Université de Montréal

The ARMC5-Cullin3-RBX1 Forms an RPB1-Specific Ubiquitin Ligase Essential for RNA Polymerase II Homeostasis

par

Linjiang Lao

Département de Médecine

Faculté de Médecine

Thèse présentée en vue de l'obtention du grade de

Philosophiæ Doctor (Ph.D.)

en Sciences Biomédicales

September 15, 2021

Université de Montréal

Faculté de Médecine

Cette thèse intitulée

The ARMC5-Cullin3-RBX1 Forms an RPB1-Specific Ubiquitin Ligase Essential for RNA Polymerase II Homeostasis

présentée par

Linjiang Lao

a été évaluée par un jury composé des personnes suivantes :

Dr. Francis Rodier
(président-rapporteur)

Dr. Jiangping Wu
(directeur de recherche)

Dr. Hongyu Luo
(codirecteur)

Dr. Claude Perreault
(membre du jury)

Dr. François Bachand
(examinateur externe)

Dr. Lea Harrington
(représentant du doyen de la FESP)

Résumé

ARMC5 est une protéine qui contient sept motifs Armadillo répétitifs organisés en tandem et un domaine BTB. Nous avons observé que cette protéine était fortement exprimée dans les organes lymphoïdes, les glandes surrénales et le cerveau. Les souris avec une délétion d'*Armc5* (souris KO) étaient de petite taille, et présentaient une diminution de la prolifération et la différenciation des lymphocytes T. L'absence d'ARMC5 entraînait une déficience de la réponse immunitaire médiée par les lymphocytes CD4$^+$ et CD8$^+$ dans les modèles expérimentaux d'encéphalomyélite auto-immune et d'infection au virus de la choriomeningite lymphocytaire, respectivement. Par la suite, plusieurs études ont révélé que la mutation *ARMC5* était associée à l'hyperplasie macronodulaire bilatérale primitive des surrénales (HMBPS), qui représente une cause rare du syndrome de Cushing. Nous avons ensuite confirmé que l'hyperplasie des glandes surrénales s'était développée chez les souris KO âgées, et qu'elle s'accompagnait d'une légère augmentation des taux sériques de glucocorticoïdes.

Comme ARMC5 ne présentait pas d'activité enzymatique, il était probable qu'elle faisait appel à d'autres protéines pour exercer sa fonction. Nous avons identifié plusieurs protéines qui se liaient à ARMC5, et plus particulièrement le complexe ARMC5/Cullin3 qui formait une ubiquitine ligase (E3) spécifique de la sous-unité RPB1 de l'ARN polymérase II. ARMC5 contrôlait le processus d'ubiquitination de RPB1 qui, par conséquent, s'accumulait dans plusieurs organes majeurs : les glandes surrénales, les ganglions lymphatiques, le cerveau, les poumons, le foie, etc. chez la souris KO. Ces résultats démontrent un rôle clé de l'ubiquitine ligase dans la dégradation de la protéine RPB1. Une accumulation similaire a également été observée dans les tissus hyperplasiques des surrénales provenant de patients atteints

d'HMBPS et porteurs de la mutation *ARMC5*, ce qui souligne la pertinence clinique de nos résultats de recherche fondamentale dans les maladies humaines. Un défaut de dégradation de RPB1 augmentait le pool d'ARN polymérase II. Par ailleurs, nous avons identifié un groupe de gènes fortement surexprimés dans les glandes surrénales déficientes en ARMC5, parmi lesquels figurent les gènes effecteurs qui seraient impliqués dans l'hyperplasie des surrénales chez les souris KO et l'HMBPS chez les patients porteurs de la mutation *ARMC5*.

Finalement, nous avons montré que la délétion ou la mutation d'*Armc5* augmentait considérablement le risque des anomalies du tube neural chez les souris et les humains. Chez les patients souffrant de myéloméningocèle, nous avons constaté neuf différentes mutations faux-sens délétères, dont une diminuait l'interaction entre ARMC5 et RPB1. L'augmentation du pool d'ARN polymérase II dans les cellules précurseurs neurales (CPN), causée par la délétion *ARMC5*, influençait un groupe particulier de gènes, dont certains (p. ex. *Folh1*) seraient susceptibles de participer au développement du tube neural.

En résumé, l'association ARMC5 et Cullin3 forme un complexe E3 qui cible RPB1 provoquant son ubiquitination et sa dégradation. En absence d'un tel mécanisme, on observe une perturbation de l'homéostasie de l'ARN polymérase II, qui mène à une diminution de la réponse immunitaire médiée par lymphocytes T, le développement d'HMBPS et un risque accru d'anomalies du tube neural.

Mots-clés: ARMC5 (Armadillo repeat containing 5), ubiquitine ligase (E3), Cullin3, ARN polymérase II, sous-unité RPB1 de l'ARN polymérase II (RPB1), pool d'ARN polymérase II, hyperplasie macronodulaire bilatérale primitive des surrénales (HMBPS), anomalies du tube neural, fonction des lymphocytes T.

Abstract

ARMC5 protein contains seven tandem Armadillo repeats and one BTB domain. We observed that *Armc5* was highly expressed in the lymphatic organs, adrenal glands, and brain. *Armc5* knockout (KO) mice were small in size and exhibited compromised T cell proliferation and differentiation. The absence of ARMC5 resulted in an impairment of the CD4$^+$ cell- and CD8$^+$ cell-mediated immune response in the experimental autoimmune encephalomyelitis model and lymphocytic choriomeningitis virus infection model, respectively. Subsequently, several studies revealed that *ARMC5* mutations were related to primary bilateral macronodular adrenal hyperplasia (PBMAH), which is a rare cause of Cushing's syndrome. We then confirmed that adrenal gland hyperplasia was indeed developed in aged *Armc5* KO mice with mildly increased serum glucocorticoid levels.

Since ARMC5 did not exhibit enzymatic activity, its function likely depends on the interaction with other proteins. We identified several proteins that binds to ARMC5, most notably ARMC5 binding to Cullin3, forming a ubiquitin ligase (E3) specific for RNA polymerase II subunit I (RPB1). ARMC5 regulated the ubiquitination of RPB1, and its deletion resulted in RPB1 accumulation in major organs (e.g., adrenal glands, lymph nodes, brain, lung, and liver), indicating the critical role of this E3 in RPB1 degradation. A similar accumulation was also found in hyperplasia tissues from adrenal glands of PBMAH patients carrying *ARMC5* mutations, underscoring the clinical relevance of our basic research findings in human disease. Defective degradation of RPB1 led to an enlarged RNA polymerase II (Pol II) pool. In addition, we have identified a group of genes strongly upregulated in

KO adrenal glands, including the effector genes which would be involved in adrenal gland hyperplasia in *Armc5* KO mice and PBMAH patients carrying *ARMC5* mutation.

Finally, we have shown that deleting or mutating *Armc5* significantly augments the risk of neural tube defects in mice and humans. In patients with myelomeningocele, we found nine deleterious missense mutations in *ARMC5*, one of which weakened the interaction between ARMC5 and RPB1. The enlarged Pol II pool in *Armc5* KO neural precursor cells (NPCs) influenced a particular group of genes, some of which (e.g., *Folh1*) are thought to be involved in the development of the neural tube.

In summary, ARMC5 and CUL3 form an E3 complex, which targets RPB1 causing its ubiquitination and degradation. In the absence of such a mechanism, there is a disturbance of RNA polymerase II homeostasis, which leads to a decrease in the T cell-mediated immune response, the development of PBMAH and an increased risk of neural tube defects.

Keywords: ARMC5 (Armadillo repeat containing 5), ubiquitin ligase (E3), Cullin3, RNA polymerase II, RNA polymerase II subunit I (RPB1), RNA polymerase II pool, primary bilateral macronodular adrenal gland hyperplasia (PBMAH), neural tube defects, T cell function

Contents

11

17

List of tables

List of figures

Liste des sigles et des abréviations

Abbreviation	Definition
3C	Chromosome Conformation Capture
AA	Amino Acid
ACN	Acetonitrile
AC	Adenylyl Cyclase
ACTH	Adrenocorticotropic Hormone
AF	Allele Frequency
AG	Adrenal Gland
AN	Allele Number
ANOVA	Analysis of Variance
AP	Affinity Precipitated
APC/C	Anaphase-Promoting Complex/Cyclosome
APPBP2	Amyloid Beta Precursor Protein Binding Protein 2
ARM	Armadillo
ARMC5	Armadillo Repeat-Containing 5
ATP	Adenosine Triphosphate
B	Bone
BM	Bone Marrow
BRCA1	Breast Cancer gene 1
BREu	TFIIB Recognition Element upstream
BREd	TFIIB Recognition Element downstream

BSA	Bovine Serum Albumin
BTB	Broad-Complex, Tramtrack, and Bric-Abrac
cAMP	cyclic Adenosine MonoPhoshate
CAr	Central Artery
Cb	Cerebellum
CDK	Cyclin Dependent Kinase
CFSE	Carboxyfluorescein Succinimidyl Ester
ChIA-PET	Chromatin Interaction Analysis by Paired-End Tag
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
CP	Core Particle
CPM	Count Per Million
CPSF	Cleavage and Polyadenylation Specificity Factor
CRBN	Cereblon
CRH	Corticotropin Releasing Hormone
CRL	Culling Ring Ligase
CS	Cushing's Syndrome
CSB	Cockayne Syndrome Group B
CstF	Cleavage Stimulatory Factor
CTCF	Corrected Total Cell Fluorescence
CTD	C-Terminal Domain
CUL	Cullin
Cx	Cortex
CYLD	Cylindromatosis
DAPI	4,6-diamidino-2-phenylindole

DAPK1	Death Associated Protein Kinase 1
DCAF	DDB1 And CUL4 Associated Factor 1
DCE	Downstream Core Element
DDB1	Damage Specific DNA Binding Protein 1
DiGly	Gly-Gly-ϵ-Lys
DMEM	Dulbecco's modified Eagle's medium
DN	Double Negative
DP	Double Positive
DPE	Downstream Promoter Element
DRB	5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole
DSIF	DRB Sensitivity Inducing Factor
DUB	Deubiquitinase
E	Embryonic day
E1	Ubiquitin-Activating Enzyme
E2	Ubiquitin-Conjugating Enzyme
E3	Ubiquitin Ligase
EAE	Experimental Autoimmune Encephalomyelitis
ECL	Enhanced ChemiLuminescence
EDTA	EthyleneDiamine Tetraacetic Acid
EGF	Epidermal Growth Factor
ELISA	Enzyme-Linked Immunoassay
ERCC1	ERCC Excision Repair 1
ES	Embryonic Stem
FA	Formic Acid
FAK	Focal adhesion kinase

FAP	Familial Adenomatous Polyposis
FBS	Fetal Bovine Serum
FC	Fold Change
FDR	False Discovery Rate
ffu	focus-forming unit
FH	Fumarate Hydratase
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GG-NER	Global Genome Nucleotide Excision Repair
gnomAD	genome Aggregation Database
GO	Gene Ontogeny
GTF	General Transcription Factor
HA	Hemagglutinin
HE	Hematoxylin/Eosin
HECT	Homologous to the E6-AP Carboxyl Terminus
HEK	Human Embryonic Kidney
HIF	Hypoxia Inducible Factor
HLRCC	Hereditary Leiomyomatosis and Renal Cell Carcinoma
HPLC	High Performance Liquid Chromatography
HRP	HorseRadish Peroxidase
HSP	Heat Shock Protein
HSV	Herpes Simplex Virus
IB	Immunoblotting
IBR	In-Between-Ring
IFN	Interferon
IL	Interleukin

INT	Integrator
IP	Immunoprecipitation
ISH	*in situ* hybridization
KCTD	Potassium Channel Tetramerization Domain
KLHL	Kelch-like
KLRG1	Killer Cell Lectin Like Receptor G1
KO	Knockout
LC-MS/MS	Liquid Chromatography with tandem Mass Spectrometry
LCMV	Lymphocytic Choriomeningitis Virus
LD	Linkage Disequilibrium
Lint	Large intestine
LT	Lymphatic Tissue
LUBAC	Linear Ub Chain Assembly Complex
MAS	McCune-Albright syndrome
MC2R	MelanoCortin 2 Receptor
Me	Medulla
MEFs	Mouse Embryonic Fibroblast Cells
MEN	Multiple Endocrine Neoplasia
METAP2	Methionyl Aminopeptidase 2
MG-	Glycine after Methionine
miRNA	microRNA
MOG	Myelin oligodendrocyte glycoprotein
MP-	Proline after Methionine
MPEC	Memory Precursor Effector Cell
MTE	Motif Ten Element

MW	Molecular Weight
MS	Mass Spectrometer
NELF	Negative Elongation Factor
NEM	N-ethylmaleimide
NER	Nucleotide Excision Repair
NF-κB	Nuclear Factor Kappa B Subunit 1
NLS	Nuclear Localization Signal
NPCs	Neural Precursor Cells
NPC	Nuclear Pore Complex
NS	Not Significant
NTD	Neural Tube Defect
OTU	Ovarian Tumor Protease
PAF1	Polymerase Associated Factor 1
PAS	Polyadenylation Signal
PBC	Predicted Biological Confidence
PBMAH	Primary Bilateral Macronodular Adrenal gland Hyperplasia
PBS	Phosphate-Buffered Saline
PCP	Planar Cell Polarity
PCR	Polymerase Chain Reaction
PDE	Phosphodiesterase
PI	Pausing Index
PIC	Pre-initiation Complex
PKA	Protein Kinase A
PMA	Phorbol 12-myristate 13-acetate
PMAH	Primary Macronodular Adrenal gland Hyperplasia

Pol II	RNA polymerase II
PPNAD	Primary Pigmented Nodular Adrenocortical Disease
pre-mRNA	pre-Messenger RNA
pre-rRNA	pre-Ribosomal RNA
PRKX	Protein Kinase X-Linked
PROTAC	Proteolysis-Targeting Chimera
PSAQ	Protein Standard Absolute Quantification
PTM	Posttranslational Modification
pVHL	The Von Hippel-Lindau tumor suppressor
RBR	RING-Between-RINGs
RBX	Ring-Box
RIG	Retinoic acid-Inducible Gene
RING	Really Interesting New Gene
RIPA	Radioimmunoprecipitation assay
RIPK2	Receptor Interacting Serine/Threonine Kinase 2
RNAP	RNA Polymerase
RNF	Ring Finger Proteins
RP	Regulatory Particle
RPAP	RNA polymerase II Associated Protein
RPB	RNA polymerase II Subunit
RPB1-IIa	RPB1 with hypo-phosphorylated CTD
RPB1-IIo	RPB1 with hyper-phosphorylated CTD
RT-qPCR	Quantitative reverse transcription PCR
SCF	Skp, Cullin, F-box containing complex
SDS	Sodium Dodecyl Sulfate

SDS-PAGE	Sodium Dodecyl-Sulfate PolyAcrylamide Gel Electrophoresis
SE	Standard Error
SEM	Standard Error of the Mean
SEC	Super Elongation Complex
Shh	Sonic Hedgedog
SI	Supplementary Information
SILAC	Stable Isotope Labeling with Amino acids in Cell culture
SIM	Substrate Interaction Motif
siRNA	small interfering RNA
Sk	Skin
SLEC	Short-Lived Effector Cells
snoRNA	Small nucleolar RNA
snRNA	small nuclear RNA
SNV	Single Nucleotide Variant
SP	Single Positive
St	Stomach
STAR	Steroidogenic Acute Regulatory Protein
TAF	TBP-Associated Factor
TBP	TATA box-Binding Protein
TC-NER	Transcriptional-Coupled Nucleotide Excision Repair
TCR	T-cell receptor
TES	Transcription Ending Site
TF	Transcription Factor
Th	Thymus
TRAF	TNF Receptor-Associated Factor

TRIM	The Tripartite Motif Family
tRNA	transfer RNA
TSS	Transcription Starting Site
Ub	Ubiquitin
UbA	Ubiquitin-Associated
UBB	Ubiquitin B
UBC	Ubiquitin C
UBD	Ubiquitin-Binding Domain
UbL	Ubiquitin-Like
UCH	Ubiquitin C-terminal Hydrolase
UPS	Ubiquitin-Proteasome System
USP	Ubiquitin-Specific Protease
UV	Ultraviolet
Vb	Vertebrae
WT	Wild-Type
Y2H	Yeast Two Hybrids
ZUP	Zinc finger-containing Ubiquitin Peptidase

This thesis is dedicated to

My parents, my sister, and my dear wife Qi for your love and support.

Remerciements

I feel tremendously grateful for having Dr. Jiangping Wu and Dr. Hongyu Luo as my supervisor and co-supervisor. Thanks for your patience, motivation, encouragement, and guidance throughout my Ph.D. study. You provide me an unfettered science journey to fulfill my curiosity and self-challenging. I really enjoyed the experience of digging the unknown knowledge, although the whole process was accompanied by the numerous times we overturned our hypotheses.

I would like to express my deepest appreciation to my thesis committee: Dr. Francis Rodier, Dr. Claude Perreault, Dr. François Bachand, and Dr. Lea Harrington for their insightful comments and meaningful encouragement. Thank you for spending time reading and commenting on my thesis. Besides, thanks to Dr. Marika Sarfati for translating and revising the French abstract of this thesis.

I would also like to extend my sincere thanks to my colleagues, Dr. Wei Shi, Xiao He, Dr. Yan Hu, Dr. Yujia Wang, Dr. Shanshan Lan and Dr. Bing Yang for the help and mental support you kindly offered. Many thanks to Dr. Junzheng Peng, as the completion of these projects would not have been possible without your help and suggestions. Special thanks to Dr. Wei Shi for this invaluable life experience we had shared in Canada all these years.

