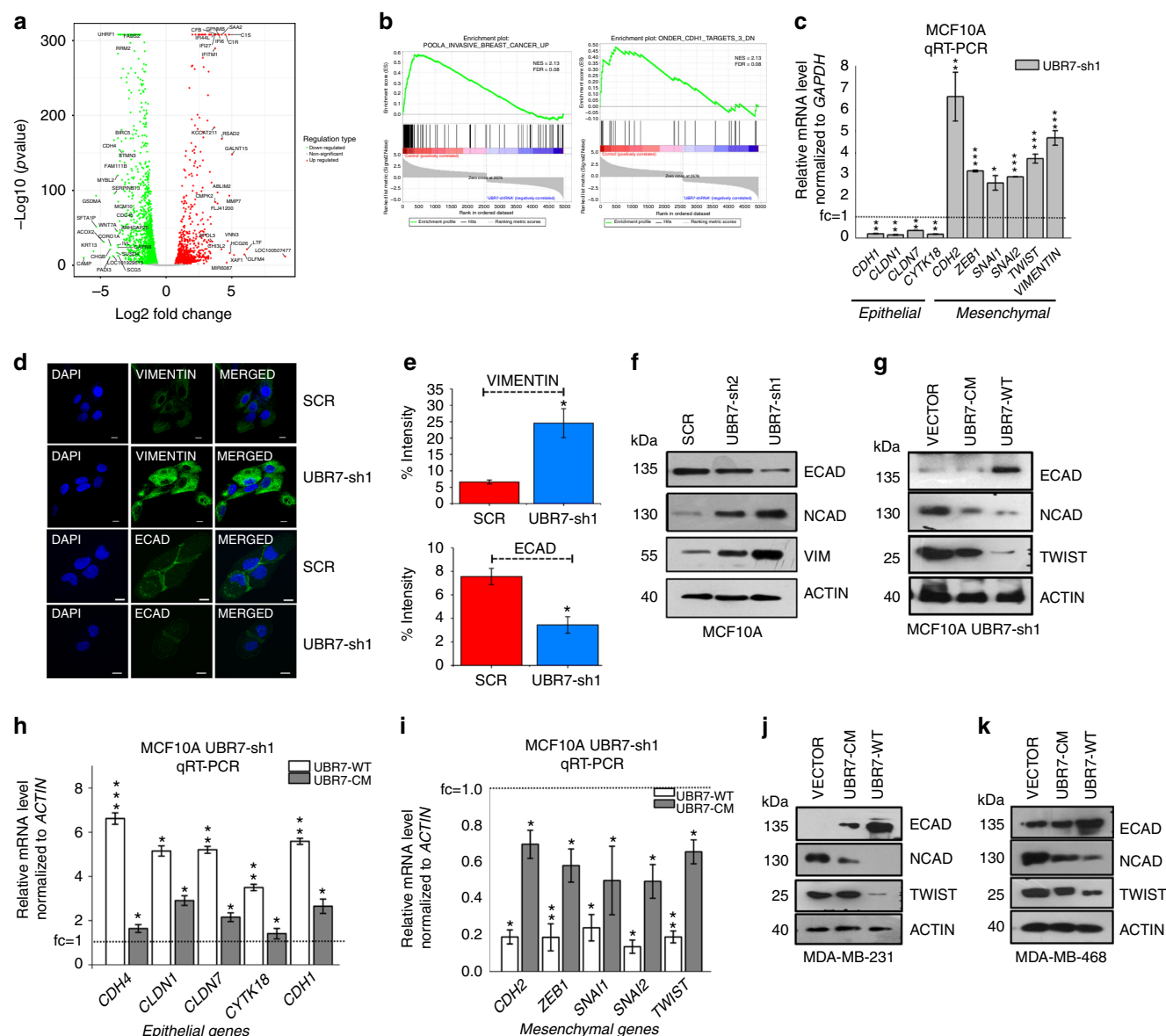


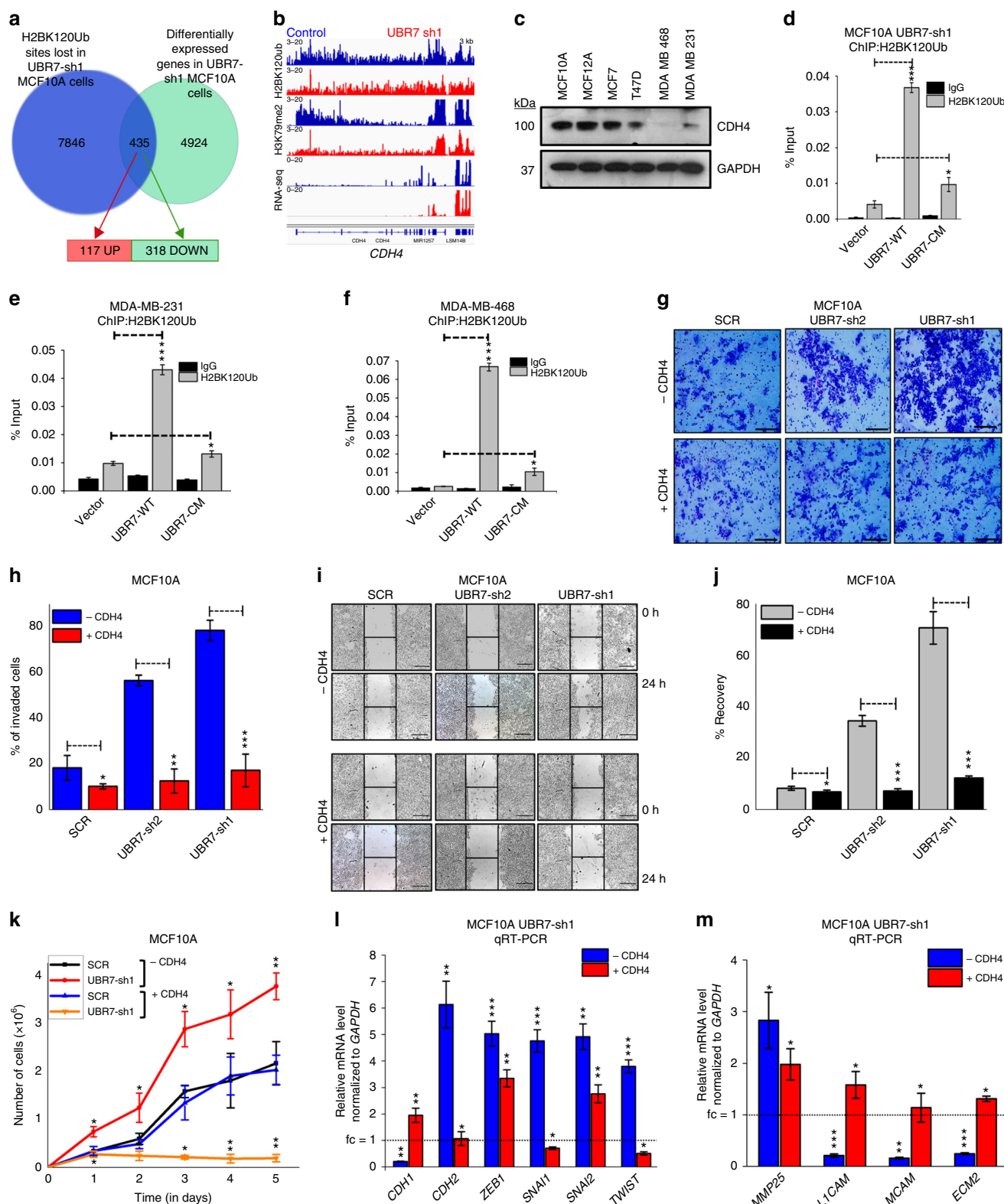
Fig. 5 UBR7 is an epithelial-to-mesenchymal transition (EMT) suppressor. **a** Volcano plots of the differentially expressed genes upon knockdown of UBR7 (using UBR7-sh1) in MCF10A cells scored via RNA-sequencing (RNA-seq) analysis. Green, red and grey indicates down-, up-regulated and non-significant, respectively. **b** Gene Set Enrichment Analysis (GSEA) output images of two chosen pathways displaying a correlation of differentially regulated genes in UBR7-knockdown MCF10A cells with the “poola_invasive_breast_cancer_upregulated” set and “Onder_CDHI_targets_downregulated” set. **c** Quantitative real-time PCR (qRT-PCR) analysis of EMT signature gene expression upon UBR7 knockdown (UBR7-sh1) in MCF10A cells. **d**, **e** Immunofluorescence studies showing expression of VIMENTIN and ECAD after UBR7 knockdown (UBR7-sh1) in MCF10A cells (**d**). Scale bar indicates 100 μ m. Percent intensity is quantified (**e**). **f**, **g** Immunoblots of MCF10A cells expressing scrambled (SCR) or UBR7 short hairpin RNAs (shRNAs) (UBR7-sh1 and sh2) (**f**) or MCF10A UBR7-sh1 expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM) (**g**), monitoring expression of candidates regulating EMT. ACTIN was used as a loading control. **h**, **i** qRT-PCR analysis of epithelial (**h**) and mesenchymal (**i**) genes in MCF10A UBR7-sh1 expressing WT (UBR7-WT) and CM (UBR7-CM). **j**, **k** Immunoblots of MDA-MB-231 (**j**) and MDA-MB-468 (**k**) cells expressing WT (UBR7-WT) and CM (UBR7-CM) monitoring expression of candidates regulating EMT. ACTIN was used as a loading control. In **c**, **d**, **e**, **h**, **i**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

Methods

Cell lines and cell culture. HEK293T cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco) and 1% antibiotic-antimycotic (Gibco) at 37 °C and 5% CO₂. MCF10A and MCF12A cells were maintained in DMEM/Ham’s F12 supplemented with 5% horse serum (Gibco), epidermal growth factor (EGF), insulin, hydrocortisone, cholera toxin (Sigma), and 1% antibiotic-antimycotic. MDA-MB-231, MDA-MB-468, MCF7, and T47D cells were maintained in RPMI-1640 medium (Gibco) supplemented with 10% FBS, insulin, and 1% antibiotic-antimycotic. All cell lines were purchased from ATCC. 16N, 21PT, and 21MT2 (provided by R. Weinberg, Whitehead Institute, Massachusetts Institute of Technology) were maintained in DMEM supplemented with 10% FBS, insulin, hydrocortisone, EGF,

and 1% antibiotic-antimycotic. 4T1 and 4T07 (provided by R. Weinberg) were maintained in DMEM/Ham’s F12 medium supplemented with 10% FBS, insulin, hydrocortisone, and 1% antibiotic-antimycotic. All cell lines used in the study were negative for mycoplasma. All cell lines were validated by MD Anderson Cancer Center Characterized Cell line core facility via DNA fingerprinting. For transient transfection, cells were counted and seeded in 12-well or 6-well or 6-cm dishes and then subjected to overexpression using Lipofectamine-2000 (Invitrogen) as per the manufacturer’s protocol.

Protein purification. The full-length UBR7 complementary DNA (cDNA) sequence, UBR domain alone, or PHD finger alone was cloned in a pDEST15



vector (GATEWAY cloning system, Invitrogen) and sequence verified. The protein used in *in vitro* assays were purified as described previously⁴⁵. Briefly transformed cells were grown till 0.8 optical density (OD) and induced with 1 mM isopropyl β -D-1-thiogalactopyranoside (Sigma) at 20 °C for 16 h. The pellets were resuspended in lysis buffer and lysed mildly followed by glutathione sepharose beads binding and washing with wash buffer. The proteins were eluted and purified further to homogeneity via gel filtration chromatography using a Superdex75 column (GE Healthcare). CM derivative (H163S/H166S) were generated using a QuikChange site-directed mutagenesis kit (Stratagene) as per the standard protocol⁴⁶.

Nucleosome isolation. Nucleosomes were prepared freshly from HeLa cells as described previously⁴⁵. Briefly, the nuclear pellet from HeLa cells was digested with MNase (0.2 Units/ μ l; Sigma) and extracted with TE buffer for 1 h. The mononucleosome was separated via sucrose gradient (5–40%) ultracentrifugation using Sorvall WXUltra100 (Thermo Fischer Scientific) with AH650 rotor for 16 h at 207,203 \times g. For further analysis, these fractions were pooled and concentrated.