I gratefully acknowledge the assistance of Dr. Fragoso Gabriela for the patient and professional guidance of the pathological experiments. I had a great pleasure and honor of working with our collaborators, Dr. Isabelle Bourdeau, Dr. Lucia Gagliardi, Dr. Bingbing Hao, Dr. Minjia Tan, Dr. David Torpy, Dr. Hamish Scott, and Dr. Andre Lacroix. Thanks for your suggestions and contribution to this project.

I very much appreciate my M.Sc. supervisor Dr. Lizhong Du for introducing me to Dr. Wu's lab. This is a wonderful suggestion considering my research interests and career planning.

Finally, I would like to expand my gratitude to my parents, my sister, and all my family members for your unconditional love and support. To the most important, my wife Qi Wang, you have witnessed every second of happiness and sorrow during my Ph.D. life. Without your love and support, I would not be able to hold on to the last.

Thank you all for giving me belief and strength to carry on, and I really appreciate it.

Chapter 1

Introduction

1.1. The general features of ARMC5

1.1.1. ARMC5 protein and gene

Armadillo repeat-containing 5 (ARMC5) was a protein with an unknown function. We believed that it might have functions related to T cell activation when our lab started to investigate it in 2007. Recently, it has been reported that some mutations of *ARMC5* are linked to an increased risk of primary bilateral macronodular adrenal hyperplasia (PBMAH)[1,2,3] and meningiomas[4,5].

ARMC5 gene is located at chromosome 16p11.2. ARMC5 protein contains seven Armadillo repeats and one broad-complex, tramtrack, and bric-abrac (BTB) domain (**Figure 1.1**). The Armadillo repeat consists of approximately 40 amino acids. It was first identified in *Drosophila* segmentation gene (the homolog of mammalian *β-catenin*). Each repeat is composed of three α-helices[6]. Tandem repeats form a right-hand superhelix of helices and create a groove to bind to other proteins. The BTB domain consists of approximately 120 amino acids and also acts as a protein-protein interaction module. Most of the BTB domain-containing proteins are evolutionary conserved, and involved in various biological processes. Many of them are related to ubiquitin-dependent protein modification and degradation[7].

Human *ARMC5* has four complete transcription isoforms according to the Ensembl database (`www.ensembl.org`): *ARMC5*-201, *ARMC5*-202, *ARMC5*-203, and *ARMC5*-204. The distribution of their exons and introns is shown in **Figure 1.1**. *ARMC5*-201 is the most common one and cited by many studies[1,8]. *ARMC5*-202 and *ARMC5*-204 have two extra short exons and one exon, respectively, before the first exon of *ARMC5*-201. *ARMC5*-203 does not contain the last two exons, which are presented in the other three isoforms[9].
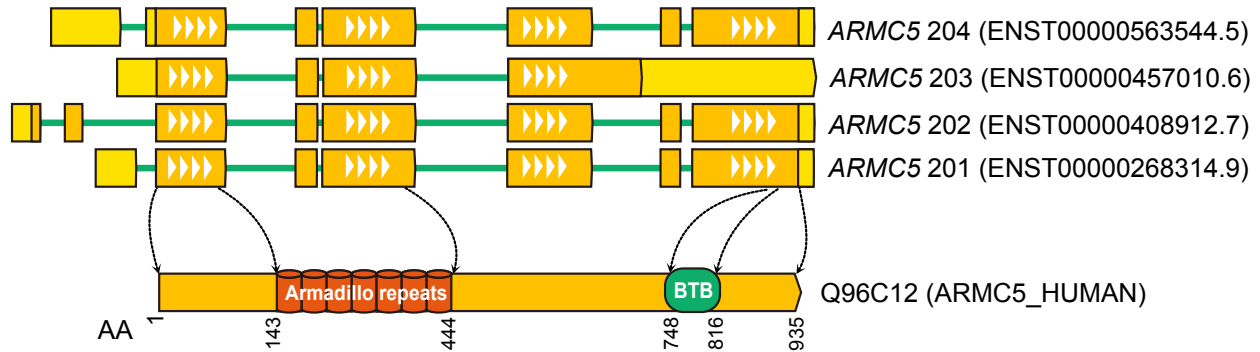
Fig. 1.1. The different isoforms of *ARMC5* transcripts and the distribution of domains in ARMC5 protein The distribution of domains in ARMC5 protein is based on UniProtKB (Q96C12). The protein is transcribed from *ARMC5*-201 (ENST00000268314.9). The position of exons (yellow squares) in *ARMC5* (201, 202, 203, and 204) refers to the Ensembl database. The dark yellow regions indicate the protein coding sequence. AA: amino acid.

1.1.2. ARMC5 function

In this section, I will review the known function of ARMC5, including the findings from my published paper, which is part of my thesis and will be detailed later, as the published paper is already part of the literature.

Armc5 was among the upregulated genes during T cell activation, according to our study[8]. It was highly expressed in the thymus, adrenal glands, lymphatic tissues, bone marrow, cerebellum, and skin, based on *in situ Armc5* hybridization. We generated *Armc5* knockout (KO) mice by replacing the first three exons of *Armc5* with the *neo* gene. Only 10% live pups were delivered in the C57BL/6J × 129/sv background with a heterozygous × heterozygous mating strategy, which is below the expected 25%[8]. Berthon et al. also tried to generate *Armc5* deficiency mice with a mixed background of C57/BL6, 129 Sv/J

and Swiss. Only 0.025% of *Armc5* KO mice survived in their case[10]. Almost all *Armc5* KO fetuses died during the early stage of embryonic development. They showed in that genetic background, *Armc5* KO mice failed to form mesoderm and gastrulation at E7.5[10].

Both our lab and Stratakis's team found that *Armc5* KO mice were smaller than wild-type (WT) controls from the embryonic stage to adulthood[8,10]. However, when we examined serum growth hormone levels in KO mice, no significant difference was found compared to that of WT ones.

Since ARMC5 was later found to be linked to primary bilateral macronodular adrenal hyperplasia (PBMAH) and Cushing's syndrome[1,2,3], we assessed the sizes of adrenal glands and serum glucocorticoid levels in young and aged *Armc5* KO mice. No significant difference was found in young mice, whereas in old (>15 months) KO ones, enlarged adrenal glands and mildly increased glucocorticoid levels were observed.

Our results also showed that ARMC5 played a vital role in immune responses. CD4$^+$ T cells, CD8$^+$ T cells, and B cells in *Armc5* KO mice had compromised proliferation after activation. The differentiation of naïve KO CD4$^+$ cells into Th1 and Th17 was also defective. The differentiated Th1 and Th17 cells showed the same slower proliferation phenotype as CD4$^+$, CD8$^+$ and B cells. Cell cycle analysis revealed that *Armc5* KO T cells were partially blocked during G1/S progression. In addition, T cells were prone to apoptosis induced by FasL.

To better understand the role of ARMC5 in CD4$^+$ and CD8$^+$ cell-mediated immune responses, our lab employed an experimental autoimmune encephalomyelitis (EAE) model and a lymphocytic choriomeningitis virus (LCMV) infection model in *Armc5* WT and KO mice, respectively. The onset of EAE clinical symptoms was delayed by seven days in KO mice. Meanwhile, the maximum disease score of KO mice was also lower than that of WT controls. KO mice had fewer antigen-specific T cells in the draining lymph nodes and fewer infiltrating mononuclear cells in the brain and spinal cord after myelin oligodendrocyte glycoprotein (MOG) immunization. On the other hand, the clonal expansion of CD8$^+$ cells in KO mice was compromised after LCMV antigen stimulation. Viral titers in the kidneys,

liver, and spleen were significantly higher in KO mice after LCMV infection, indicating the defect function of viral clearance.

Since ARMC5 does not exhibit enzymatic activity, its possible functions depend on the interactions with other proteins. To identify the binding partners, our lab conducted a yeast two-hybrid (Y2H) with human ARMC5 protein as bait and a human thymocyte cDNA expression library as prey. We identified 16 possible binding patterns with high scores, such as DAPK1, ARMC5 itself, STK24. TTF1, RPB1, and CUL3. By studying the interaction with these possible partners, we may better understand ARMC5's molecular mechanism.

1.2. Primary bilateral macronodular adrenal hyperplasia

As *ARMC5* mutations are related to an increased PBMAH risk, a short review of this disease is given in the following section.

1.2.1. Cushing's syndrome

Cushing's syndrome (CS) is caused by prolonged tissue exposure to excessive glucocorticoids from either endogenous or, more commonly, exogenous sources. Characteristic features of Cushing's syndrome include weight gain around the midsection and upper back, exaggerated facial roundness, slow healing of cuts, pink or purple stretch marks on the abdomen skin, osteoporosis, and decreased growth velocity in children[11].

Three categories of primary tumors can cause endogenous CS: ectopic ACTH or corticotropin-releasing hormone (CRH)-secreting neuroendocrine tumors, adrenocorticotropic hormone (ACTH)-secreting pituitary adenomas (Cushing's disease), and primary adrenal tumors. Cushing's disease is the most common cause of endogenous CS, which comprises $60-70\%$. Approximately $20-30\%$ of endogenous CS cases are caused by primary adrenocortical hyperfunction, while ectopic ACTH or CRH-secreting tumors account for the remaining $5-10\%$[11].

Among the primary adrenocortical hyperfunction cases, benign cortisol-producing adenomas account for $10 - 15\%$, and adrenal hyperplasia occupies 20%, which is almost always bilateral. In contrast, adrenocortical carcinomas are much less frequent $(< 5\%)$[11].

1.2.2. The general features of PBMAH

Bilateral adrenal hyperplasia is characterized by multiple nodules and can be classified into two subgroups based on the size of most nodules. Micronodular adrenal hyperplasia refers to adrenal hyperplasia with nodules less than 1 cm, while macronodular adrenal hyperplasia, such as PBMAH, with nodules larger than 1 cm and frequently reaching to $3 - 4$ cm in diameter. PBMAH is considered as a rare $(< 2\%)$ cause of endogenous CS. The bilateral nature of PBMAH suggests a genetic origin.

Several other terms have been used to describe PBMAH disease, such as massive macronodular hyperplasia[12], ACTH-independent massive bilateral adrenal disease[13], macronodular adrenal hyperplasia[14], nodular hyperplasia of the adrenal glands[15]. However, currently, the prevailing nomenclature for this disease is PBMAH.

Most of PBMAH is diagnosed in patients between 40- to 70-year-old based on the clinical signs of cortisol excess. The excess level of glucocorticoids is usually mild. Therefore, Cushing's symptoms are also mild. Macroscopic examination shows many yellowish nodules of different sizes. The histological analysis presents with several island-like structures in non-pigmented nodules.

1.2.3. The cause of PBMAH

1.2.3.1. ARMC5 and PBMAH.

In 2013, Assie et al.[1] genotyped germline and tumor DNA from patients with PBMAH and identified *ARMC5* mutations in tumors obtained from 18 out of 33 ($\sim$55%) patients. Both alleles of *ARMC5* were mutated, as one germline mutation and the other somatic mutation. Later, a cohort of 34 PBMAH patients from the National Institutes of Health (NIH) Clinical Research Center showed germline *ARMC5* mutations in 15 patients ($\sim$44%) and

all patients carrying pathogenic *ARMC5* mutations manifested clinical CS[14]. In another large cohort of PBMAH patients with subclinical or clinical CS, ARMC5-damaging mutations were identified in 24 patients out of total 98 patients (∼24%). The *ARMC5* mutation carriers showed a severer hypercortisolism and larger adrenal nodules[16].

More PBMAH families have been identified carrying *ARMC5* mutations[17,18,2,19,3,20]. The PBMAH patients with *ARMC5* mutation were more like to have overt CS, more and larger adrenal nodules, compared to the ones without *ARMC5* mutation[16].

Stop codon mutations, missense mutations, frame-shift mutations, deletion mutations, and point mutations of *ARMC5* have been documented at the germline or somatic level. All mutations found to date are summarized and illustrated in **Figure 1.2**.
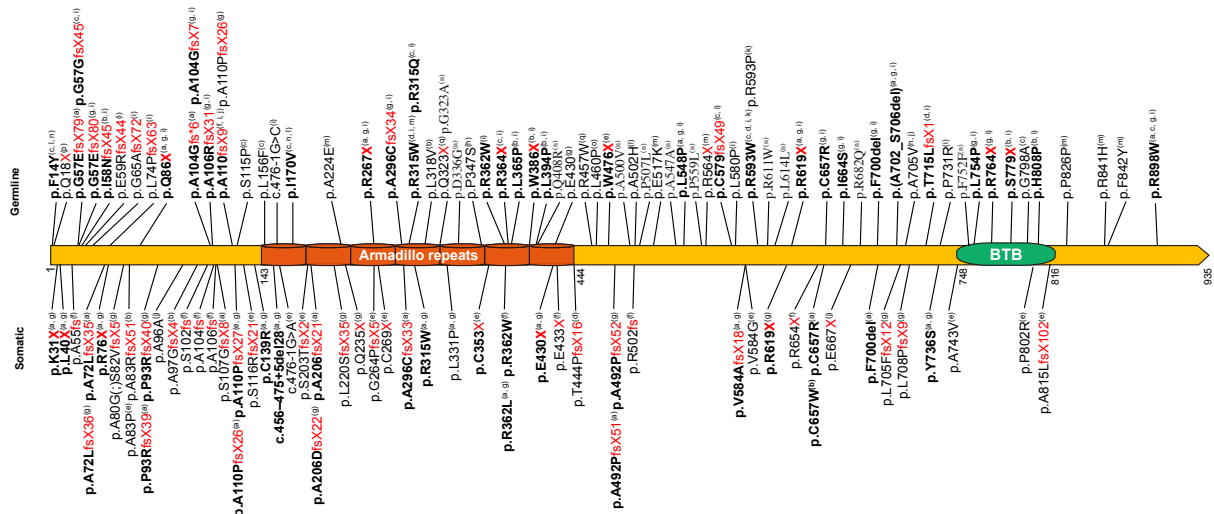
Fig. 1.2. *ARMC5* germline and somatic mutations related PBMAH
The mutations up to date in ARMC5 protein are based on the transcript *ARMC5*-201 (ENST00000268314.9). Above: germline mutation. Below: somatic mutation only or somatic mutation with undetermined germline status. The red font indicates stop codon or frameshift mutation. The bold font indicates the mutation sites have been found in more than two independent patients. Studies from (a) Paris, France[1], (b) Sao Paulo, Brazil[18], (c) Maryland, United States[14], (d) Adelaide, Australia[3], (e) Maryland, United States[21], (f) Berlin, Germany[19], (g) Paris, France[16], (h) Chiba, Japan[22], (i) Padua, Italy[17], (j) Quebec, Canada[2], (k) Oxford, UK[23], (l) Padova, Italy[24], (m) Beijing, China[20], (n) United States[25], (o) Porto, Portugal[5], (p) Liaoning, China[26], (q) Wuhan, China[27], are included.

1.2.3.2. The cAMP-PKA pathway and PBMAH.

Several other genes related to PBMAH were investigated before the discovery of *ARMC5*. The abnormal cyclic adenosine monophosphate (cAMP)-protein kinase A (PKA) signaling has been implicated in most benign cortisol-producing adrenal tumors[28,29]. It is activated by ACTH via melanocortin 2 receptor (MC2R), which is a G-protein melanocortin receptor located at the adrenal cell membrane. Up to date, two patients with *MC2R* mutation have been documented, but only one of them was diagnosed with PBMAH[30]. Thus, *MC2R* mutation can be considered as a rare causative factor of PBMAH.

The binding between ACTH and MC2R releases the α subunit (encoded by *GNAS*) from the $\beta\gamma$ dimer of the G protein. The α unit then binds to and activates adenylyl cyclase (AC), and converts adenosine triphosphate (ATP) to cAMP. Some *GNAS* mutations, that lead to overactivation of the cAMP-PKA pathway, are one of the genetic causes of McCune-Albright syndrome (MAS)[31]. MAS is manifested by symptoms related to the skeleton (Fibrous dysplasia), the endocrine organs (e.g., early puberty, excessive growth hormone), and the skin (café-au-lait spots)[32]. Primary biomorphic adrenocortical disease has also been reported in association with MAS[33].

PKA is a stable and inactive heterotetramer, consisting of two catalytic subunits and two regulatory subunits. The activation of PKA is triggered by the binding between cAMP molecules and regulatory subunits, which leads to the dissociation of the two catalytic subunits. When activated, the catalytic subunits phosphorylate specific nuclear factors, which facilitate steroidogenic gene transcription. The activity of PKA is highly related to cell proliferation, cell metabolism, and gene transcription. Humans have four isoforms of the catalytic subunits (Cα, Cβ, Cγ, and PRKX) and four isoforms of the regulatory subunits (R1α, R1β, R2α, and R2β). *PRKAR1A* encodes R1α of PKA. Inactivation of PRKAR1A leads to the loss of the regulatory function of PKA, therefore activating the cAMP-PKA pathway. Germline inactivating mutations of *PRKAR1A* have been found in ∼37% of sporadic Carney complex patients and more than 70% of familial Carney complex patients[34]. The Carney complex is a dominantly inherited syndrome characterized by multiple endocrine

neoplasia, including primary pigmented nodular adrenocortical disease (PPNAD). Not only the Carney complex, *PRKAR1A* mutations have also been reported in cortisol-producing adenomas and adrenocortical carcinoma[35]. In addition to *PRKAR1A*, gain-function mutation of *PRKACB* has been found in a patient with Carney complex[36], which activated the PKA signaling.