In vitro ubiquitination assay. *In vitro* ubiquitination reactions were set up with purified UBR7-PHD, UBR7-WT, or UBR7-CM (H163S/H166S) as E3 enzymes and

Fig. 6 UBR7 suppresses epithelial-to-mesenchymal transition (EMT) through activation of CDH4. **a** Venn diagram showing overlaps of H2BK120Ub enriched and differentially regulated genes after UBR7 knockdown. **b** Integrative Genomics Viewer (IGV) view of H2BK120Ub and H3K79me2 chromatin immunoprecipitation sequencing (ChIP-seq) and RNA-sequencing (RNA-seq) tracks on the *CDH4* gene in Control (SCR) or UBR7-sh1 short hairpin RNA (shRNA) expressing MCF10A cells. **c** Immunoblot showing the expression of CDH4 across different normal breast cell and breast cancer cell lines. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a loading control. **d–f** Bar plots showing H2BK120Ub ChIP in the *CDH4* gene locus in MCF10A UBR7-sh1 (**d**) or MDA-MB-231 (**e**) or MDA-MB-468 (**f**) cells expressing wild-type (UBR7-WT) and catalytic-mutant (UBR7-CM). **g, h** Invasion of MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1 and UBR7-sh2) of UBR7 in a Matrigel chamber was photographed and quantitated. **i, j** Wound healing by MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1 and UBR7-sh2) of UBR7 was photographed and the percent recovery was measured over time. **k** Proliferation of cultured MCF10A cells overexpressing CDH4 in the presence (SCR) and absence (UBR7-sh1) of UBR7. **l, m** Quantitative real-time PCR (qRT-PCR) analysis of EMT signature genes (**l**) and cell adhesion-linked genes (**m**) in UBR7-sh1 MCF10A cells upon CDH4 overexpression. In **g, i**, scale bar indicates 10 μ m. In **d–f, h, j–m**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. * $P < 0.05$; ** $P < 0.001$; *** $P < 0.0001$

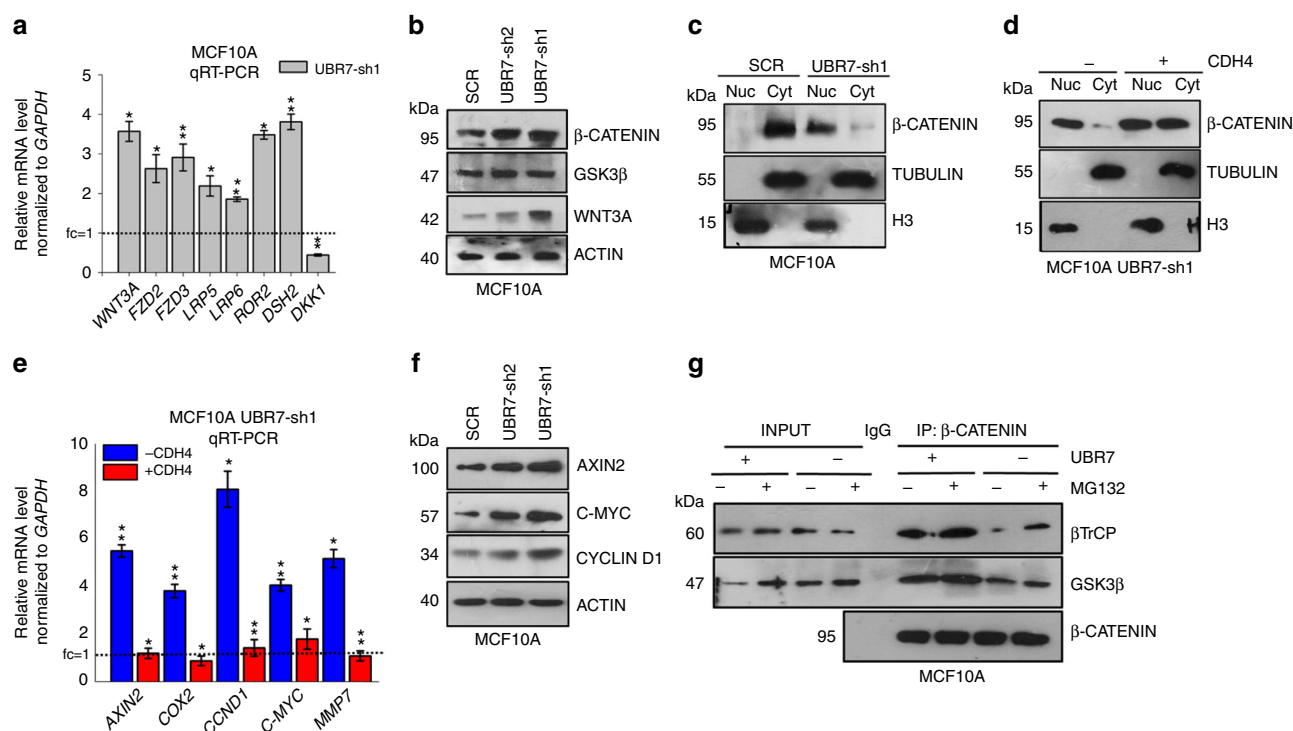


Fig. 7 UBR7 silencing activates the Wnt/ β -catenin signaling pathway. **a** Quantitative real-time PCR (qRT-PCR) analysis of key regulators of the Wnt/ β -catenin signaling pathway upon loss of UBR7 (UBR7-sh1). **b** Immunoblots monitoring expression of β -CATENIN, glycogen synthase kinase 3 β (GSK3 β), WNT3A, and ACTIN (loading control) in MCF10A cells expressing scrambled (SCR) or UBR7-shRNAs. **c** Immunoblots monitoring expression of β -CATENIN in the nuclear and cytoplasmic fractions of MCF10A cells expressing scrambled (SCR) or UBR7-sh1shRNA. Histone H3 and TUBULIN were used as loading controls for nuclear and cytoplasmic lysates, respectively. **d** Immunoblots monitoring expression of β -CATENIN in the nuclear and cytoplasmic fractions of CDH4-expressing MCF10A UBR7-sh1 cells. Histone H3 and TUBULIN were used as loading controls for nuclear and cytoplasmic lysates, respectively. **e** qRT-PCR analysis of β -catenin target genes expression in UBR7-sh1 MCF10A cells upon CDH4 overexpression. **f** Immunoblots showing expression of β -catenin target genes upon UBR7 knockdown (UBR7-sh1 or -sh2) in MCF10A cells. **g** Co-immunoprecipitation of β -catenin from UBR7-knockdown MCF10A cells in the presence or absence of MG132 (20 μ M for 18 h) showing association of β -catenin with GSK3 β and β -transducin repeat-containing protein (β -TrCP). In **a, e**, error bars indicate standard deviation (s.d.); $n = 3$ technical replicates of a representative experiment (out of three experiments). P values were calculated using two-tailed t tests. * $P < 0.05$; ** $P < 0.001$

recombinant H2B, H2A/H2B dimer, core histones octamer, or purified nucleosome as a substrate using Ubiquitinylation kit (Cat #: BML-UW9920, Enzo Life Sciences) as per the manufacturer's protocol. Briefly, the reaction was carried out in ubiquitinylation buffer containing 100 U/ml inorganic pyrophosphatase, 1 mM dithiothreitol (DTT), and 5 mM EDTA (negative control). 2.5 μ M of ubiquitin (biotinylated) was incubated with 100 nM E1, 2.5 μ M E2 (for UBR7, UbcH6 acts as the E2), 5 mM Mg-ATP, and 100 nM E3 along with 1 μ M substrate at 37 $^{\circ}$ C for 1 h. The reaction was stopped, and trichloroacetic acid precipitation was performed and analyzed via western blotting with antibodies against H2BK120Ub, H2B, H2A, H3, and H4 antibodies.

Peptide pull-down assay. A peptide pull-down assay was performed as described previously⁴⁵. Briefly, equivalent amounts of peptides and protein were incubated in immunoprecipitation (IP) buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 0.05%

NP-40, 1 mM DTT). The complex was pulled down with streptavidin beads, washed with the IP buffer, and eluted and analyzed using western blotting.

GST pull-down assay. GST and GST-fusion proteins were incubated with the recombinant histones H3, H4, H2A, and H2B at equimolar ratios in a IP buffer overnight as described previously⁴⁵. The complex was pulled down with glutathione sepharose bead (GE Healthcare), washed with IP buffer, and analyzed using western blotting with specific antibodies. Ten percent of the histone proteins (H3, H4, H2A, and H2B) were used as inputs.

Co-immunoprecipitation. Cells were subjected to co-immunoprecipitation as delineated previously⁴⁵. In brief, cells were lysed in lysis buffer (20 mM Tris pH 8.0,

150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), 1 mM EDTA) and pulled down with specific antibodies or FLAG M2 beads, followed by washes with the lysis buffer. The immunoprecipitant was analyzed using western blotting. Ten percent of the lysate with which IP was set was used as the input.

RNA interference through lentiviral production. shRNA plasmids for UBR7 with pLKO.1-puro backbone (Sigma-Aldrich) were screened for efficient knockdown. Two of seven shRNAs were selected for subsequent experiments. Their sequences are as follows: UBR7-sh1 5'-CAGTGCACCCAGGGTTATTG-3' and UBR7-sh2-5'-GCTTAAAGCTAAGCAGCTTAT-3'. UBR7-sh1 targets the 3'-UTR of the gene, so the overexpression constructs were resistant to the shRNAs. HEK293T cells were plated at a density of 3×10^5 cells in 10-cm dishes. Eight micrograms of shRNA and packaging vectors were transfected into the cells as described previously⁴⁷. Transduced cells were selected using puromycin (10 µg/ml; Sigma) for 3 days.

UBR7 overexpression via lentiviral production. WT and H163S/H166S CM of UBR7 were cloned into the pHAGE-CMV-fulleF1a-IRES-ZsGreen (from Jeng-Shin Lee; Dana-Farber/Harvard Cancer Center). 293T cells were plated at density of 3×10^5 in 10-cm dishes. Recombinant lentiviral particles were produced using Lipofectamine 2000-mediated transient transfection in HEK293T cells. Briefly, 8 µg of overexpression vectors, a packaging vector (psPAX2), and an envelope vector (pMD2.G) were transfected into 293T cells plated in 10-cm dishes. The viral supernatant was harvested 48 and 72 h after transfection and filtered. Cells were infected three times in 48 h with the viral supernatant containing 10 µg/ml Polybrene. Green fluorescent protein-positive cells were sorted and cultured for other experiments.

Quantitative real-time PCR. Total RNA was isolated using TRIzol reagent (Invitrogen) and reverse transcribed using a Revertaid First Strand cDNA Synthesis kit (Thermo Fischer Scientific) according to the manufacturer's protocol followed by qRT-PCR using ABI-SYBR GREEN mix (Applied Biosystems). qRT-PCR was performed using StepONE plus FAST Real-time PCR machine. Each sample was analyzed independently three times and the results of one representative experiment, with technical triplicates, are shown. List of primers is provided in Supplementary Table 1.

Western blot analysis. Whole-cell lysates were prepared in lysis buffer (20 mM Tris, pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 1 mM EDTA) as described previously⁴⁵. The lysates were electrophoresed on 7.5%, 11%, or 15% SDS-polyacrylamide gel electrophoresis gels. Blots were probed with specific antibodies. The membrane was blocked with 5% bovine serum albumin (BSA) or non-fat dry milk in TBST (Tris-buffered saline, 0.1% Tween-20). The antibodies used are listed in Supplementary Table 2. All uncropped images of the blots are provided in Supplementary Fig. 10.

Immunofluorescence. Immunofluorescence was carried out as described previously⁴⁵. Cells were fixed with 4% paraformaldehyde or methanol, permeabilized with 1% Triton X-100 or 100% methanol, blocked with 3% BSA, and stained with the indicated antibodies. Coverslips were mounted after staining with DAPI (4',6-diamidino-2-phenylindole) and photographed using a Nikon T1E confocal microscope with an AIRMP Scanner Head.

Migration or wound healing assay. Migration assay was performed as described previously⁴⁸. Cells were plated in 6-well dish in triplicates and grew to 90% confluence and the cell surface was scratched with a sterile 20 µl tip. Images were captured with Nikon T1 E100 microscope at 0 h and post 24 h after the scratch to measure the wound healing by the cells, indicating their migratory capacity and migration rate. The images were analyzed using the Image J software program (National Institutes of Health; NIH), to measure the wound recovery by the cells indicating their migratory ability.

Proliferation assay. Transduced and transfected cell lines were plated at a density of 1×10^6 cells in triplicates in a 24-well plate. MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide) was added at indicated time points and MTT assay was performed as described previously⁴⁹. The OD was measured using an Epsom Plate reader. The number of proliferating cells was calculated from the standard curve.