The cAMP signal is terminated by phosphodiesterases (PDEs), which degrade cAMP to AMP. The inactivating mutations of *PDE11A* and *PDE8B* increase the level of cAMP and lead to persistant activation of the cAMP-PKA signaling. *PDE11A* contains many polymorphic variants in the general population, and some rare ones with reduced function have been implicated in PPNAD and PBMAH[37,38]. In addition, *PDE8B* mutation have also been documented in PBMAH, PPNAD, and non-secreting adenomas[39].

The schematic diagram of the cAMP-PKA pathway is shown in **Figure 1.3**.
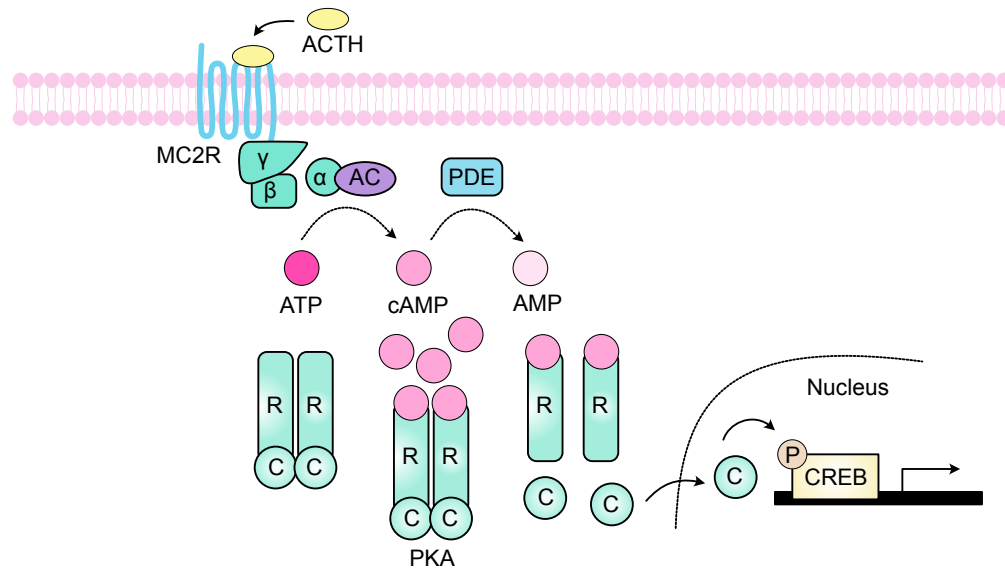
Fig. 1.3. The cAMP-PKA pathway The binding between ACTH and MC2R releases the α subunit from the $\beta\gamma$ dimer of the G protein. The α unit activates adenylyl cyclase (AC), and converts ATP to cAMP. Two cAMP bind to each regulatory subunit (R) of PKA, leading to the dissociation of two catalytic subunits (C). Phosphodiesterases (PDEs) terminate the cAMP signal by converting cAMP to AMP. The catalytic subunit phosphorylates CREB transcription factor, which facilitates gene transcription.

1.2.3.3. Mutations of other genes associated with PBMAH.

Multiple endocrine neoplasia type 1 (MEN1) is a dominantly inherited syndrome caused

by *MEN1* mutations. The frameshift or nonsense mutation in *MEN1* leads to a truncated protein, and results in reduced levels of MEN1 protein. MEN1 is manifested by multiple tumors of endocrine glands (e.g., parathyroid gland and pituitary gland). Adrenocortical tumors or hyperplasia presents in 30 – 40% of MEN1 patients[40]. PBMAH is also found in some MEN1 patients, but at a very low incidence rate[41].

Familial adenomatous polyposis (FAP) is another inherited disease caused by inactivating mutations of *APC* gene. FAP is characterized by numerous colorectal adenomatous polyps. It is also associated with endocrine tumors[42] and adrenocortical tumors[43]. Approximately 13% (14/107) of FAP patients have adrenal masses larger than 1 cm. However, none of them shows clinical symptoms of endocrine disturbance[43].

Germline mutations in *FH* gene caused hereditary leiomyomatosis and renal cell carcinoma (HLRCC). FH protein is a Krebs cycle enzyme that converts fumarate to malate. The inactivation of FH causes the accumulation of fumarate in the cytoplasm. HLRCC patients tend to develop several benign tumors, including smooth muscle tumors and/or papillary renal carcinoma. Twenty out of 255 HLRCC patients had primary adrenal lesions. The pathological examination reveals macronodular adrenal hyperplasia in all nodules[44].

1.3. Neural tube defects

In addition to PBMAH, we will later prove that *ARMC5* mutations are also related to increased risk of neural tube defects (NTDs). A brief review of NTDs is given in this section.

NTDs, including spina bifida, anencephaly, and craniorachischisis, are severe birth defects with a prevalence of 0.5 to 10 per 1000 established pregnancies[45]. NTDs are caused by the failure of neural tube closure during embryonic development. The failed closure leads to the exposure of the neural tube to the amniotic fluid, thereby causing neuroepithelial degeneration *in utero* and loss of massive neural tissue. The exposure of the brain in NTDs, such as craniorachischisis and anencephaly, are frequently lethal for fetuses, whereas spina bifida is not lethal but often causes physical and intellectual disabilities.

1.3.1. The embryonic basis of NTDs

To understand better of the morphological process of the failed neural tube closure, I will briefly introduce the neurulation process during embryogenesis.

The brain and spinal cord are developed from the neural tube through a process named neurulation. The neural tube is folded and fused from the neural plate, which is originated from the dorsal ectoderm. The closure process initiates sequentially from different sites of the embryo-sagittal axis. Several regions, termed as neuropores, that are open during the folding and fusing process, will be completely closed at the end of the entire neurulation.

In mice, the primary neurulation initiates from Closure 1 on embryonic day 8.5 (E8.5) at the boundary of the cervical position and hindbrain, followed by the fusion spreading bidirectionally into the hindbrain and the posterior neuropore. The failure of Closure 1 leads to craniorachischisis, the most severe NTD type, with both the brain and entire spinal cord remaining open.

The second and third closure occur on E9 at the midbrain/forebrain boundary and the rostral part of the forebrain. The failure of Closure 2 leads to anencephaly, which is also a severe NTD in which a baby is born without parts of the skull and brain. If Closure 3 is not well completed, which is uncommon, the resulting phenotype is a split face, commonly with forebrain anencephaly.

The anterior neuropore (between Closure 2 and Closure 3) is gradually shortened and closed on E9, followed by the closure of the hindbrain neuropore (between Closure 1 and 2). The fusion process continues in one direction along the spinal axis, reaching the posterior neuropore, and finally closes at the second sacral segment level. The impaired closure of the posterior neuropore causes a persistently open region, termed as open spina bifida, which is also known as myelomeningocele. The entire primary neurulation completes on E10.

The secondary neurulation at the lower sacral segments is accomplished by a population of tail bud-derived cells with neural fate, which then forms the lumen of the tube and coccygeal regions.

The primary and secondary neurulation processes are almost conserved among mammalians. Compared to the neurulation in mice, the process in humans appears to lack Closure 2 at the midbrain/forebrain boundary. Therefore, Closure 3 from the rostral part and Closure 1 from the hindbrain directly fuse at the rostral neuropore. The schematic diagram of neurulation in mice and humans is shown in **Figure 1.4**.
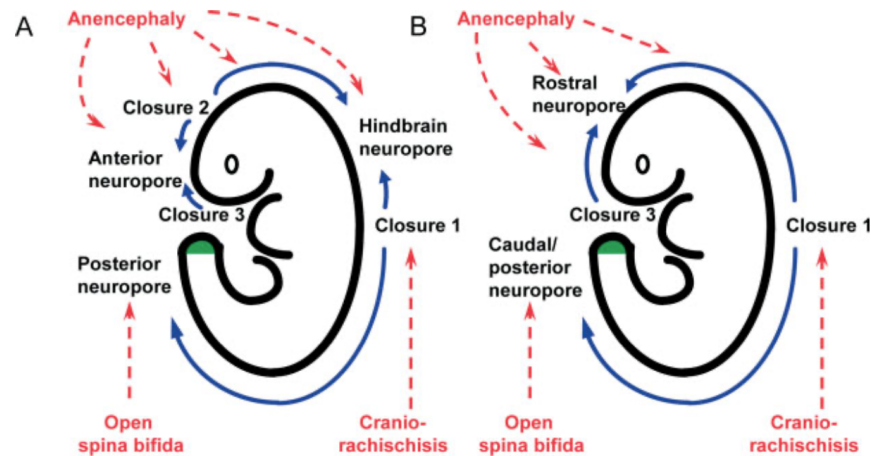
Fig. 1.4. The neural tube closure and its related NTDs in mice and humans **(a)** mouse embryos, **(b)** human embryos. The complete neural tube closure process consists of three steps of closure (Closure 1, 2, and 3) and several fusion events at neuropores. The neural plate is zippered into the neural tube unidirectionally or bidirectionally, as indicated in blue arrows. Different NTDs caused by the failure of different events are indicated by red arrows. Secondary neurulation occurs from the closed posterior neuropore level (green region). The schematic diagram is adapted from the publication of Copp and Greene[46].

1.3.2. Multifactorial causation of human NTDs

Epidemiological studies have suggested that NTDs were caused by environmental and genetic factors in humans.

1.3.2.1. Environmental factors.

In humans, various environmental risk factors have been associated with NTDs, such as folate supplementation[47], maternal obesity and diabetes[48], and teratogenic agents (e.g., the anticonvulsant drug valproic acid[49] and the fungal product fumonisin[50]).

It was initially reported that the B-vitamin folate levels were lower than normal in mothers of NTD fetuses[47]. A multiple-center randomized controlled trial later proved that maternal folate supplementation (4 mg daily) reduced the incidence of NTDs significantly[51]. Furthermore, a community-based intervention in China documented the effectiveness of folic acids (0.4 mg daily) supplementation in the prevention of NTDs in some high incidence areas[52].

However, the mechanism by which maternal folic acid supplementation prevents NTDs remains unclear. Although folate deficiency is a high-risk factor, in most people, the folate level is rarely clinically deficient. Only in individuals with susceptible genetic backgrounds, suboptimal levels of folic acid may lead to NTDs[53].

Maternal obesity and diabetes are two risk factors related to NTDs. The underlying mechanism remains uncertain due to the complexity of the diabetic milieu, although it has been proven that hyperglycemia alone is sufficient to cause NTDs in mice. Oxidative stress and cell apoptosis under hyperglycemic conditions may have roles in inducing NTDs[54,55].

1.3.2.2. Genetic factors.

Although the morphological changes of neural tube closure have been well studied, the genetic factors of NTDs remain unclear. Most NTDs occur sporadically. Nevertheless, NTDs have multifactorial and polygenic root causes[56]. Instead of focusing on one gene, most NTDs genetic studies have explored multiple candidate genes, including their modifier genes, epigenetic-related genes, and environment-related genes.

The folate supplementation has been proved its efficiency in the prevention of NTDs. Folic acid is highly related to the folate cycle of one-carbon metabolism, which is a complex network involving several biosynthesis processes, including nucleotide synthesis and methylation reactions. The primary focus of NTDs studies has been on genes related to this network. For example, knockout of *Mthfd1S* gene, that involved in one-carbon metabolism, has been proved lethal before the neural tube closure stage in mice[57]. In addition, the mutations in folate receptor protein (encoded by *Folr1*) causes null embryos in mice and leads to NTDs even supplemented with sufficient folic acids in the diet. Furthermore, the homocysteine

remethylation gene (*MTHFR*), which has two genetic polymorphisms, is highly related to NTDs in the Hispanics population[58]. Aberrant thymidylate and purine synthesis has also been implicated in mouse NTDs models[59] and several NTD patients[60]. Besides, some genes related to glucose metabolism, such as *Pax3*, have also been studied to determine the culprit genes in diabetes-related NTDs[55].

Several signaling pathways, such as the planar cell polarity (PCP) signaling pathway, the *Sonic hedgehog* (Shh) signaling pathway, and the BMP signaling pathway have been implicated in NTDs. For example, mutations of genes related to the PCP pathway, such as *DVL2*, *VANGL1*, *VANGL2*, and *FZD6*, have already been confirmed as risk factors of NTDs in humans. Dysregulation of the *Shh* signaling genes, such as *Fkbp8* and *Gli2*, has also been reported to associate with certain types of NTDs[46].

1.3.3. Prevention of NTDs

Once the closure of the neural tube has failed, the following damage to the exposed neural tissue is permanent. Thus, the prevention before and during pregnancy is an optimal approach to reduce the risk of NTDs. For open spinal bifida fetuses, some challenging *in utero* surgeries have been practiced to protect the neural tissue from degeneration.

According to recent public health recommendations, all preconceptions and pregnant women should consume 1.0 mg folic acid daily, in which 0.2 – 0.3 mg through diet, and 0.7 – 0.8 mg by supplementation[61]. However, folic acid supplementation cannot prevent all NTDs. Approximately one-third of NTDs maybe folic acid-resistant[62]. Since NTDs have a multifactorial causation, the optimal prevention would also require a combination of multiple interventions. In addition to folic acid, vitamin B12 and vitamin C may help to reduce the risk of NTDs as well[63,64]. In mouse models, formate, 5-methylTHF, and thymidine/purine precursors are also used as remedies for NTDs prevention[65,66,67].

1.4. Protein ubiquitination and degradation

Since CUL3 is one of the potential binding partners of ARMC5 as indicated in Y2H results, they may corporately serve as a ubiquitin ligase and catalyze the ubiquitination of other proteins. Thus, I would like to introduce the ubiquitin-proteasome system (UPS) in the following sections.

1.4.1. The general aspects of the ubiquitin system

Ubiquitin (Ub) is a short protein containing 76 amino acids, which is highly conserved among all eukaryotes. In most cases, it is synthesized either as polyubiquitin cassettes (encoded by *UBB* and *UBC*)[68] or as a single Ub fused to the ribosomal proteins L40 and S27a (encodes by *UBA52* and *RPS27*, respectively)[69,70]. The monomeric Ub is cleaved off by deubiquitinases (DUBs) from these fusion proteins.

Ubiquitination is catalyzed by a three-step enzymatic cascade, that comprised of ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and ubiquitin ligases (E3s). It starts with the activation of Ub by E1s in an ATP-dependent manner to generate a thioester bond between the E1 and the C-terminus of Ub[71]. Secondly, E1 transfers the activated Ub to an E2 via a transthiolation reaction. Finally, an E3 ligase, which interacts with both the Ub-loaded E2 and a substrate, most commonly either forms an iso-peptide bond at the C-terminus of Ub onto the lysine site of the substrate or forms a peptide bond with the amino terminus of the substrate. The schematic diagram is shown in **Figure 1.5**.

1.4.2. E1s

There are two E1s (UBE1 and UBA6) for activating Ub in metazoan[72,73]. UBE1 (Uba1 protein in yeasts) is the major one. In proliferating cells, UBE1 is fully charged with Ub, while UBA6 is about half charged under a similar condition[73].

Compared to UBE1, UBA6 is a more specific one. It can only transfer Ub to certain E2s, such as UBC5 and UBC13[74]. It also activates ubiquitin-like (UbL) proteins (e.g.,
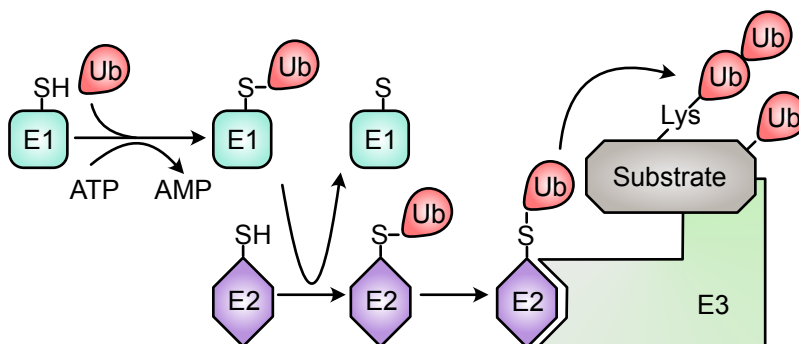
Fig. 1.5. The ubiquitin cascade Ubiquitin (Ub) carried by an ubiquitin-activating enzyme (E1) is transferred to an ubiquitin-conjugating enzymes (E2) and added to the substrate by an ubiquitin ligase (E3s) via a three-step enzymatic cascade reaction.

FAT10)[75,73,74]. Meanwhile, a special E2 (USE1) can only be charged by UBA6[73]. This phenomenon indicates that UBA6 may be specialized in the ubiquitination of a particular group of proteins.

1.4.3. E2s

There are about 40 E2s existing in humans. An activated E2 contains a core Ub-conjugating domain, which carries a catalytic Cys residue. The Ub-conjugating domains from different E2s share a similar structure consisting of four α-helices, one antiparallel β-sheet, and one short 3_{10}-helix[76]. The job of E2s is transferring Ub from E1s to E3s. They are often considered as a simple role as "Ub carriers". However, recent studies have shown that they played a more active role in determining the length and topology of Ub chains and selecting E3s, which would finally influence the outcome of ubiquitination.

The first important task of E2s is receiving Ub from E1s. E2s have a significantly increased affinity to E1s that carrying Ub[71]. Sequentially, the L1/L2 loops and the α-helix-1 on the surface of E2 bind to specific E3s. One E2 is able to match with several E3s. For example, the UBE2D family, including UBE2D1, UBE2D2, and UBE2D3 can transfer Ub to many E3s[77]. It would be good to identify all the physiological E2-E3 pairs, however, the weak and transient interactions between E2s and E3s make this job challenging.