Invasion assay. Invasion assay was performed as described previously⁴⁷. For this assay, 1×10^6 cells were cultured in the upper well of Matrigel chambers (Corning) containing serum-free medium and allowed to migrate towards serum-enriched medium in the bottom wells. After 20 h of incubation, invading or migrating cells were stained with 0.5% Giemsa, photographed with Nikon T1 E100 microscope, and counted using the Image J software program.

Soft agar assay. Transduced human breast epithelial and cancer cell lines were resuspended in 3 ml of soft agar (medium containing 0.3% noble agar [Affymetrix]) warmed to 45 °C. The cell suspension was layered onto 3 ml of bottom agar (medium containing 0.8% noble agar) in a 6-well plate (six replicates). 2 ml of the medium was added to the top agar and changed every 3 days. Visible colonies were scored after 4–5 weeks and stained with 0.5% Giemsa. The number of colonies and mean area of colonies was calculated using the Image J software program.

Animal experiments and in vivo imaging. Animals used in these experiments were all female nude mice aged 8–10 weeks (The Jackson Laboratory). All cells used for in vivo injections were labeled with red fluorescent protein luciferase. The mice were anesthetized with isoflurane and injected with 2×10^6 MDA-MB-231 cells in 50 µl of phosphate-buffered saline (PBS) in the abdominal mammary fat pads on both sides ($n = 5$ mice). Tumor growth was monitored weekly via caliper measurement and bioluminescent imaging once every 2 weeks. Once the largest tumor diameter was reached (1.5 cm, which is the maximal tumor diameter allowed under our institutional protocol), the animals were sacrificed. For the tail vein injections, the mice ($n = 4$) were injected with 100 µl of 1×10^6 cells were injected into the medial tail vein. All mouse experiments were performed with the approval of the MD Anderson Institutional Animal Care and Use Committee.

For in vivo imaging, cells were infected with EF1-RFP-T2A-Luciferase (System Biosciences) to enable stable expression of firefly luciferase. All in vivo bioluminescent imaging was carried out at the MD Anderson Small Animal Imaging Facility. For this imaging, the animals were anesthetized with isoflurane. They were injected intraperitoneally with 3 mg of D-luciferin (Perkin Elmer) and imaged using the IVIS Spectrum Imaging System (Perkin Elmer). Analysis after acquisition was done using the Living Image software program (version 4.3; Perkin Elmer).

Tissue microarrays. Patients: Breast tumors and normal breast tissue were obtained from patients who underwent surgery at MD Anderson Cancer Center from 2005 to 2015. Information on their hormone and Her2 statuses, as well as proliferation fraction (Ki-67), was retrieved from pathological and clinical reports, ER information was available for 367 (99%) patients, PR information was available for 368 (99%) patients, Her2 information was available for 362 (98%) patients, and proliferation fraction (Ki-67) information was available for 208 (56%) patients.

Breast cancer TMAs from MD Anderson Cancer Center and a commercially available TMA BR2082a (US Biomax, Inc. Rockville, MD, USA) were used. The TMAs from MDACC included 538 breast cancers and 15 normal tissues. Breast cancers were represented in 389 (60%) cases with 3 punches of 1 mm, in 96 (17.8%) cases with 3 punches of 0.6 mm, and in 53 cases (9.9%) with 3 punches of 1 mm and 3 punches of 0.6 mm. The normal tissue samples included six punches of 0.6 mm from three normal lymph nodes, three normal breast tissues, three normal kidneys, three normal colon, and three normal lung samples. The composition of the commercial TMA is shown at the vendor's website (<http://www.biomax.us/tissue-arrays/Breast/BR2082a>).

Immunohistochemistry. For validation of the anti-UBR7 antibody, Western blot of UBR7 expression on the whole cell line lysate of MCF10A transfected with a control small interfering RNA (siRNA) and siRNA against UBR7, and of MCF7 transfected with siRNA against UBR7 was performed. These cell lines were paraffin embedded and the final protocol was established to perform staining on 5-µm-thick TMA sections. In brief, epitope retrieval was performed with citrate buffer at pH 6.0 for 20 min, followed by peroxidase blocking for 5 min. A polyclonal rabbit anti-UBR7 antibody (Bethyl Laboratories) was then incubated for 60 min using a 1:2000 dilution followed by polymer (goat-anti-rabbit immunoglobulin G (IgG)) and 3,3'-diaminobenzidine incubation for 8 and 10 min, respectively. Slides were counterstained with hematoxylin. The staining was performed using supplies and an autostainer from Leica Biosystems.

For evaluation of UBR7 staining, nuclear staining was scored semi-quantitative providing the percentage of stained cells and staining intensity (0 = no staining, 1+ = weak staining, 2+ = moderate staining, and 3+ = strong staining). Representative staining examples are shown in Supplementary Fig. 4a. The study was approved by the MD Anderson Institutional Review Board.

ChIP assay. ChIP assays were performed as described earlier⁵⁰. Cells were cross-linked with 1% formaldehyde and the chromatin was sheared and immunoprecipitated with the UBR7 antibody (Bethyl Laboratories), H2BK120Ub antibody (Millipore), H2B antibody (Abcam), or as a negative control IgG. ChIP DNA was analyzed by qPCR using gene specific primers. Each ChIP experiments were performed three independent times with technical triplicates.

ChIP-seq assay. ChIP assays were performed as described previously⁵⁰ with minor modifications. Briefly, $\sim 2 \times 10^7$ cells were harvested via cross-linking with 1% (wt/vol) formaldehyde for 10 min at 37 °C with shaking. After quenching with 150 mM glycine for 10 min at 37 °C with shaking, cells were washed twice with ice-cold PBS and frozen at –80 °C for further processing. Cross-linked pellets were thawed and lysed on ice for 30 min in ChIP harvest buffer (12 mM Tris-Cl, 1 × PBS, 6 mM EDTA, 0.5% SDS) with protease inhibitors (Sigma). Lysed cells were sonicated with

a Bioruptor (Diagenode) to obtain chromatin fragments (~200–500 bp) and centrifuged at 15,000 × g for 15 min to obtain a soluble chromatin fraction. In parallel with cellular lysis and sonication, antibodies (5 µg/3 × 10⁶ cells) were coupled with 30 µl of magnetic protein G beads in binding/blocking buffer (PBS + 0.1% Tween + 0.2% BSA) for 2 h at 4 °C with rotation. Soluble chromatin was diluted five times using ChIP dilution buffer (10 mM Tris-Cl, 140 mM NaCl, 0.1% dissolved organic compound, 1% Triton X, 1 mM EDTA) with protease inhibitors and added to the antibody-coupled beads with rotation at 4 °C overnight. After washing, samples were treated with elution buffer (10 mM Tris-Cl, pH 8.0, 5 mM EDTA, 300 mM NaCl, 0.5% SDS), RNase A, and Proteinase K, and cross-links were reversed overnight. ChIP DNA was purified using AMPure XP beads (Agencourt) and quantified using the Qubit 2000 (Invitrogen) and Bioanalyzer 1000 (Agilent). Libraries for Illumina sequencing were generated following the New England BioLabs (NEB) Next Ultra DNA Library Prep Kit protocol. A total of 10 cycles were used during PCR amplification for the generation of all ChIP-seq libraries. Amplified ChIP DNA was purified using double-sided AMPure XP to retain fragments (~200–500 bp) and quantified using the Qubit 2000 and Bioanalyzer 1000 before multiplexing.

ChIP-seq data processing. Raw fastq reads for all ChIP-seq experiments were processed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), and quality reads were aligned to the hg19 reference genome using Bowtie version 1.1.2⁵¹ with the following criteria: -n 1 -m 1 -best-strata. Duplicate reads were marked using SAMBLASTER⁵² before compression to BAM files. To directly compare Control and UBR7-shRNA ChIP-seq samples, uniquely mapped reads for each mark were normalized by total reads per condition, sorted, and indexed using samtools version 0.1.19⁵³.

Model-based analysis of ChIP-seq (MACS) (version 1.4.2; peak calling algorithm with a *p* value threshold of 1e-7)⁵⁴ was used to identify H2BK120Ub enrichment over “input” background. Unique H2BK120Ub binding sites were identified using the concatenate, cluster, and subtract tools from the Galaxy/Cistrome web-based platform⁵⁵. Briefly, a shared peak set was first generated by clustering intervals of H2BK120-Control peaks that directly overlapped H2BK120-UBR7-shRNA peaks by a minimum of 1 bp. Unique peaks were then identified by subtracting the total number of H2BK120 peaks in each condition by the shared peak set. Venn diagrams were generated using the Venn Diagram tool in Galaxy. To visualize ChIP-seq libraries on the IGV browser, we used deepTools version 2.4.060 to generate bigWig files by scaling the bam files to reads per kilobase per million (RPKM) using the following criteria: bamCoverage -b-normalizeUsing RPKM-smoothLength 300-binSize 30-extendReads 200 -o.

A list of known genes was obtained from the UCSC Genome browser (<http://genome.ucsc.edu/>). Proximal promoters were defined as ±5 kb from the transcription start site (TSS) and the genebody was defined as all genic regions outside of the +5 kb promoter region. Intergenic regions were defined as all regions outside both the proximal promoter and genebody. H2BK120 peaks were assigned to genes if they overlapped the promoter or genebody by a minimum of 1 bp. These H2BK120Ub “enriched” regions were further used for the generation of read density plots for all ChIP-seq data. All read density plots were generated using the *thengs.plot* package in R⁵⁶.

ChIP-seq analysis for UBR7, RNF20, and RNF40 comparison. To directly compare Control SCR, UBR7, RNF20, and RNF40 shRNA ChIP-seq samples, uniquely mapped reads for H2BK120Ub in all conditions were normalized to ~10 million reads. For generation of Supplementary Fig. 3b–d, Control replicate1 was used for the UBR7-shRNA-1 comparison and Control replicate2 was used for the RNF20/RNF40 shRNA comparisons. Normalization of Control replicate1 and UBR7-shRNA-1 samples to ~10 million reads displayed little effect on the average density profile of H2BK120Ub (Supplementary Fig. 3b). For further UBR7, RNF20, and RNF40 comparisons, read counts for Control (SCR) H2BK120ub and Input replicates were merged together, normalized to ~10 million reads, and peaks were called using MACS (*p* value 1e-7). To identify unique H2BK120 binding sites that were lost upon knockdown of either UBR7, RNF20, or RNF40, a shared peak set was first generated by clustering intervals of Control (SCR) H2BK120Ub sites that directly overlapped either UBR7, RNF20, or RNF40 shRNA H2BK120Ub peaks by a minimum of 1 bp. Unique peaks were then identified by subtracting the total number of H2BK120Ub peaks in each condition by their associated shared peak set. A final Control (SCR) shared peak set was further generated using the Control (SCR) H2BK120Ub binding sites from Supplementary Fig. 3e–g, which were lost from depletion of each factor (UBR7, RNF20, or RNF40). A final Control (SCR) unique peak set was identified by subtracting the total number of H2BK120Ub peaks in each condition by the shared peak set. H2BK120Ub peaks were assigned to genes if they overlapped the promoter (±5kbTSS) or genebody by a minimum of 1 bp and these H2BK120Ub “enriched” regions were used for the generation of average density profiles (Supplementary Fig. 3b–d, 3h) and pathway analysis (Supplementary Fig. 3i).

Chromatin state calls. ChromHMM⁵⁷ was used to identify combinatorial chromatin state patterns based on the histone modifications studied. Normalized bam files were converted into binarized data at a 1000 bp resolution using the

BinarizeBam command with a *p* value cut-off of 1e-5. We specified that ChromHMM should learn a model based on 10 chromatin states. As we considered models between 8 and 20 chromatin states, we chose a 10-state model because it is large enough to identify important functional elements while still being small enough to interpret easily. Overlap enrichment was used to compute differential enrichment in each of the 10 chromatin states between Control and UBR7-shRNA samples. The ChromHMM segment files from the 10-state model contain the genomic locations of each chromatin state called in both the Control and UBR7-shRNA samples. To determine which chromatin states were enriched between conditions, we further compared the genomic locations by using the Control segments file as input for the segment directory, and by further separating the UBR7-shRNA segments file into 10 individual states and using it as input for the external coordinate directory. The UBR7-shRNA segment file was separated into individual chromatin states for the external coordinate directory with the following command:

```
awk -F/t '{print >> $4;close($4)}' ~/path_to/UBR7-sh1_segments.bed
Overlap enrichment was ran using the following command:
java -mx4000M -jar ChromHMM.jar OverlapEnrichment ~/path_to/Control_segments.bed ~/path_to/UBR7-shRNA_segments_separated OverlapEnrichment_Control_vs_UBR7
```