Different E2s are in-charge of the diverse tasks. Certain E2s are only responsible for adding the first Ub on the substrate, such as the UBE2D family. The elongation of Ub

chains is often controlled by some other E2s, which could decide the position where the next Ub is added on. Thus, they can control the types of polyubiquitin chains. For example, UBE2N-UBE2V1 is a K63-specific chain-elongating E2, and UBE2S is the one for K11 polyubiquitin chains[78,79].

In a nutshell, E2s act as the bridge to transfer the correct E1 to a suitable E3. It helps to determine the length and topology of Ub chains, finally affecting the consequence of ubiquitination.

1.4.4. E3s

The human genome encodes ~700 E3s, according to the bioinformatic analysis. The accuracy of this estimation is not certain, because we do not know about the characteristics of the sequences in all possible E3s. The function and substrates of most putative E3s remain unknown.

The E3s are critical parts of the ubiquitin system owing to their specificity of recruiting substrates and determining the types of ubiquitination. They have been classified into three main families, Homologous to the E6-AP Carboxyl Terminus (HECT)-type E3s, Really Interesting New Gene (RING)-type E3s, and RING-between-RINGs (RBR)-type E3s, by their distinguished structures and the Ub-transferring mechanism. A schematic diagram of the Ub-transferring patterns of three E3 families is shown in **Figure 1.6**. The mechanisms by which different E3s target substrates vary. Some E3s contain domains directly binding to substrates. Some E3s form complexes with other adaptor proteins or non-protein molecules, which help to recruit substrates.

1.4.4.1. HECT E3s.

There are 28 known HECT E3s in humans[80]. In general, they are characterized by a bi-lobar structure. The HECT domain with catalysis function is located at the C-terminal lobe, whereas some other domain like the tryptophan-tryptophan (WW) motif in NEDD4 E3 or the chromosome condensation 1 (RCC1)-like domain (RLD) in HERC E3 family is located at the N-terminal lobe. The N-terminal lobe can interact with a Ub-loaded E2. Two lobes

are linked with a flexible hinge that allows adjusting the orientation and distance between Ub carried by the E2 and the ubiquitination site of the substrate. HECT E3s catalyze Ub transfer through a two-step reaction. Ub is first transferred by E2 to a catalytic site on the HECT E3, then from the E3 to the substrate.

1.4.4.2. RING E3s.

The bioinformatic analysis predicts that there are more than 600 RING E3s in humans[81]. They are characterized by containing a zinc-binding domain (a RING finger or a U-box domain). A RING finger is a small domain that coordinates two zinc ions, while a U-box domain is similar but without zinc-binding function. The activated RING domain or U-box domain interacts with a Ub-loaded E2 and stimulates the E2 to directly transfer Ub to the substrate. RING E3s can function as monomers, homodimers, or heterodimers. Some RING E3s consist of several subunits, such as Culling-RING ligases (CRLs) and the anaphase-promoting complex/cyclosome (APC/C) E3.

1.4.4.3. RBR E3s.

There are 14 RBRs identified in humans up to date[80]. RBR E3s are composed of two RING domains (RING1 and RING2), which are separated by an in-between-RING (IBR) domain. Well-known E3 ligases, such as PARKIN, HHARI, and HOIP belong to this family. The Ub transfer mechanism is similar to HECT E3s and occurs in a two-step transfer manner. Firstly, the RING1 domain recruits the Ub-loaded E2 and then transfers the Ub to the catalytic RING2 domain. Secondly, the RING2 domain deliver the Ub to the substrate.

1.4.5. The Cullin-RING ligases (CRLs)

Since CUL3 is the potential binding partner of ARMC5, a more detailed literature review of CRLs is presented in this section.

1.4.5.1. The components of CRLs.

CRLs is a highly diverse family sharing some common features. A Cullin protein (CULLIN) serves as a scaffold unit to assemble multiple other units, such as a RING domain
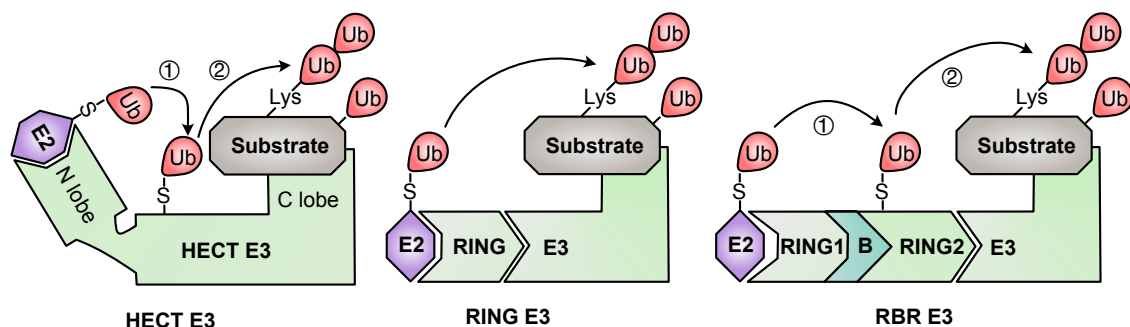
Fig. 1.6. The ubiquitin transferring patterns in HECT E3s, RING E3s, and RBR E3s (Left). HECT E3 first transfers Ub from E2 to a catalytic cysteine site on its main body, then from itself to the recruited substrate. (Middle). RING E3 directly transfers Ub to the substrate without contacting itself. (Right). The RING1 domain of RBR E3 transfers Ub from E2 to the RING2 domain, then to the substrate.

protein (RBX1 or RBX2) and diverse adaptor proteins that recruit substrates. The adaptor module can be composed of one protein or several proteins.

There is a key lysine residue at the C-terminus of the Cullin protein, which can be modified by NEDD8 protein. This modification, termed as neddylation, is essential for CRLs activation[82]. The neddylation refers to sequential enzymatic actions similar to ubiquitination, employing a NEDD8 activating enzyme, a NEDD8-conjugating enzyme, and a NEDD8 E3 ligase, to transfer NEDD8 onto the Cullin proteins.

CAND1 binds to Cullin protein when the E3 is under resting status. It acts as an inhibitor and blocks the binding site for the adaptor modules[83,84]. The Cullin protein neddylation disrupts its binding to CAND1, thus, activates the E3. Besides, neddylation causes the conformation change of CRL, reduces the distance from E2 to the specific residues on the substrate, and finally starts the ubiquitination process[85]. The general components, activation and inhibition of the CRLs model is shown in **Figure 1.7**.

1.4.5.2. The Cullin family.

There are seven Cullin proteins (CUL1, 2, 3, 4A, 4B, 5, and 7) in humans. The CRLs are divided into six families based on different Cullin scaffold proteins: CRL1s, CRL2s, CRL3s, CRL4s, CRL5s, and CRL7s.
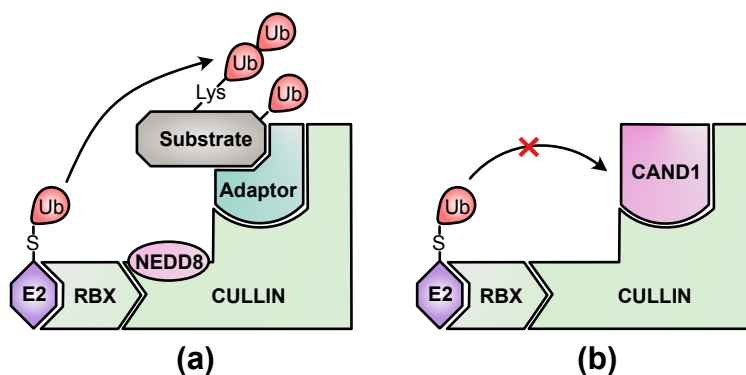
Fig. 1.7. The general components of Cullin-RING ligases (CRLs) model (a). A CULLIN protein serves as a scaffold protein. An RBX protein recruits an Ub-loaded E2, while an adaptor protein or complex recruits the substrate to CULLIN. The activity of CRLs requires NEDD8 modification on the C-terminus of CULLIN. **(b).** CAND1 acts as an inhibitor and blocks the binding site for adaptor modules in CULLIN. NEDD8 modification can dissociate CAND1 from CULLIN, and activate CRLs.

The adaptor protein for CUL1 and CUL7 is SKP1. However, SKP1 does not contain the substrate interaction motif. It binds to the F-box protein to cooperatively serve as an adaptor module to recruit the substrate (**Figure 1.8a and f**)[86,87]. There are about 69 F-box proteins in humans, which can be divided into three categories according to the substrate interaction domain: WD40 domains (FBXW family), leucine-rich repeats (FBXL family), and other diverse domains (FBXO family)[88]. In addition to the substrate recognition domain in F-box proteins, they all contain a F-box domain, which is responsible for the binding to SKP1. CUL1-SKP1 binds to most of these F-box proteins, thus constituting diverse E3s, while CUL7-SKP1 forms two E3s by binding to FBXW8 or FBXW11 protein[89].

Elongin B and Elongin C cooperatively act as a similar role of SKP1 in CUL2 and CUL5-based E3s. They do not contain the substrate interaction motif either. In CUL2-based E3, Elongin B/Elongin C binds to pVHL, which is able to recognize the substrates, such as HIF-1α[90]. In CUL5-based E3, they bind to SOCS protein to carry out the same function (**Figure 1.8b and e**). For CUL4A- and CUL4B-based E3s, DDB1 binds to the DCAF family and acts as the substrate recognizers (**Figure 1.8d**)[91]. CUL4A/B-based E3s have been implicated in the protein ubiquitination in response to DNA damage[92,93].

The adaptor modules in these CRLs, such as the SKP1/F-box complex and Elongin B/Elongin C/pVHL complex, are formed by several proteins. However, in CUL3-based E3s, BTB domain-containing protein by itself alone is able to serves both roles of binding to the Cullin protein and recruiting the substrate (**Figure 1.8c**). BTB domain-containing proteins own various interaction motifs, which makes CUL3-based E3s to be the largest E3 family in CRLs.
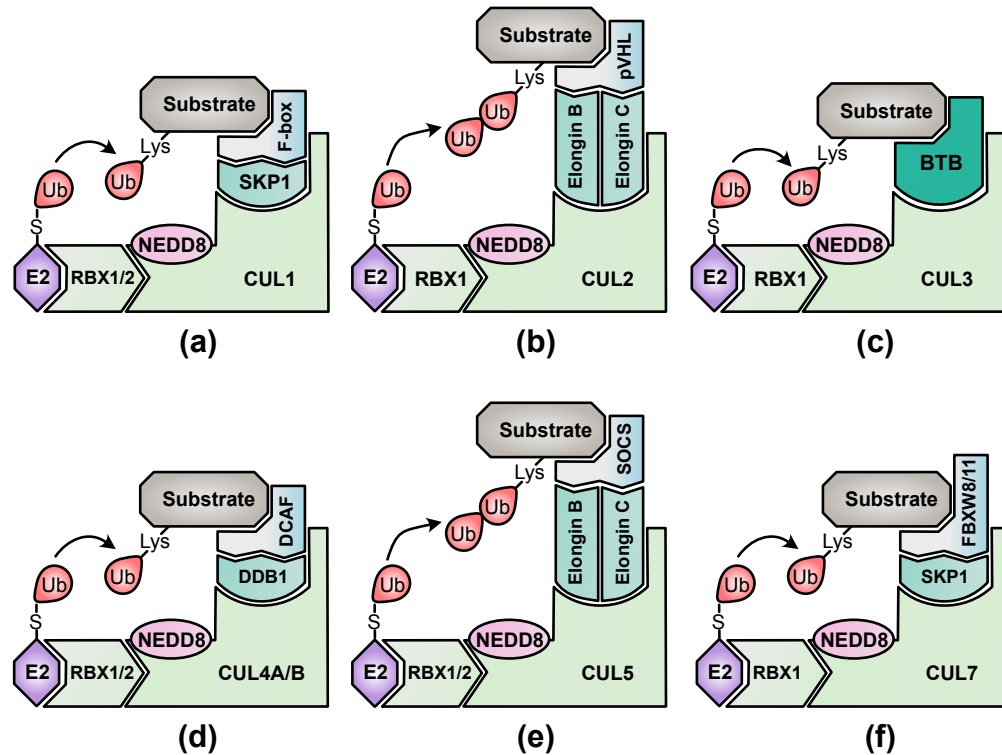
Fig. 1.8. The models of Cullin-RING ligase (CRL) family (a). The SKP1/F-box proteins act as the adaptor modules in CUL1-based E3s. **(b).** The Elongin B-Elongin C-pVHL complex serve as an adaptor module in CUL2-based E3. **(c).** The BTB domain-containing proteins serve as an adaptor in CUL3-based E3s. **(d).** The DDB1-DCAF complex acts as the adaptor modules in CUL4A/4B E3s. **(e).** The Elongin B-Elongin C-SOSC complex serves as an adaptor module in CUL5-based E3s. **(f).** The SKP1-FBXW8 or SKP1-FBXW11 complex acts as an adaptor module in CUL7-based E3s.

1.4.6. Cullin3-RING ubiquitin ligase (CRL3)

CUL3-based E3s play fundamental roles in various biological process. It has been reported that deletion of *Cul3* caused early embryonic lethality in mice[94]. Depending on the

different adaptor proteins and various substrates, CUL3-based E3s are involved in many physiological and pathological processes, including cell development[95], anti-oxidative responses[96], blood pressure control[97], cell proliferation[98], cell apoptosis[99], tumorigenesis[100] and autism[101].

CUL3 interacts with the BTB domain-containing proteins to catalyze E3 function. The human genome encodes ∼200 BTB domain-containing proteins[102]. Only a few of them have been characterized as substrate recognizers up to date. Many BTB domain-containing proteins with unknown function remain unexplored.

BTB domain-containing proteins commonly have other modules for protein-protein interactions, other than BTB domain. According to the shared common domain architecture, they can be classified as the Kelch-like (KLHL)/Kelch repeat-BTB (KBTBD) family (shared Kelch domain), the zinc finger-BTB (ZBTB) family (shared zinc finger domain), the MATH-BTB family (shared MATH domain), the potassium channel tetramerization domain (KCTD)-BTB family (shared potassium channel domain), the Rho-BTB family (shared Rho domain), the Ankyrin-BTB family (shared Ankyrin domain), and the BTB-only family[103]. In addition, interacting domains, such as Armadillo repeat and Pent domain, form the unnamed family of BTB domain-containing proteins. ARMC5 is such a protein containing seven Armadillo repeats and one BTB domain.

Moreover, BTB domain-containing proteins are often dimerized through the BTB domains, which leads to the dimerization or polymerization of the CUL3-based E3s. The structure analysis of human CUL3-KEAP1 E3 showed that two KEAP1 formed a homodimer to bind one NRF2 substrate at two different interaction sites to get an optimal orientation for ubiquitination[100].

CUL3 and other essential subunits (such as RBX1 and NEDD8) that constitute E3 complexes are widely expressed in different tissues and remain unchanged during cell differentiation[104]. However, BTB domain-containing proteins exhibit distinctive expression levels in different cell types and tissues. For example, KLHL10 plays an important role in spermatogenesis[105], while KCTD13 is a major driver of neural developmental phenotypes[106]. For

ARMC5, it is highly expressed in the adrenal glands, thymus, stomach, bone marrow, and lymphoid tissues[8]. ARMC5 may play a more critical role in these tissues than other ones with lower expression levels.

1.4.7. Deubiquitinating enzymes (DUBs)

DUBs are proteases that cleave the bonds between two Ub molecules or between Ub and a ubiquitinated protein. There are seven families with different catalytic domains: ubiquitin-specific proteases (USPs), ovarian tumor proteases (OTUs), ubiquitin C-terminal hydrolases (UCHs), Josephins (MJDs), motif interacting with ubiquitin-containing novel DUB family (MINDYs), Zinc finger-containing ubiquitin peptidase 1 (ZUP1), and the Jab1/MPN/MOV34 (JAMM) family. The first six families are cysteine proteases, and the last one is zinc-dependent metalloproteinases[107].

In addition to the catalytic domain, DUBs are distinguished by other domains, such as UBL, DUSP, and CAP[108]. These domains guide DUBs to their specific substrates. Some DUBs also recognize specific polyUb chain architectures and may not be able to remove all attached Ub molecules. For example, OTUB1 DUB, OTUD4 DUB, and A20 DUB prefer K48-linkage polyUb chains, while cylindromatosis (CYLD) DUB, AMSH DUB, and BRCC6 DUB are more specific for K63-linkage chains[107].

DUBs have key roles in maintaining protein homeostasis and regulating signaling. As mentioned, Ub is synthesized either as polyUb cassettes[68] or as a single Ub fused to L40 and S27a[69,70]. These newly synthesized Ub molecules are released as single ones by DUBs. In mammalian cells, more than half of Ub molecules are monomeric ones and are conjugated to lysine residues of the substrates. A further $10 - 20\%$ of them form chains, while the rest of them exist as free ones[109]. By removing Ub from the attached protein, DUBs can directly regulate some signaling complexes and recycle Ub to maintain the free Ub pool. More importantly, DUBs can rescue the ubiquitinated proteins from the proteasomal degradation.

1.4.8. Proteasomal recognition of ubiquitinated proteins

Most of the ubiquitinated proteins are delivered to the proteasome for degradation. The proteasome is a self-compartmentalized protease with proteolytic activity. The complete process requires the cooperation of all proteasome subunits.