RNA-sequencing. RNA was isolated using RNeasy kit and libraries prepared using Illumina mRNA-Seq library kit. Raw FASTQ reads for all RNA-seq experiments were processed using FastQC and quality reads were aligned with the hg19 reference genome using TopHat(version 2.0.14)⁵⁸ with a Bowtie2 (version 2.2.3)⁵⁹ index based on UCSC annotations using the following criteria: -G -g 1 -r 150-mate-std-dev 50-library-type fr-unstranded. These criteria preserved only the best reads that uniquely mapped to the genome with one or fewer mismatches. To visualize RNA-seq libraries on the IGV browser, we used deepTools version 2.4.060 to generate bigWig files by scaling the bam files to RPKM using the following criteria: bamCoverage -b-normalizeUsing RPKM-smoothLength 300-binSize 30 -o.

For identification of differentially expressed genes and gene set enrichment analysis (GSEA), raw counts were obtained by assigning reads at the gene level across the UCSC hg19 reference genome using featureCount⁶⁰ in the Rsubread package. DESeq2⁶¹ was employed for normalization and identification of differentially expressed genes in UBR7-shRNA and Control samples. All plots were generated using the ggplot and ggrepel packages in R. GSEA⁶² was run with normalized counts from all identified differentially expressed genes using the hallmark, curated, and gene ontology gene sets with default settings.

Statistical analysis. TCGA data analysis: Gene expression data for UBR7 from TCGA data were analyzed using the UCSC Xena functional genomics browser. Cancer subtypes for gene expression data were extracted from PAM50 version.

ChIP-qPCR and qRT-PCR: All qRT-PCR, ChIP, and other quantification data were collected in experiments performed in technical triplicate. Each experiment was repeated at least three times, and statistically significant results were obtained. An unpaired two-tailed Student's *t* test was performed using the Prism software program (GraphPad Software) to assign the significant differences between groups. Significant differences were considered when *P* < 0.05, **P* ≤ 0.05, ***P* ≤ 0.001, and ****P* ≤ 0.0001. Error bars indicate the standard deviation of the mean for the technical replicates, as indicated in the legend.

TMA: Two different TMAs were used: one for analyses of primary versus metastatic invasive ductal carcinoma tumors and the other for the association of UBR7 staining with ER, PR, and HER2 status. The intensity and percentage were measured three times for each tumor, and the means of these measures were used for data analyses (missing values were excluded). All statistical analyses were performed using R (version 3.3.1).

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability

ChIP-Seq and RNA-Seq data can be accessed at GEO using the accession number: [GSE93759](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93759). All relevant data are available from the authors upon request.

Received: 26 September 2017 Accepted: 9 February 2019

Published online: 28 March 2019

References

1. Polyak, K. Heterogeneity in breast cancer. *J. Clin. Invest.* **121**, 3786–3788 (2011).
2. Palchik, J. D. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
3. Viale, G. The current state of breast cancer classification. *Ann. Oncol.* **23**, x207–10 (2012).