The 26S proteasome is a huge complex consisting of a 20S core particle (CP) and one or two 19S regulatory particles (RPs). The 20S CP can be capped by 19S RPs at one end or both ends. The CP is barrel-shaped, composed of two inner β catalytic rings ($\beta 1 - 7$) and two outer α rings ($\alpha 1 - 7$). The narrow central channel of the α rings acts as an entrance for degrading proteins. However, when CP is alone without RP and other activators, this channel is closed. Only the fully assembled proteasome owns an activated entrance.

The RP, which consists of 6 ATPase subunits (RPT1 – 6) and 13 non-ATPase subunits (RPN1 – 3, 5 – 13, and 15), plays an important role in navigating the protein to proteasomal degradation. The RP is biochemically divided into two parts, "base" and "lid". The "base" consists of the ATPase ring (RPT1 – 6), two homologous subunits (RPN1 and RPN2), and Ub receptor units (RPN10 and RPN13). The "lid" is composed of 9 non-ATPase subunits (RPN3, 5 – 9, 11 – 12, and 15), which surrounds the ATPase ring[110]. The structure and key subunits of the 26S proteasome are shown in **Figure 1.9**.

When the ubiquitinated substrates are delivered to the proteasome, RPN1, RPN10, and RPN13, those located near the periphery of PR serve as Ub receptors and capture the ubiquitinated targets. Next, the Ub chains are removed by RPN11 in the "lid", which is a metalloprotease DUB. The substrate will be unfolded by ATPase at the "base" and translocated to CP for destruction[111]. Within CP, the substrate is attacked by three distinct proteolytic activities (chymotryptic digestion, tryptic digestion, and caspase-like digestion) and cleaved into peptides around 4 – 25 amino acids in length[110].

Taken together, the proteasome is a complex carrying various biochemical activities (e.g., the recognition of ubiquitinated substrates, deubiquitination, protein unfolding and translocation, protein cleavage, and protein destruction).
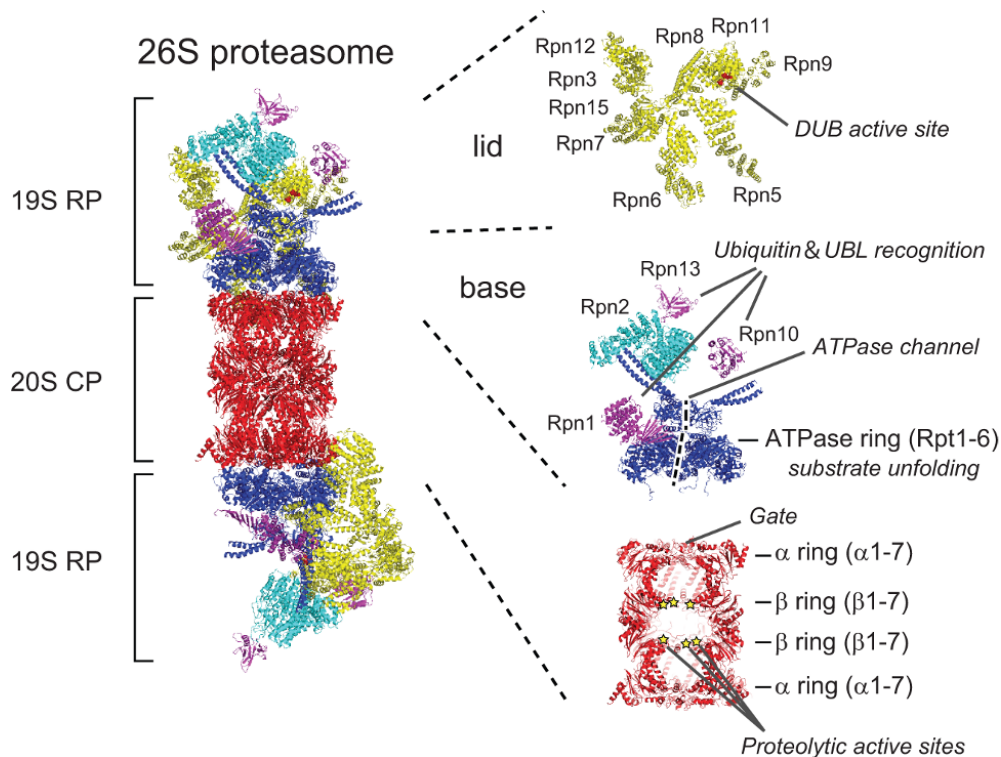
Fig. 1.9. The structure and key subunits of the 26S proteasome The 26S proteasome consists of a 20S core particle (CP) and one or two 19S regulatory particles (RPs). The CP is composed of two inner β catalytic rings and two outer α rings. The RP is biochemically divided into two parts, the "base" and the "lid". The "base" contains ATPase ring, ATPase channel, and ubiquitin recognition units. The "lid" is composed of 9 non-ATPase subunits. The figure is adapted from the publication of Saeki[111].

1.4.9. Ubiquitin codes

Protein ubiquitination not only directs the substrate to the degradation machinery, but also can regulate their functions. The signaling is controlled by diverse ubiquitin codes.

Ub contains seven lysine sites in amino acid position 6, 11, 27, 29, 33, 48, and 63 (**Figure 1.10**). A polyUb chain is formed by linking Ub molecules one to another via an iso-peptide binding between one of the lysine sites of acceptor Ub and the C-terminal position 76 glycine (G76) site of donor Ub. Ub molecules are also able to conjugate to one another by a head-to-tail pattern (G76 residue of donor Ub binds to the N-terminal Methionine site [M1] of the acceptor Ub), thus forming M1-linked (i.e., linear) polyUb chains.
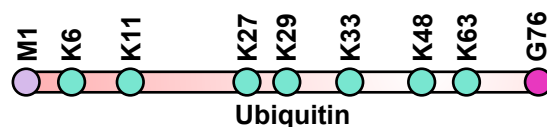
Fig. 1.10. The conjugating sites in ubiquitin molecule Ubiquitin has seven lysine (K) sites in amino acid (AA) position 6, 11, 27, 29, 33, 48, and 63. The first AA is Methionine (M) and the last one is glycine (G) in AA position 76.

Based on the number of Ub molecules added to the substrate and the conjugating position of each Ub within the chain, the complexity of ubiquitin codes is unimaginable. It can be roughly divided into two categories: mono-ubiquitination and poly-ubiquitination.

1.4.9.1. Mono-ubiquitination.

Mono-ubiquitination means adding a single Ub through the bond at G76 to the substrate. It can occur at one residue (**Figure 1.11a**) or multiple residues of one substrate (**Figure 1.11b**). One classic example is the mono-ubiquitination of histone H2A/H2B. In mammals, histone H2A is mono-ubiquitinated by BMI1-RING1 E3 at lysine 119[112]. Histone H2B mono-ubiquitination occurs at lysine 120, written by RAD6A/RAD6B and RNF20/RNF40 E3s[113,114]. The mono-ubiquitination of histones plays a pivotal role in DNA damage response, especially DNA double-strand break repair mechanism[115].

1.4.9.2. Poly-ubiquitination.

A polyUb chain can be formed by one Ub molecule adding to another. They are normally attached to the substrates, but some of them also exist as an unanchored form (**Figure 1.11h**)[116]. The types of polyUb chains are determined by the conjugating position in Ub molecules. Most chains are synthesized homogeneously. One chain only contains one type of linkage, which can occur only at one site (**Figure 1.11c**) or at multiple sites with different types of chains (**Figure 1.11d**) of the substrate.

In addition, mass spectrometry analysis indicates that there are various of heterogeneous chains existing in cells[117], which contain at least two types of linkages. The heterogeneous chains are divided into two categories, the mixed chains and the branched chains. In the mixed chains, Ub molecules are connected with different linkages, and each Ub only contains one linkage (**Figure 1.11e**). However, if a single Ub contains two or more linkages, a

branched chain is generated from that Ub (**Figure 1.11f**). Furthermore, recent studies indicate Ub itself can be phosphorylated or acetylated (**Figure 1.11g**)[118,119]. This highly increases the diversity of the Ub codes.
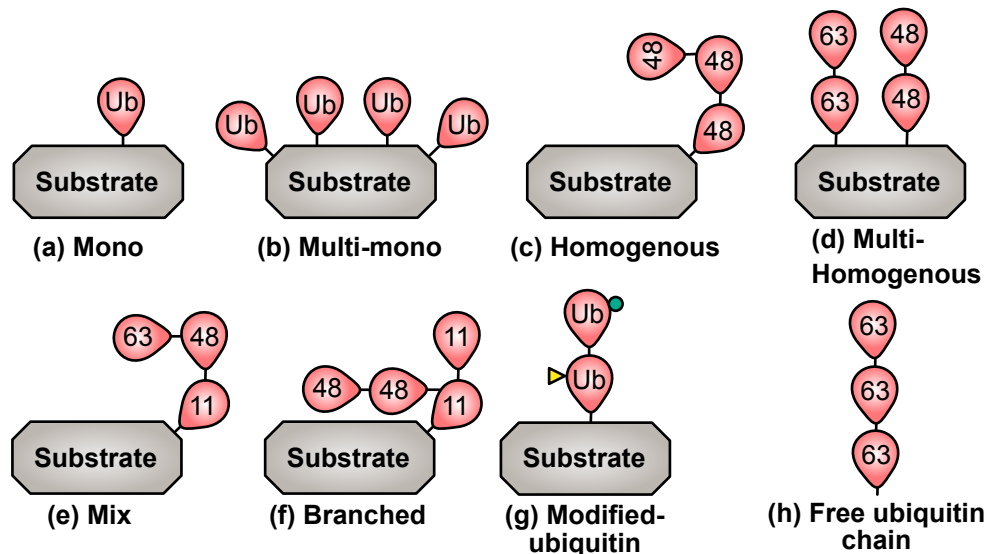
Fig. 1.11. The Ubiquitin codes (a). Mono-ubiquitination; **(b).** Mono-ubiquitination at multiple residues; **(c).** Homogenous poly-ubiquitination; **(d).** Homogenous poly-ubiquitination with different linkages; **(e).** Heterogeneous poly-ubiquitination with mixed linkages; **(f).** Heterogeneous poly-ubiquitination with branched chains; **(g).** Acetylated or phosphorylated ubiquitin; **(h).** unanchored ubiquitin chains. Yellow triangle: phosphorylation; Green circle: Acetylation.

1.4.9.3. The "canonical" ubiquitin chains.

According to the analysis of the cellular Ub pool using the protein standard absolute quantification (PSAQ) method, about 23% of Ub molecules present as free Ub, ∼65% of them exists as mono-ubiquitinated ones, and approximately 11% of them are polyUb chains in HEK293 cells, as well as in mouse embryonic fibroblast cells (MEFs)[109]. In human frontal cortex tissue, free Ub takes account for 82%, about 14% of Ub molecules are mono-ubiquitinated ones, and polyubiquitin chains take the rest 4%[109]. These studies suggest only a few of Ub molecules existing as the form of polyUb chain.

Among the polyUb chains, K48-linkage is the most abundant type, which accounts for ∼75% of all linkages in yeast cells. K29-linkages and K63-linkage share ∼13% and ∼8%,

respectively. A small minority (∼2.78%) is K11-linkage. Each of the rest linkages (K6-, K27-, K33-, and M1-polyUb chain) constitutes less than 2% of all polyUb chains[120].

K48-linkage and K63-linkage are considered as the "canonical" poly-ubiquitination types, as they were the well studied ones when the research of this field began[121]. K48-linkage is believed to be the major signal for proteasome degradation. It was discovered in the degradation of short-lived proteins[122]. Now it is known to be involved in the degradation signaling of most proteins, such as p53[123]. CRLs, HUWE1 E3, and APC/C E3 are identified as K48-linkage-related E3s[124,125,126]. Originally, an *in vitro* reconstituted system has proved that the proteasome recognized K48-linked tetra-Ub chain as a minimal signal and the efficiency markedly increased when K48-linked octa-Ub chains were added[127]. However, compared to the length of each Ub chain, recent studies suggest that the amount of K48-linked chains added to the substrates may be more important than the number of ubiquitin in a single chain[128]. Lu et al. demonstrated that two di-Ub modifications added on two lysine sites of cyclin B1 protein had a more efficient proteasomal degradation signaling than a single tetra-Ub added on one lysine site[128].

K48-linked Ub chains are recognized by RPN10 and RPN13 subunits of the 26S proteasome. Deletion of the ubiquitin-binding domains (UBDs) in Rpn10 and Rpn13 proteins in the yeast proteasome reduces its affinity to K48-linked chains. However, some K48-linked chains could still bind to the proteasome in the *RPN10/RPN13* double mutant yeast cells, suggesting the receptors of K48-linked chains is not limited to RPN10 and RPN13 in proteasome[129].

Since K48-linkage is the major degradation signal to the proteasome, does K48-linkages selectively bind to UBDs in the proteasome, compared to other types of Ub chains, especially K63-linkage? The answer is probably not. In a reconstituted proteasome system, K63-linkages showed a similar affinity to the proteasome as K48-linkages[130,129,131]. However, some ubiquitin-like/ubiquitin-associated proteins (UbL/UbA), such as Rad23 in yeasts (HHR23A in humans), did show a selective affinity for K48-linkages than K63-linkages[132,133]. Therefore, it can be speculated that some other proteins, such as Rad23, play roles in selectively

recognizing K48-linkages and directing K48-polyubiquitinated substrates to the proteasomal degradation pathway.

Rather than triggering degradation, K63-linked polyUb chains serve as signals that allow rapid and reversible modification of many signaling complexes. It has been reported that K63-linkages participated in regulating NF-κB activation[134,135], Akt kinase activation[136], innate immune responses[137], clearance of damaged mitochondria[138], DNA damage repair[139] and oxidative stress responses[140]. Several K63-linkage-specific E3s have been identified, such as TRAF2, TRAF6,[141,142], UCHL1[143], ITCH[144], and NEDD4L[145].

K63-linked chains are even able to perform functions without attaching substrates[146,116]. The unanchored K63-linked chains synthesized by TRAF6 E3 and UBCH5C E2 can directly activate the TAK1-IκK signaling by binding to the Ub receptor TAB2. Disassembly of K63-linked chains by a K63-linkage-specific DUB (CYLD) terminates the activation of TAK1 and IκK activation[146]. In addition, the unanchored K63-linked Ub chains can activate RIG-1, which is important for detecting invading viral RNA[116].

1.4.9.4. The "non-canonical" ubiquitin chains.

The M1-linked chains are produced by the linear Ub chain assembly complex (LUBAC), which has been reported in the NF-κB signaling pathway[147,148]. LUBAC is an RBR E3 member consisting of HOIP, HOIL-1L, and Sharpin. OTULIN protein is a known DUB that exclusively removes linear polyUb chains. OTULIN binds to M1-linked polyUb chains with high affinity and catalyzes the process with the assistance of substrate[149]. Therefore, the deficiency of LUBAC and OTULIN causes several M1-linkage-related phenotypes, including embryonic lethality, impaired NF-κB signaling and Wnt signaling, vascularization defects, and chronic proliferative dermatitis[150,151,152].

It has been reported that the majority of M1-linked chains are covalently attached to the K63-linked chains upon IL-1 stimulation[153]. In this study, if K63-linkage was inhibited by deleting the specific K63-linkage E2 (Ubc13 in yeasts), correspondingly, IL-1 induced M1-linkage chains also reduced, suggesting that the M1-linked polyUb were added upon K63-linked chains[153]. Other than IL-1 signaling, M1/K63 heterogeneous chains are also

implicated during the activation of tumor necrosis factor receptor 1 (TNFR1), indicating the role of hybrid chains in the innate immune signaling[154]. It has been proved that the decoration of K63-linkages with M1-linked polyUb prevents the deubiquitination by K63-linkage specific DUBs, thereby stabilize the signaling pathway[155].

K11-linked polyUb chains are known to be assembled by the APC/C E3 and its specific E2 UBE2S[156,157,158]. The abundance of K11 linkage significantly increases when APC/C E3 is activated during mitosis[159]. UBE2S does not simply extend K11-linked chains, but branches other linkages on the Ub molecules[160]. Recent studies indicate that, compared to homogenous K11- or K48-chains, K11/K48-branched chains significantly enhance the substrate recognition by the proteasome[160]. Branched K11/K48 tri-Ub forms a hydrophobic interface, which increases the affinity to RPN1 in the 26S proteasome. K11/K48-branched chain has been implicated in cell cycle regulation and protein quality control[161,162].

In addition to K48/K11-linkage, other K48-related branched chains, such as K29/K48-[163] and K48/K63-linkages[164], can channel substrates to efficient proteasome degradation as well. K48/K63-branched chains also play a role in regulating the signaling pathway. In the NF-κB pathway, K48/K63-branched chains, that were synthesized by TRAF E3 (K63-linkage) and HUWE1 E3 (K48-linkage), amplified the signaling by protecting K63-linkages from DUBs in response to IL-1β stimulation[165].

The world of K27-linkage, K29-linkage, and K33-linkage are less explored yet. K27-linked ubiquitination is a major chromatin marker of DNA damage. RNF168 catalyzes K27-linked ubiquitination of H2A upon DNA damage, which helps to recruit other response mediators to the chromatin[166]. Smurf1 E3 ubiquitinates Axin through K29-linked chains. The K29-linked ubiquitination of Axin does not lead to proteasome degradation. Instead, it disrupts the interaction between Axin and Wnt coreceptor LRP5/6, which subsequently represses the Wnt/β-catenin signaling[167]. K33-linked ubiquitination has been identified in T cell receptor-ζ[168] and AMP-activated protein kinase[169]. It serves as a non-proteolytic signal regulating receptor-mediated signal transduction. In addition, K33-linked ubiquitination has also been implicated in post-Golgi protein trafficking[170].

1.5. RNA polymerase II

RNA polymerase II subunit I (RPB1) protein has been indicated as one of the potential binding partners of ARMC5 in Y2H results. A brief review of RNA polymerases is provided in this section.

1.5.1. DNA-dependent RNA polymerases

DNA-dependent RNA polymerases (RNAPs) are enzymes that synthesize RNA from template DNA. RNA polymerase is essential to all organisms. In different organisms, an RNA polymerase can be a single protein or a huge protein complex. The single protein polymerase can be found in phages as well as eukaryotic chloroplasts and mitochondria. In eukaryotes, bacteria, and archaea, RNAPs are multi-unit protein complexes, sharing a similar structure and mechanism. Bacteria and archaea only have one kind of RNAP, while eukaryotes contain multiple types of RNAPs.

In eukaryotes, RNA polymerase I (Pol I) synthesizes pre-ribosomal RNA (pre-rRNA). RNA polymerase II (Pol II) is responsible for transcribing all precursors of messenger RNA (pre-mRNA) and most small nuclear RNA (snRNA) and microRNA (miRNA). RNA polymerase III (Pol III) transcribes 5S rRNA, transfer RNA (tRNA), and other small RNA. RNA polymerase IV and V are found in plants that synthesize some siRNA.

1.5.2. The assembly of Pol II

After the structure of Pol II was first resolved by Dr. Roger D. Kornberg's lab[171,172], whose work was awarded the 2006 Nobel Prize in Chemistry, structure biologists are continuously exploring the detailed architecture of Pol II-related complexes at different transcription stages[173,174,175,176,177,178,179].

Pol II is a 550 kDa complex of 12 subunits, which are named from RPB1 to RPB12. The Pol II core is composed of three subassemblies: RPB1 subassembly (RPB1, RPB5, RPB6, and RPB8), RPB2 subassembly (RPB2 and RPB9), and RPB3 subassembly (RPB3, RPB10, RPB12, and RPB11). The RPB1 and RPB2 subassemblies form the active core

cleft. The distinct RPB3 subassembly bridges RPB1 and RPB2. In addition, RPB4 and RPB7 form a stalk protruding from the surface of RPB1. The structure of Pol II is illustrated in **Figure 1.12**.
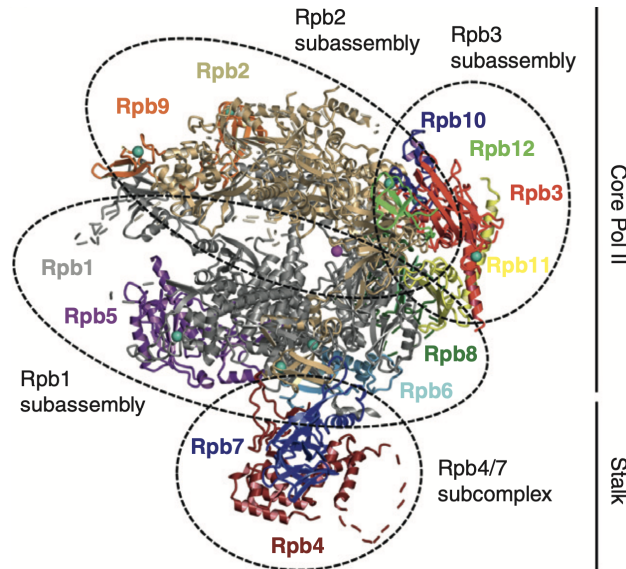
Fig. 1.12. The components and structure of RNA polymerase II The Pol II core is composed of RPB1 subassembly (RPB1, RPB5, RPB6 and RPB8), RPB2 subassembly (RPB2 and RPB9), and RPB3 subassembly (RPB3, RPB10, RPB12, and RPB11). The RPB1 and RPB2 subassemblies are connected by the RPB3 subassembly and form the active core cleft. RPB4 and RPB7 form a stalk protruding from the surface of RPB1. The figure is adapted from the publication of Wild and Cramer[180].

1.5.3. The synthesis, assembly, and transportation of Pol II

Although the structure and function of Pol II have been deeply investigated, the mechanisms of Pol II synthesis, assembly, and transportation from the cytosol to the nucleus remain unclear. Several recent studies have revealed part of the processes. All Pol II subunits are synthesized in the cytoplasm. Deletion of any Pol II subunit lead to the accumulation of RPB1 in the cytosol, suggesting that Pol II requires to be fully assembled before entering the nucleus[181]. Furthermore, it has been proven that many assembly factors such as R2TP/heat shock protein 90 (HSP90) co-chaperone complex, RNA polymerase II associated protein 1 (RPAP1), RPAP2, GPN-loop GTPase 1 (GPN1), GPN2, GPN3, and GrinL1a, have participated in the Pol II assembling process[181,182,183,184]. Depleting of GPN1 or GPN3 also leads

to the cytoplasmic accumulation of RPB1[183,182,185], indicating the important roles of GPN1 and GPN3 in mediating the nuclear import of Pol II.

However, neither the subunits of Pol II nor GPN proteins contain a nuclear localization signal (NLS). Likely, other factor containing NLS provides the importing signal. IWR1 is the one identified carrying a nuclear importing signal for Pol II[186], as it also binds to the center cleft of Pol II. It mediates Pol II binding to importin-α/β and facilitates the transportation through the nuclear pore complex (NPC). The deletion of *IWR1* in yeasts results in the cytoplasmic accumulation of Rpb1 and Rpb3 proteins[186]. Therefore, it can be speculated that Iwr1 protein binds to the fully assembled Pol II in the cytoplasm, and helps Pol II to enter into the nucleus. This could be a checkpoint to ensure that only the correctly assembled Pol II is imported into the nucleus. When Pol II enters the nucleus, it will release the assembly factors and move to the DNA template with the help of the general transcription factors (GTFs). The assembly factors, such as GPN protein and Iwr1, are then exported and recycled to the cytoplasm by the exporting protein Crm1[181,183]. However, the model of Pol II assembly and transportation still needs further exploration, since the depletion of *GPN1* only leads to the cytoplasmic accumulation of Rpb1 and Rpb2 proteins instead of all Pol II subunits.

1.5.4. The C-terminal repeat domain of RPB1

There is an unusual structure in the C-terminal repeat domain (CTD) of RPB1, which is evolutionarily conserved in eukaryotes. The CTD contains multiple tandemly repeated heptapeptides with the consensus sequence: Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7 (Y1-S2-P3-T4-S5-P6-S7). The number of repeats varies from five in *Plasmodium yoelii* to twenty-six in yeasts, forty-five in *Drosophila* and fifty-two in mammals[187]. In most organisms, the majority of repeats match the consensus repeats, but in mammalians, the first 26 repeats closely match the consensus sequence, and the latter 26 only contains three consensus repeats.

Deletion of the CTD in mice, *Drosophila*, and yeasts is lethal. However, the full-length CTD is not required for maintaining the basic growth rate in yeasts. The minimal length

could be reduced to eight repeats to keep the viability of yeasts[188]. But still, a full-length CTD is preferred for maintaining the normal function of cells and dealing with various stress conditions, although a reduced number of repeats or some mutations of CTD is tolerated in yeasts[189,190].

The residues in CTD are subjected to multiple posttranslational modifications (PTMs), including phosphorylation, proline isomerization, and O-GlcNAcylation[191,192,193]. Some residues such as Lys7 and Arg7 within the non-consensus CTD can also be methylated, acetylated, or ubiquitinated[194,195,196,197]. These modifications are highly related to the Pol II function. The schematic diagram is shown in **Figure 1.13**.
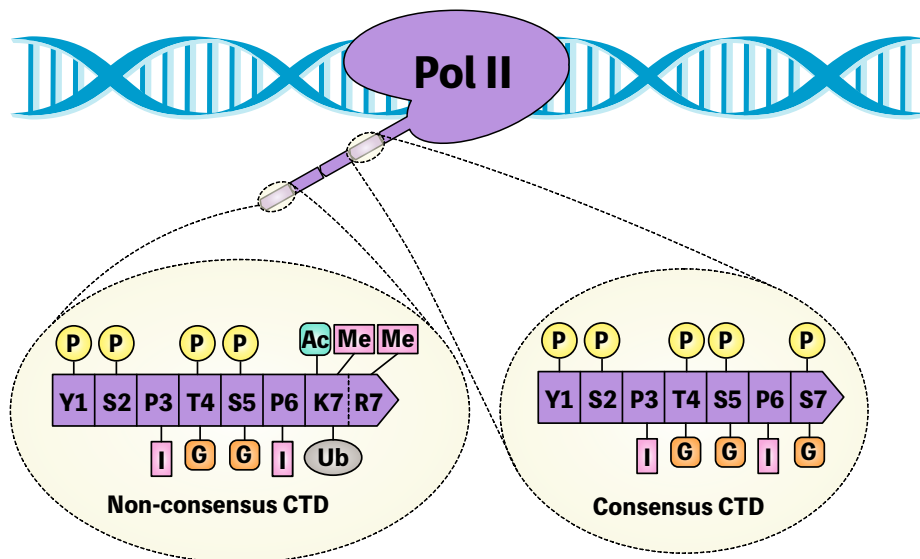
Fig. 1.13. Posttranslational modification of the consensus and non-consensus CTD In mammals, the first 26 repeats in CTD closely match the consensus, and the latter 26 repeats do not, especially Ser7 in non-consensus CTD, which can be replaced by Lys7 or Arg7. The residues in CTD are known to be subjected to phosphorylation, proline isomerization, and O-GlcNAcylation. Lys7 and Arg7 within the non-consensus CTD can additionally be methylated, acetylated, or ubiquitinated. (P), phosphorylation; (G), glycosylation; (I), isomerization; (Ac), acetylation; (Me), methylation; (Ub), ubiquitination.

By far, phosphorylation of the CTD has been extensively studied. There are 52 repeats in human CTD, and each repeat contains three serines (S2, S5, and S7), one tyrosine (Y1), and one threonine (T4) that can be phosphorylated (**Figure 1.13**). The complexity of the phosphorylated CTD pattern is beyond imagination. However, recent work has demonstrated

that not all repeats in CTD are phosphorylated[198,199]. Thus, the combination of actual phosphorylation CTD states is still manageable, about hundreds rather than an astronomical number. RPB1 has two main forms that migrated on an SDS-PAGE gel: the IIa form in which the CTD is hypo-phosphorylated and the IIo form that has a hyper-phosphorylated CTD. Phosphorylated forms of RPB1 other than IIa and IIo have also been described. An IIm form is developed during serum stimulation or during somatic and oxidative stress[200], and an IIe (embryonic) form can be found in early embryos[201], while an IIi (intermediary) form is generated after the herpes simplex virus (HSV) infection[202].

The CTD is located next to the RNA exit channel, where can directly or indirectly influence RNA synthesis[172]. The phosphorylation of CTD is essential during transcription. It influences the interaction between Pol II and other proteins, such as transcription factors and RNA processing enzymes. According to the RPB1 ChIP-Seq profiling, Ser5P and Ser7-P occur early at the transcription starting site (TSS) region, whereas Ser2P and Thr4P occur later[203] during the elongation. The average profiling of CTD phosphorylation in humans and yeasts is illustrated in **Figure 1.15**. The function of phosphorylated CTD will be reviewed later in combination with the transcription process.

1.5.5. The major steps of transcription

Transcription is the rate-limiting step of the entire gene expression machinery. Within a simplified model of transcription, it contains three major stages in eukaryotes: initiation, elongation, and termination.

1.5.5.1. Transcription initiation.

The transcription initiation starts with the formation of the pre-initiation complex (PIC) in the gene promoter. PIC is assembled by Pol II and several GTFs. Followed by DNA template unwinding, PIC turns from a closed state to an open state, and a nascent RNA chain starts to be synthesized. When the RNA chain reaches to a certain length, Pol II is released by the dissociation of the initiation factors and forms an elongation complex with

other factors. The schematic diagram of the transcription initiation machinery is shown in **Figure 1.14**.
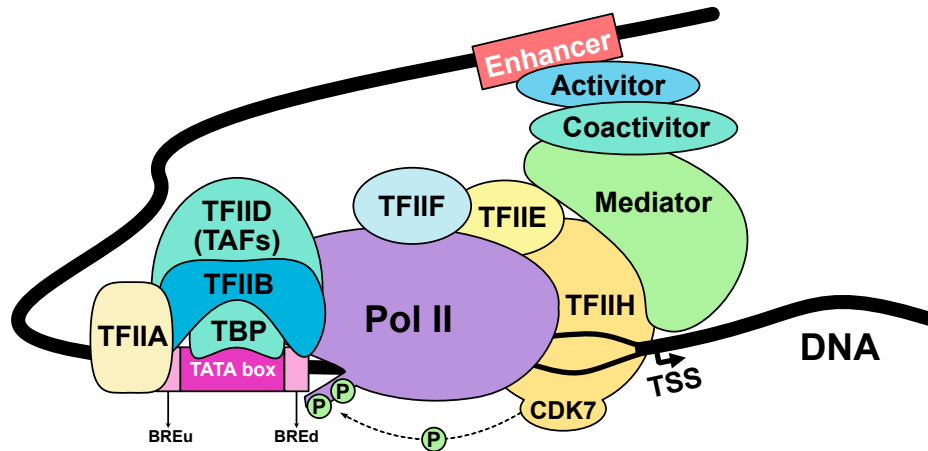
Fig. 1.14. The transcription initiation machinery A pre-formed TFIID-TFIIB-TFIIA complex locates at the TATA box and BREu/BREd regions of the promoter DNA and recruits the Pol II-TFIIF complex. TFIIE binds to Pol II and recruits TFIIH, which is able to unwind DNA. The transcription activators targeting on the enhancer DNA elements recruit coactivitors and Mediator to the pre-initiation complex (PIC). Cyclin-dependent kinase (CDK7) in TFIIH phosphorylates Ser5 and Ser7 of the CTD in RPB1 and starts elongation. BREu: B recognition element upstream; BREd: B recognition element downstream

In a canonical model, the first step of initiation is the binding of TFIID to the promoter region[204,205,206]. TFIID contains TATA box-binding protein (TBP) and 13 – 14 TBP-associated factors (TAFs). The TATA box is located at the upstream region of TSS and owns a consensus sequence of TATAWAWR[207]. TBP owns a saddle-shaped structure, which is able to bind the minor groove of TATA box and bend DNA template[208].

However, TATA box only exists in about 10 – 20% of metazoan genes[209]. How other genes without the TATA box are transcribed? The answer is not clear yet. Some other promoter elements recognized by TAFs may help to explain the selective expression of certain genes. The initiator element, which can be recognized by TAF1 and TAF2, locates at the TSS[210]. The motif ten element (MTE) and the downstream promoter element (DPE) are located at +30 bp downstream of TSS and are probably recognized by TAF6 and TAF9[211,212,213]. The downstream core element (DCE) has also been found at the downstream of TSS and may

bind by TAF1[214]. The mechanism of transcription initiation at TATA-less promoters still needs more exploration in the future.

In addition to TFIID, TFIIB facilitates TBP's binding to DNA and recruits Pol II to the promoter. There are two regions flanking the TATA box, B recognition element upstream (BREu) and BRE downstream (BREd). They serve as TFIIB binding sites and guide the orientation of the PIC. TFIIB contains an N-terminal B-ribbon domain and a C-terminal B-core domain. The B-core domain binds to TBP and promoter, while the B-ribbon domain binds to the docking domain of Pol II. TFIIA is not essential for the basal transcription, but can stabilize the TFIID-DNA complex. Thus, the pre-assembled TFIID-TFIIB-TFIIA-DNA complex is ready to recruit the Pol II-TFIIF complex.

Sequentially, the Pol II-TFIIF complex recruits TFIIE and TFIIH, which help to open and stabilize the promoter DNA[215,216]. TFIIH contains DNA-dependent ATPase activity. It is a ten-subunit factor that consists of a six-subunit core module (XPD, p62, p52, p34, p8, and p44), an ATPase XPB, and a three-subunit kinase module (CDK7-Cyclin H-MAT1)[217]. The ATPase XPB is responsible for promoter opening, whereas the helicase XPD is required for DNA unwinding.

Another conserved coactivator complex, termed Mediator is also involved in the regulation of transcription initiation. Mediator is recruited to PIC by the transcription activator, which binds to the upstream enhancer element. It can stabilize PIC and stimulate TFIIH kinase activity[218]. Mediator consists of a 4-subunit kinase and a core complex, which is composed of the head, middle, and tail modules; in total, it is a large complex comprising 30 subunits in humans[179]. The activity of Mediator partially depends on the binding to Pol II and DNA binding transactivators[219,220].

According to human RPB1 ChIP-seq profiling, the level of Ser5P in the CTD, as well as Ser7P and Tyr1P, peaks early in transcription process, around the TSS region, suggesting that the phosphorylation of these CTD residues plays a critical role in transcription initiation (**Figure 1.15**)[203].

When Pol II is recruited by TFIIB to the promoter, the CTD is unphosphorylated[221]. The unphosphorylated CTD has a high affinity to Mediator. During the transition from initiation to elongation, cyclin-dependent kinase (CDK7) in TFIIH phosphorylates Ser5 and Ser7 of the CTD, which reduces the affinity of Pol II to the Mediator and starts Pol II promoter-escape[222].