4. Stephens, P. J. et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**, 400–404 (2012).
5. Diepenbruck, M. & Christofori, G. Epithelial–mesenchymal transition (EMT) and metastasis: yes, no, maybe? *Curr. Opin. Cell Biol.* **43**, 7–13 (2016).
6. Felipe Lima, J., Nofech-Mozes, S., Bayani, J. & Bartlett, J. M. EMT in breast carcinoma—a review. *J. Clin. Med.* **5**, 65 (2016).
7. Taube, J. H. et al. The H3K27me3-demethylase KDM6A is suppressed in breast cancer stem-like cells, and enables the resolution of bivalency during the mesenchymal–epithelial transition. *Oncotarget* **8**, 65548–65565 (2017).
8. Malouf, G. G. et al. Architecture of epigenetic reprogramming following Twist1-mediated epithelial–mesenchymal transition. *Genome Biol.* **14**, R144 (2013).
9. Chervona, Y. & Costa, M. Histone modifications and cancer: biomarkers of prognosis? *Am. J. Cancer Res.* **2**, 589–597 (2012).
10. Elsheikh, S. E. et al. Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.* **69**, 3802–3809 (2009).
11. Bhatnagar, S. et al. TRIM37 is a new histone H2A ubiquitin ligase and breast cancer oncoprotein. *Nature* **516**, 116–120 (2014).
12. Bochar, D. A. et al. BRCA1 is associated with a human SWI/SNF-related complex: linking chromatin remodeling to breast cancer. *Cell* **102**, 257–265 (2000).
13. Rai, K. et al. Dual roles of RNF2 in melanoma progression. *Cancer Discov.* **5**, 1314–1327 (2015).
14. Shema, E. et al. The histone H2B-specific ubiquitin ligase RNF20/hBRE1 acts as a putative tumor suppressor through selective regulation of gene expression. *Genes Dev.* **22**, 2664–2676 (2008).
15. Espinosa, J. M. Histone H2B ubiquitination: the cancer connection. *Genes Dev.* **22**, 2743–2749 (2008).
16. Cao, J. & Yan, Q. Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front. Oncol.* **2**, 26 (2012).
17. Fuchs, G. et al. RNF20 and USP44 regulate stem cell differentiation by modulating H2B monoubiquitylation. *Mol. Cell* **46**, 662–673 (2012).
18. Kato, A. & Komatsu, K. RNF20-SNF2H pathway of chromatin relaxation in DNA double-strand break repair. *Genes (Basel)* **6**, 592–606 (2015).
19. Kwon, Y. T. et al. The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **95**, 7898–7903 (1998).
20. Tasaki, T. et al. A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. *Mol. Cell Biol.* **25**, 7120–7136 (2005).
21. Tasaki, T. et al. The substrate recognition domains of the N-end rule pathway. *J. Biol. Chem.* **284**, 1884–1895 (2009).
22. Lee, M. J. et al. Synthetic heterovalent inhibitors targeting recognition E3 components of the N-end rule pathway. *Proc. Natl Acad. Sci. USA* **105**, 100–105 (2008).
23. Sanchez, R. & Zhou, M. M. The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **36**, 364–372 (2011).
24. Kim, J., Hake, S. B. & Roeder, R. G. The human homolog of yeast BRE1 functions as a transcriptional coactivator through direct activator interactions. *Mol. Cell* **20**, 759–770 (2005).
25. Mallery, D. L., Vandenberg, C. J. & Hiom, K. Activation of the E3 ligase function of the BRCA1/BARD1 complex by polyubiquitin chains. *EMBO J.* **21**, 6755–6762 (2002).
26. Xia, Y., Pao, G. M., Chen, H. W., Verma, I. M. & Hunter, T. Enhancement of BRCA1 E3 ubiquitin ligase activity through direct interaction with the BARD1 protein. *J. Biol. Chem.* **278**, 5255–5263 (2003).
27. Maunakea, A. K., Chepelev, I. & Zhao, K. Epigenome mapping in normal and disease States. *Circ. Res.* **107**, 327–339 (2010).
28. Zhu, J. et al. The UCSC Cancer Genomics Browser. *Nat. Methods* **6**, 239–240 (2009).
29. Jurikova, M., Danihel, L., Polak, S. & Varga, I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. *Acta Histochem.* **118**, 544–552 (2016).
30. Liu, F., Gu, L. N., Shan, B. E., Geng, C. Z. & Sang, M. X. Biomarkers for EMT and MET in breast cancer: an update. *Oncol. Lett.* **12**, 4869–4876 (2016).
31. Agiostratidou, G. et al. Loss of retinal cadherin facilitates mammary tumor progression and metastasis. *Cancer Res.* **69**, 5030–5038 (2009).
32. Pohl, S. G. et al. Wnt signaling in triple-negative breast cancer. *Oncogenesis* **6**, e310 (2017).
33. Cowin, P., Rowlands, T. M. & Hatsell, S. J. Cadherins and catenins in breast cancer. *Curr. Opin. Cell Biol.* **17**, 499–508 (2005).
34. Li, J. & Zhou, B. P. Activation of beta-catenin and Akt pathways by Twist are critical for the maintenance of EMT associated cancer stem cell-like characters. *BMC Cancer* **11**, 49 (2011).
35. Aasland, R., Gibson, T. J. & Stewart, A. F. The PHD finger: implications for chromatin-mediated transcriptional regulation. *Trends Biochem. Sci.* **20**, 56–59 (1995).
36. Kwan, A. H. et al. Engineering a protein scaffold from a PHD finger. *Structure* **11**, 803–813 (2003).
37. Borden, K. L. & Freemont, P. S. The RING finger domain: a recent example of a sequence-structure family. *Curr. Opin. Struct. Biol.* **6**, 395–401 (1996).
38. Kim, J. et al. RAD6-mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell* **137**, 459–471 (2009).
39. Werner, M. & Ruthenburg, A. J. The United States of histone ubiquitylation and methylation. *Mol. Cell* **43**, 5–7 (2011).
40. Chatterjee, C., McGinty, R. K., Fierz, B. & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nat. Chem. Biol.* **6**, 267–269 (2010).
41. Norton, L. & Massague, J. Is cancer a disease of self-seeding? *Nat. Med.* **12**, 875–878 (2006).
42. Halbleib, J. M. & Nelson, W. J. Cadherins in development: cell adhesion, sorting, and tissue morphogenesis. *Genes Dev.* **20**, 3199–3214 (2006).
43. Glukhova, M. A. & Streuli, C. H. How integrins control breast biology. *Curr. Opin. Cell Biol.* **25**, 633–641 (2013).
44. Berx, G. & Van Roy, F. The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. *Breast Cancer Res.* **3**, 289–293 (2001).
45. Adhikary, S. et al. Selective recognition of H3.1K36 dimethylation/H4K16 acetylation facilitates the regulation of all-*trans*-retinoic acid (ATRA)-responsive genes by putative chromatin reader ZMYND8. *J. Biol. Chem.* **291**, 2664–2681 (2016).
46. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91 (2008).
47. Su, X. et al. Tap63 suppresses metastasis through coordinate regulation of Dicer and miRNAs. *Nature* **467**, 986–990 (2010).
48. Liang, C. C., Park, A. Y. & Guan, J. L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. *Nat. Protoc.* **2**, 329–333 (2007).
49. Venkatanarayan, A. et al. IAPP-driven metabolic reprogramming induces regression of p53-deficient tumours in vivo. *Nature* **517**, 626–630 (2015).
50. Blecher-Gonen, R. et al. High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein–DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
52. Faust, G. G. & Hall, I. M. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503–2505 (2014).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
55. Liu, T. et al. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83 (2011).
56. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genom.* **15**, 284 (2014).
57. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
58. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
59. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
60. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
61. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
62. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

We thank Prof. Robert G. Roeder and Prof. Moshe Oren for providing us with the FLAG H2B wild-type and K120R mutant plasmids and RNF20 shRNA plasmid, respectively. We thank Drs. Michelle C. Barton, Jessica Tyler, and Tapas K. Kundu for critical comments on the manuscript. We acknowledge Scientific Editing team at MD Anderson Cancer Center for proofreading our manuscript. This work was supported in part by research grants from Biomolecular Assembly, Recognition and Dynamics Project (Grant 12-R&D-SIN-5.04-0103) from the Department of Atomic Energy, Swarnajayanti Fellowship, Department of Science and Technology and Ramalingaswami Fellowship, Department of Biotechnology, Government of India to C.D.; CSIR-Network Project (UNSEEN) and Ramanujan fellowship, Department of Science and Technology,

Government of India to S.R.; and National Cancer Institute grant (CA160578) and The University of Texas MD Anderson Cancer Center (start-up funds) to K.R. S.A., A.D., and D.K.S. thank the Council for Scientific and Industrial Research, University Grants Commission and Indian Council for Medical Research, Government of India, respectively, for funding their fellowship. D.C. is supported by the Triumph post-doctoral training program at MD Anderson Cancer Center supported by CPRIT (RP170067).

Author contributions

S.R., K.R., and C.D. conceived the study, designed experiments, and analyzed the data. S.A., D.C., C.T., I.S., and M.M. designed, performed experiments, and analyzed data. E.T., F.Y., and C.T. stained and scored T.M.A.s. A.A.S. provided reagents. J.M. and R.B. performed statistical analyses related to the TMAs. A.D., D.K.S., C.T., and A.T.R. performed the bioinformatic analysis. C.D., K.R., S.R., S.A., D.C., and C.T. wrote the paper.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-019-08986-5>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Journal peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019