In addition to initiation, the CTD also plays an important role in the regulation of mRNA 5'-capping. The 5'-capping occurs as early as the time when a nascent pre-mRNA is just synthesized and leaves the RNA exit channel of Pol II. The 5'-capped structure protects nascent mRNA from degradation by exonucleases and assists the downstream mRNA processing, including splicing, polyadenylation, and nuclear export of mRNA[223]. The Ser5P CTD recruits the capping complex and couples pre-mRNA capping to early elongation[224]. The importance of Ser5P CTD in mRNA 5'-capping has been proved by the lethal phenotype caused by Ser5-CTD mutation in yeasts, which could be rescued by tethering the capping complex to the CTD[190].
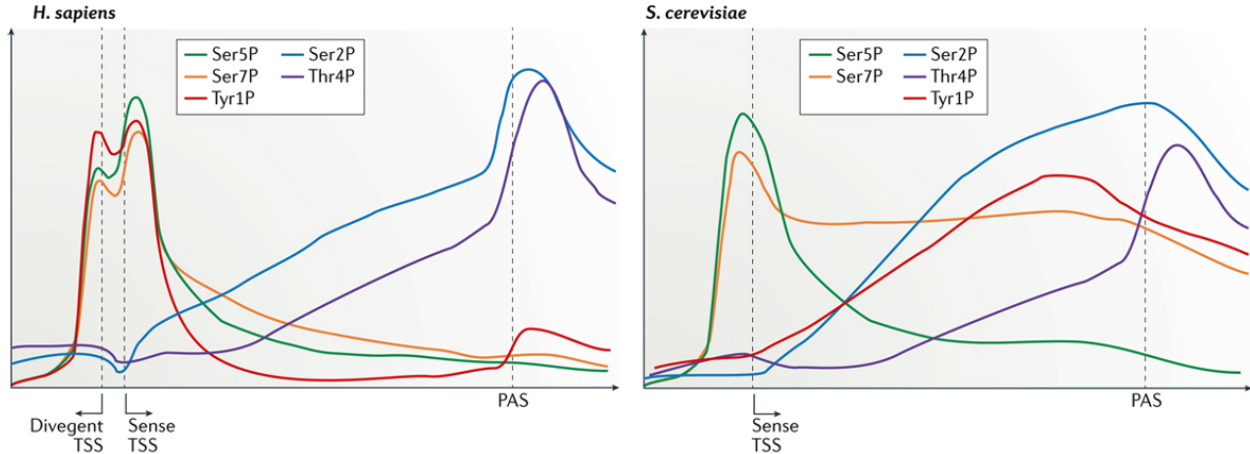
Fig. 1.15. The profiling of CTD phosphorylation of RPB1 across protein-coding genes in humans and yeasts The signals are revealed by RPB1-specific chromatin immunoprecipitation (ChIP)-seq studies. In humans and yeasts, both Ser5P and Ser7P CTD signals peak near the TSS and promoter region, while Ser2P and Thr4P CTD signals start to increase after TSS and reach to the top around PAS. The Tyr1P CTD signal acts differently between humans and yeasts. It is similar to Ser2P in yeasts, although its level drops before the PAS. By contrast, the Tyr1P signal peak around the TSS in humans. TSS: transcription start site. PAS: polyadenylation site. The figure is adapted from the publication of Harlen and Churchman[203].

1.5.5.2. Transcription elongation.

After the formation of PIC and promoter escape, Pol II enters the early stage of elongation. However, in most metazoan genes, Pol II will pause after transcribing 20 – 120 nucleotides downstream of TSS. This status is termed promoter-proximal pausing, which is a transition stage between early and productive elongation.

The promoter-proximal pausing state is stabilized by the binding of negative elongation factor (NELF) complex and DRB sensitivity-inducing factor (DSIF). Knockdown of either factor increases the elongation rate significantly[225]. Positive transcription elongation factor B (P-TEFb) is required to release the pausing Pol II. P-TEFb, which is an essential CDK, phosphorylates NELF, causing its disassociation from the elongation complex[226], and phosphorylates DSIF, converting it into a positive elongation factor[227]. P-TEFb also phosphorylates Ser2 of the CTD, which helps Pol II to recruit several elongation and chromatin-modifying factors. The phosphorylation process by P-TEFb during the promoter-proximal pausing stage is illustrated in **Figure 1.16**.
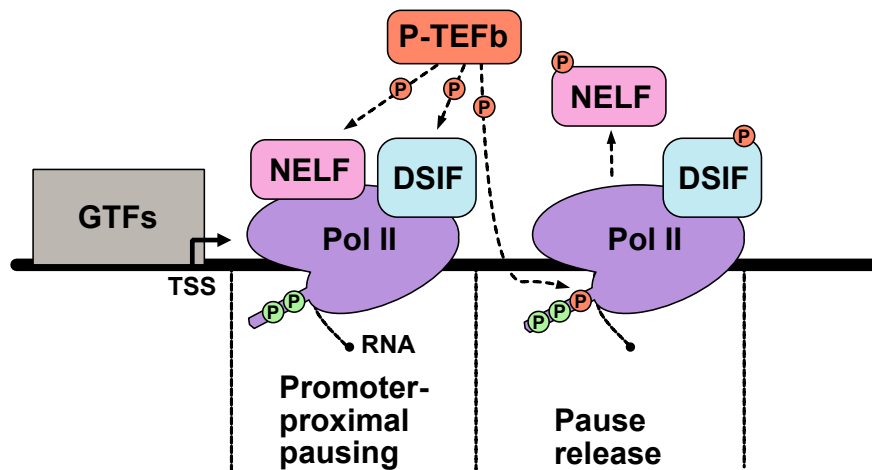
Fig. 1.16. The promoter-proximal pausing and releasing The promoter-proximal pausing state is stabilized by the binding of negative elongation factor (NELF) complex and DRB sensitivity-inducing factor (DSIF). Positive transcription elongation factor B (P-TEFb) phosphorylates NELF, causing its disassociation to the elongation complex, phosphorylates DSIF, converting it into a positive elongation factor, and phosphorylates Ser2 CTD of RPB1, which helps to recruit other elongation and chromatin-modifying factors. Then, Pol II is released from the promoter-proximal pausing state.

In addition to P-TEFb, recent studies have revealed that bromodomain-containing protein 4 (BRD4), the super elongation complex (SEC), and the transcription elongation complex RNA polymerase II-associated factor (PAF1) also contribute to transcriptional elongation[228,229].

The degree of promoter-proximal pausing is generally estimated by the pausing index[230]. Pausing index is the ratio of Pol II signal around TSS to the signal density within the genebody. A higher pausing index indicates a greater enrichment of promoter-proximal pausing Pol II.

1.5.5.3. Pausing, Stalling, and Backtracking upon DNA lesions.

During the elongation, Pol II has a strong propensity to enter a pausing or stalling state on the DNA template. Pol II pausing is a transient state and will restart if given time. If pausing persists and cannot be overcome by time, especially when elongation is interrupted by DNA sequence problems or misincorporation of nucleotides, it will decay into a stalling state.

Under this circumstance, Pol II can move backward on the DNA track, which is termed "backtracking", that normally occurs when Pol II faces obstacles such as nucleosomes. When Pol II persistently stalled, the extended backtracking, with the help of other factors, such as the RNA-cleavage stimulatory factor TFIIS, will rescue Pol II. TFIIS is able to stimulate the intrinsic RNA-cleavage function of Pol II, thereby removing the backtracked RNA fragment and synthesizing a new 3'-end of RNA[231].

DNA damage is one of the reasons causing Pol II stalling. Environmental factors, DNA-reactive chemicals, and irradiation are the reasons causing a broad range of DNA lesions[232]. When DNA lesions occur in the non-template strand, Pol II can rapidly bypass it without any consequences. However, when the lesions occur in the template strand, they do affect Pol II transcription. There are abundant of small DNA lesions originated from normal cellular activities, such as alkylation and oxidized nucleotides[233]. Most small lesions in the template DNA can often be bypassed by Pol II, which causes misincorporation of nucleotides, thereby

resulting in transcriptional mutagenesis. As long as lesions persist, the accumulated mutant transcripts will finally affect cell functions[234].

The large lesions will cause even bigger problems for transcription. The forward translocation of Pol II is completely obstructed, and the expression is totally interrupted. These lesions can be induced by UV irradiation[235], various carcinogens[236], and some intracellular metabolites[237]. Nucleotide excision repair (NER) system is the major pathway to repair bulky DNA lesions. For repairing genome-wide lesions, the sub-pathway global genome NER (GG-NER) will be involved. If the lesion is in the transcribing strand, transcription-coupled NER (TC-NER) will participate in the repairing mechanism. Pol II stalling is thought to trigger TC-NER. Cockayne syndrome group B (CSB) protein works with Pol II to sense the blockage of transcription[238]. When Pol II is unable to move forward, the TC-NER factors, such as CRL4A-DDB1-CSA complex, UV-stimulated scaffold protein A (UVSSA), and USP7 are recruited to the lesion site and initiate the repair process. TFIIH is recruited by UVSSA and may induce Pol II backtracking[239]. TFIIH, together with XPA, verifies the lesion site and recruits endonucleases ERCC1-XPF and XPG to remove the lesion DNA. The resulting gap will be filled by DNA polymerase and sealed by DNA ligase[240].

1.5.5.4. Splicing during elongation.

A majority of the splicing processes occurs simultaneously during elongation and almost immediately when the 3' splice site is transcribed[241,242]. Splicing refers to the removal of introns from the nascent pre-mRNA by the spliceosome. The spliceosome is a huge enzyme complex, composed of five small nuclear RNAs (snRNAs) (U1, U2, U4, U5, and U6) and approximately 100 proteins[243].

The co-transcriptional splicing is partially regulated by the CTD status of RPB1. Pol II with a high level of Ser5P CTD interacts with the spliceosome during transcription[244]. Phosphorylation of Ser2 CTD by P-TEFb activates not only elongation but also the splicing process, whereas the mutant Ser2 CTD inhibits the recruitment of spliceosome[245]. The Ser2P and Thr4P CTD signals, according to ChIP-seq profiling (**Figure 1.15**), start to increase

in the genebody and peak at the polyadenylation sites (PAS), suggesting that they acted important roles during the entire elongation stage.

1.5.5.5. Transcription termination.

A regular transcription termination for most protein coding genes occurs when the elongation complex encounters a functional polyadenylation signal (PAS). A PAS usually includes an AAUAAA (or a variant) sequence in company with upstream U-rich and downstream U/GU rich sequences[246].

When Pol II pauses at PAS, the 3'-end of the nascent mRNA undergoes cleavage and polyadenylation. Several proteins participate in the termination process, along with Pol II. The human cleavage and polyadenylation specificity factor (CPSF) binds to the body of Pol II to induce its pausing and recognizes AAUAAA sequence that transcribed in the nascent RNA. When the downstream U/GU-rich sequence is transcribed, the cleavage stimulatory factor (CstF) is recruited by both this sequence and Ser2P CTD of RPB1, which leads to the CPSF-mediated cleavage and finally release Pol II. Afterwards the upstream cleaved RNA is added 200 – 300 with adenosine nucleotides at the 3'-terminus by Poly(A) polymerase, whereas the downstream product is degraded by 5' – 3' exoribonuclease 2 (XRN2). Pol II is then released from the DNA template and recycled for the next round of transcription, or degraded. The process of transcription termination is illustrated in **Figure 1.17**.
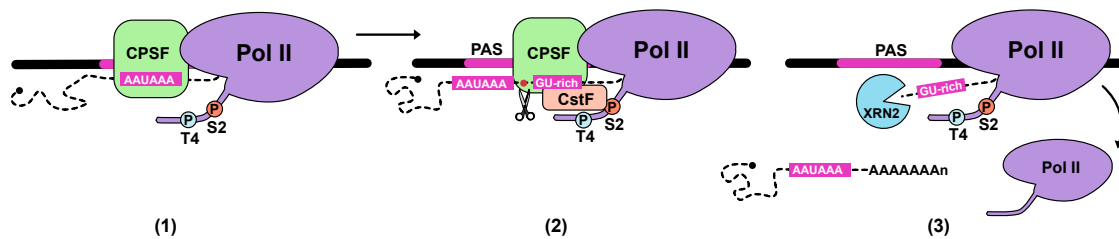
Fig. 1.17. The termination machinery in metazoans (1). The human cleavage and polyadenylation specificity factor (CPSF) binds to the body of Pol II and recognizes a functional PAS (e.g., AAUAAA sequence) in the nascent RNA. **(2).** The downstream U/GU-rich sequence and Ser2P CTD recruit the cleavage stimulatory factor (CstF), which leads to the CPSF-mediated cleavage before the U/GU-rich region. **(3).** The released upstream RNA is added with 200 – 300 adenosine nucleotides at the 3'-terminus by Poly(A) polymerase, whereas the downstream product is degraded by 5' – 3' exoribonuclease 2 (XRN2). Pol II is then released from the DNA template. PAS: polyadenylation signal.

The phosphorylation of CTD also plays an important role in transcription termination. Most mechanism studies are conducted in yeasts. During the elongation stage, Tyr1P CTD prevents the binding of Pcf11 and Rtt103 proteins to Pol II, since Pcf11 and Rtt103 are the termination factors that function on the CTD.

When it enters to termination stage, loss of Ser5P, Tyr1P, and increase of Ser2P in CTD will recruit Pcf11 to Pol II. Pcf11 acts as a cleavage and polyadenylation factor that terminates RNA synthesis. Besides, Ser2P and Thr4P CTD recruit Rtt103 in complex with Rat1 (homolog exoribonuclease as XRN2) to Pol II[247,189], which helps to degrade the downstream RNA.

According to RPB1-specific ChIP-seq signals (**Figure 1.15**), Ser2P and Thr4P peak at PAS, whereas Ser5P reduces to baseline at PAS, suggesting that the overall phosphorylation status of the CTD is finely regulating the transcription machinery.

However, most non-coding RNAs employ an alternative Pol II termination pathway. The 3'-end of snRNA is produced by endoribonucleolytic and/or exoribonucleolytic cleavages, and they do not require a poly(A) tail in their mature forms[248]. The process involves a distinct group of proteins, such as the Integrator (INT) in metazoans. The termination sites of genes encoding snRNA typically locate at a distance <1 kb from TSS, where the signal of Ser2P and Thr4P does not reach the peaks, indicating it may be a Ser2P/Thr4P independent termination mechanism. It has been proved that INT terminates transcription by recognizing Ser7P CTD and a sequence at the 3'end of snRNA[249]. In yeasts, Nrd1 (a similar role as INT) tends to bind Ser5P CTD[250], unlike Pcf11 and Rtt103 that preferentially bound to Ser2P CTD in a poly(A)-dependent pathway[251,252].

1.5.5.6. Gene-looping.

The canonical view of where Pol II goes after transcription termination is that Pol II dissociates from the template DNA and then is either degraded by the proteasome or recycled to the pool of Pol II. Some recent work updates this model with a new mechanism called "gene-looping". Pol II initiation factors and termination factors are more intertwined than one might expect. TFIIB interacts with both the CPSF 3'-end processing complex and PIC at

the promoter and guides Pol II to the promoter region from the 3'-end termination site. With the introduction of Chromosome Conformation Capture (3C) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), recent studies indicate that gene-looping is a widespread phenomenon existing in diverse species, such as *S.cerevisiae*[253,254], HIV[255], mice[256], and humans[257]. However, the detailed function of gene-looping needs to be explored deeply.

1.5.5.7. Premature termination.

Transcription termination not only occurs at the end of a gene, it can also happen at the upstream, middle, and downstream of the gene body. The termination occurring at TSS or further down in the gene body is called premature termination. Premature termination generates transcripts that are either rapidly degraded or polyadenylated. The stable premature transcript tends to produce noncoding RNA or truncated protein-coding mRNA[258]. They negatively regulate the full-length gene expression and contribute to the transcriptome diversity. Premature termination was thought to be harmful. However, recent studies showed the rapid turnover of the initiating and promoter-pausing Pol II[259,260], suggesting that only a small portion (∼1%) of the initiating Pol II complete the entire transcription cycle, whereas most of them (∼99%) are released from DNA without transcribing mRNA[260]. These premature terminated Pol II may have high potential to regulate mRNA expression in various types of cells.

1.5.6. The ubiquitination of RPB1

Over the decades, it has become evident that the ubiquitination of RPB1 is a crucial process full of complexity. It occurs via a multiple-step ubiquitination process involving various E3s, different Ub linkages, and multiple ubiquitination sites. RPB1 ubiquitination guides Pol II to the proteasome for degradation under different conditions. The ubiquitination of RPB1 not only occurs upon cellular stress, such as DNA damage, nucleotide depletion, and chromatin impediment, but also occurs in unstressed cells to maintain Pol II's basic function and turnover[261,262].

Many ubiquitination sites of RPB1 have been identified by a proteome-wide quantitative assay applying the Lys-ε-Gly-Gly (diGLY) antibody enrichment method[262]. Ten lysine sites in RPB1 are indicated as the potential ubiquitination residues in non-stressed cells. Another thirty-four lysine sites are identified in DNA damaged cells[261], including K1268, which is an important site in response to UV damage[263,264]. Several ubiquitination sites, such as K619, K940, and K1008, also indicated their association with the activity of CRLs. In addition, another six ubiquitination sites of Rpb1 protein in *S. cerevisiae* are listed in PTMFUNC database (`http://ptmfunc.com`)[265].

Most of the ubiquitination studies about Rpb1 protein are conducted in yeasts. The first identified E3 implicated in Rpb1 degradation is Rsp5 protein in yeasts[266]. Rsp5 protein has diverse roles in cells, including transcriptional activation and Pol II degradation. The WW domain of Rsp5 binds to the CTD of Rpb1. Deletion of *RSP5* causes a lack of ubiquitination of Rpb1, resulting in its accumulation upon DNA damage[267,268].

Ubiquitination assay of yeast Pol II *in vitro* using Uba1 (E1), Ubc5 (E2), and Rsp5 proves that Rsp5 adds K63-linked polyUb chains to Rpb1. However, K63 linkage is not a typical proteasomal degradation signal, which probably cannot be the direct reason for the accumulation of Rpb1 after Rsp5 inactivation. Later, they identified a K48 linkage-specific E3, Elc1/Ela1 complex, targeting on Rpb1 protein by a sequential reaction. First, Rpb1 is ubiquitinated by Rsp5 in a K63-linkage manner; followed by Rsp5-associated ubiquitin protease (Ubp2) trimming K63-linked chains to a single moiety on Rpb1[269], Elc1/Ela1 complex adds K48-linked chains on the mono-ubiquitinated Rpb1 and directs its degradation[268].

Other than Rsp5 in yeasts, Asr1 protein, a RING finger E3 ligase, can also bind to Pol II via the Ser5P CTD of Rpb1 protein and ubiquitinate at least two subunits of Pol II (Rpb1 and Rpb2), leading to the inactivation of Pol II[270]. Ubp3 DUB is capable of removing both mono- and poly-ubiquitinated Rpb1 in yeasts. It can save Rpb1 protein from degradation[271]. These results in yeasts indicate that Rpb1 ubiquitination is controlled by a complicated machinery involving various proteins in UPS.

The discovery process of RPB1-specific E3s in mammalian cells has not all been smooth going, since there are lots of E3-related proteins directly or indirectly affecting the ubiquitination of RPB1. Several E3s have been implicated in this process. However, some of them have been proved later that they only act as indirect roles in regulating the ubiquitination of RPB1.

In the initial studies, BRCA1-BARD1 E3 complex was indicated as an E3 ligase for RPB1. At that time, a smeared band above the hyper-phosphorylated (IIo) band, which was detected by the antibody against IIo form of RPB1, was believed as the ubiquitination signal of RPB1, which was not a very strong evidence as we now consider[272,273]. In addition, Cockayne syndrome type A (CSA) protein and Cockayne syndrome type B (CSB) protein, which are parts of CUL4-based E3 complex, have also been implicated in RPB1 ubiquitination process, as a reduced level of ubiquitinated RPB1 was observed in CSA- and CSB-deficient cells after DNA damage[274]. The authors hypothesized that CSA-containing CUL4 E3 complex directly ubiquitinates Rpb1. However, later studies have proved that the reduced RPB1 ubiquitination and the accumulation of RPB1 were caused by the shutdown of transcription after DNA damage[275,276].

Anindya et al. proved that neither CSA, CSB, nor BRCA1 was directly involved in Pol II ubiquitination, while NEDD4, a mammalian homolog of Rsp5 protein, was involved in the ubiquitination of human Pol II *in vivo* and *in vitro*[275]. NEDD4 cooperates with Elongin A/B/C containing E3 to catalyze the poly-ubiquitination of RPB1, which could be a similar process as Rsp5 collaborating with Elc1/Ela1 complex in yeasts[277,268]. However, in HEK293 cells, the ubiquitination of RPB1 catalyzed by NEDD4 only occurs when they are UV irradiated[275].

As mentioned previously, Elongin B/C not only binds to CUL5, but also acts as a component of pVHL-Elongin B/C-CUL2-RBX1 E3 complex. They participate in the ubiquitination of RPB1 with hyperphosphorylated CTD in response to DNA damage and oxidative stress[278,279].

Last, WWP2, a HECT E3, could ubiquitinate the CTD of RPB1 in the absence of artificial DNA damage[194]. However, this was only proved *in vitro* in F9 embryonic carcinoma cells, in which *WWP2* siRNA knockdown leads to increased total RPB1, as well as Ser2P and Ser5P RPB1. The activity of these RPB1-specific E3s highly depends on RPB1 CTD phosphorylation. The CTD alone can serve as the substrate of WWP2 E3 ligase[194].

A brief summary of these potential E3s are listed in **Table 1.1**.

Table 1.1. The known E3 ligases for RPB1 in *S.cerevisiae* and mammalian cells

S.cerevisiae	Mammals	Functions	Interacting domain	External factors
Rsp5[266,267,268]	NEDD4[268,275]	K63-linked chains, probably be trimmed to mono-Ub	CTD	DNA damage
Elc1/Ela1 (Cul2/Cul5?)[280,281,268]	Elongin A/B/C-CUL5-RBX2[282]	Coopreates with Rsp5/NEDD4 and polyubiquitinates RPB1	Ser5P CTD	DNA damage
	pVHL-Elongin B/C-CUL2[278,279]	Ubiquinates hyperphosphory-lated RPB1	Ser5P CTD	Oxidative Stress, DNA damage
?-Cul3[281,268]	?-CUL3	Polyubiquitinates Rpb1	?	DNA damage
Asr1[270]	-	Ubiquinates Rpb1 and Rpb2	Ser5P CTD	No
-	WWP2[194]	Proteolytic ubiquitination of RPB1	CTD	No

Note: The question-markers (?) indicate that the factors are undetermined yet.

As we have known, an E3 directing proteasomal degradation will markably affect the amount of substrate in cells. However, under the physiological condition (without inducing DNA damage), neither *Nedd4* KO B cells[283], *CSB* KO cells, nor *ELOF1* KO cells[284] showed any accumulation of RPB1, suggesting that this group of E3s might only be activated upon DNA damage, instead of controlling the regular process of RPB1 turnover.

Furthermore, since *Wwp2*[194,285] and *Vhl*[286] KO mice have been generated, there is still no report on the possible RPB1 accumulation in these mice. Either authors have not paid attention to the protein level of RPB1, or there is no failed RPB1 degradation under a

physiological condition in these KO mice. Thus, to the best of our knowledge, there is no report of E3 acting on RPB1's degradation in tissues and organs without artificially induced DNA damage.

1.6. Objectives and Hypothesis

Our lab started to explore *ARMC5* in 2007, since it was one of the genes upregulated after T cell activation. During our investigation of *ARMC5*'s function, it has been reported that some mutations in *ARMC5* are related to PBMAH in humans[1,19].

1.6.1. Objective 1

To study an unknown gene from scratch, we first need to resolve the following questions.

(1) To explore the tissue-specificity of *Armc5* expression in mice

(2) To identify the sub-cellular location of ARMC5 protein

(3) By generating *Armc5* KO mice model, we would like to record any phenotypes we observed, especially the ones related to the immune system and the endocrine system.

- To study the proliferation and differentiation of *Armc5* KO T cells
- To investigate the role of ARMC5 in T cell immune response *in vivo*
- To check whether the adrenal glands in *Armc5* KO mice develop a similar phenotype to PBMAH in humans; meanwhile, to explore adrenal endocrine function in KO mice

(4) To investigate ARMC5's molecular function by identifying its binding partners, since ARMC5 only contains two protein binding modules and itself does not exhibit enzymatic activity.

1.6.2. Objective 2

From both Y2H and IP-LC-MS/MS results, we have identified several potential binding partners of ARMC5. By confirming with Co-IP experiments, we finally targeted on ARMC5's interaction with CUL3 and RPB1.

Since CUL3 is a well-known E3 ligase, we hypothesize that ARMC5 binds to CUL3 to form an E3 complex, which ubiquitinates RPB1 protein, and the defects of RPB1 ubiquitination might contribute to the phenotypes we observed in *Armc5* KO mice and in humans carrying *ARMC5* mutations. To testify the hypothesis, we need to clarify the following points.

(1) To confirm that three molecules (CUL3, ARMC5 and RPB1) form a complex

(2) To identify the binding regions of each molecule

(3) To check whether there is a decreased ubiquitination of RPB1 in ARMC5 deficiency *in vivo* and *in vitro* systems

(4) To investigate the consequences of failed ubiquitination of RPB1 in *Armc5* KO mice, especially in lymphoid organs and adrenal glands

- Does ARMC5-related ubiquitination direct RPB1 to the proteasomal degradation pathway?
- Does ARMC5 depletion change the expression and function of RPB1 in lymphoid organs and adrenal glands?
- Does the failed ubiquitination of RPB1 affect the transcription process in *Armc5* KO tissues and cells?

(5) To explore the role of RPB1 in the pathogenesis of PBMAH in the patients carrying *ARMC5* mutations

1.6.3. Objectives 3

One of the phenotypes in *Armc5* KO mice is that they are born below the expected Mendelian ratio from heterozygous parents. Later, we also observed an increased incidence of kinky tails in the live-born KO mice. To explore the cause of embryonic lethality of KO fetuses, we found these fetus manifested anencephaly, which is a severe type of NTDs. Based on these facts, we would like to understand the following questions.

(1) What is the cellular basis (e.g., cell proliferation and cell apoptosis) that causes NTDs in *Armc5* KO mice?

(2) Based on the hypothesis that CUL3-ARMC5 serves as an E3 ligase for RPB1, we would also want to explore the role of RPB1 in the pathogenesis of NTDs in *Armc5* KO mice.

- Are the protein level and ubiquitination level of RPB1 changed in *Armc5* KO neural tubes and neural precursor cells (NPCs)?
- Does the defective ubiquitination of RPB1 change the transcriptome in *Armc5* KO NPCs?

(3) Is *ARMC5* mutation reported to be relevant to NTDs in humans?

Chapter 2

Armc5 deletion causes developmental defects and compromises T-cell immune responses

[*][1]Yan Hu, [*][1]Linjiang Lao, [1]Jianning Mao, [1]Wei Jin, [1]Hongyu Luo, [2]Tania Charpentier, [1]Shijie Qi, [1]Junzheng Peng, [3]Bing Hu, [4]Mieczyslaw Martin Marcinkiewicz, [2]Alain Lamarre, and [1,5]Jiangping Wu

From the [1]Centre de recherche (CR) and [5]Nephrology Service, Centre hospitalier de l'Université de Montréal (CHUM), 900 Rue Saint-Denis, Montréal, Québec H2X 0A9, Canada; [2]Institut national de la recherche scientifique-Institut Armand-Frappier (INRS-IAF), 531 Boul. des Prairies, Laval, Québec H7V 1B7, Canada; [3]Anatomic Pathology, AmeriPath Central Florida, 4225 Fowler Ave. Tampa, Orlando, Florida 33617, USA; [4]Cytochem Inc., 6465 Av Durocher, Outremont, Montréal, Québec H2V 3Z1, Canada

Address correspondence to: Dr. Jiangping Wu, CRCHUM, 900 Saint-Denis Street, Room R12.428, Montréal, Québec, Canada H2X 0A9. Telephone: (514) 890-8000 Extension 25164, Fax: (514) 412-7944, e-mail: jianping.wu@umontreal.ca

Author contributions: [*]Y.H. and [*]L.L. contributed equally to this work. Y.H., L.L, J.M., A.L. and J.W. conceived and designed the experiments. Y.H., L.L., J.M., W.J., H.L., T.C., S.Q., J.P. and M.M. performed the experiments. Y.H., L.L., J.M., W.J., H.L., T.C., M.M. and J.W. analyzed the data. Y.H., L.L., A.L. and J.W. wrote the manuscript.

Competing financial interests: The authors declare no competing financial interests.

2.1. Abstract

Armadillo repeat containing 5 (ARMC5) is a cytosolic protein with no enzymatic activities. Little is known about its function and mechanisms of action, except that gene mutations are associated with risks of primary macronodular adrenal gland hyperplasia. We have mapped *Armc5* expression by *in situ* hybridization, and generated *Armc5* knockout (KO) mice, which are small in body size. Here we show that these KO mice present with compromised T cell proliferation and differentiation into Th1 and Th17 cells, increased T cell apoptosis, reduced severity of experimental autoimmune encephalitis, and defective immune responses to lymphocytic choriomeningitis virus infection. Furthermore, KO mice develop adrenal gland hyperplasia in old age. Yeast 2-hybrid assay identified 16 ARMC5-binding partners. These data indicate that ARMC5 is crucial in fetal development and is pivotal in promoting T cell growth, differentiation and function as well as in adrenal gland growth homeostasis, and that the functions of ARMC5 likely depend on its interaction with molecules involved in multiple signaling pathways.

2.2. Introduction

The gene *Armadillo* was first identified in the fruit fly *Drosophila melanogaster* as a gene controlling larval segmentation with morphological similarity to armadillos[287,288]. β-Catenin is the human and mouse orthologue of fruit fly Armadillo[289]. Armadillo/β-catenin protein contains 13 and 12 conserved Armadillo (ARM) repeats, respectively: each repeat is about 40 amino acid (aa) long and consists of 3 α-helices[290]. Multiple repeats form an ARM domain which has a groove for binding various other proteins in its tertiary structure[291]. More than 240 proteins, from yeasts to humans, are known to contain an ARM domain[292,293]. Although β-catenin is believed to interact with and regulate cytoskeleton function, its roles and those of ARM domain-containing proteins, in general, are very versatile in cell biology, including cytoskeleton organization[294], cell-cell interactions[295], protein nuclear import[296], degradation[297] and folding[298], cell signaling/sensing[299,300,301], molecular chaperoning[302], cell

invasion/mobility/migration[303], transcription control[304], cell division/proliferation[305], and spindle formation[306], to name some of them.

At the tissue and organ levels, ARM domain-containing proteins are involved in T cell development[307], lung morphogenesis[308], limb dorsal-ventral axis formation[309], neural tube development[310], osteoblast/chondrocyte switch[311], synovial joint formation[312], adrenal gland cortex development[313], and tumor suppression[314].

Due to the very diverse functions of ARM domain-containing proteins, it is challenging to predict their mechanisms of action. Indeed, these aspects of many ARM domain-containing proteins remain undeciphered, and quite a number of them are given the name ARMC (ARM repeat-containing), followed by Arabic numbers (e.g., ARMC1, 2, 3 and so on). ARMC5 is one such protein.

Human and mouse ARMC5 proteins share ∼90% aa sequence homology and have similar structures[315,316]. Mouse ARMC5 is 926 aa in length and contains 7 ARM repeats. A BTB/POZ domain towards its C-terminus is responsible for dimerization or trimerization[317,318,319]. Several groups reported in 2013 and 2014 that *ARMC5* gene mutations are associated with primary macronodular adrenal hyperplasia (PMAH) and Cushing's syndrome[1,320,321]. Assié et al.[1] demonstrated that no viable HeLa cells could be obtained when they were stably transfected with *ARMC5*-expressing vectors. They suggested that the default function of wild type (WT) ARMC5 is the suppression of cell proliferation or promotion of apoptosis, which might explain the adrenal cortex hyperplasia seen in patients with *ARMC5* mutations. No other reports on ARMC5 function and mechanisms of action are available in the literature.

In the present work, we have studied the tissue-specific expression of *Armc5*. We have generated *Armc5* gene knockout (KO) mice, and revealed that ARMC5 is vital in development and immune responses. We also show that aged KO mice develop adrenal gland hyperplasia. We have identified a group of ARMC5-interacting proteins by yeast 2-hybrid (Y2H) assay, paving the way for further mechanistic and functional investigations of ARMC5.

2.3. Results

2.3.1. *Armc5* expression in mice and T cells

Armc5 mRNA expression was analyzed by *in situ* hybridization (ISH) in adult WT mice. Hematoxylin/eosin (HE) staining of a consecutive sagittal whole body section preceded ISH (**Figure 2.1a**, upper panel). *Armc5* expression, based on anti-sense riboprobe hybridization (**Figure 2.1a**, middle panel), was high in the thymus, stomach, bone marrow, and lymphatic tissues (including lymph nodes and intestinal wall). The hybridization was also apparent in the adrenal gland, and skin. Some hybridization occurred in brain structures, with noticeable levels found in the cerebellum. Control hybridization with sense (S) riboprobes revealed a faint nonspecific background (**Figure 2.1a**, bottom panel).

At the anatomical level, *Armc5* expression was high in the thymus cortex (**Figure 2.1b**, upper left panel). This was confirmed at the microscopic level (**Figure 2.1b**, bottom left panel). The higher *Armc5* signals in the cortex than in the medulla were due to higher cell density in the former. Sense riboprobes detected little background noise (**Figure 2.1b**, right left panel). Based on reverse transcription-quantitative polymerase chain reaction (RT-qPCR) results, thymic stroma cells (including epithelial cells) had *Armc5* expression similar to that of thymocytes (Supplementary **Figure 2.S.1a**). Further, there was no significant difference in *Armc5* expression among thymocyte subpopulations (CD4/CD8 double-negative (DN), CD4/CD8 double-positive (DP), and CD4 or CD8 single-positive (SP)) (Supplementary **Figure 2.S.1b and c**; gating strategy: Supplementary **Figure 2.S.2a**), and between naïve and memory spleen T cells (Supplementary **Figure 2.S.1d**; gating strategy: Supplementary **Figure 2.S.2b**).

Moderately intense *Armc5* labeling was apparent in spleen white pulp but not in red pulp (**Figure 2.1c**, upper panel). At the microscopic level, small groups of cells in WP displayed *Armc5* signals (**Figure 2.1c**, bottom panel). Sense riboprobes detected no signals (**Figure 2.1c**, middle panel).

Armc5 mRNA expression was induced rapidly in CD4$^+$ cells in 2 h after anti-CD3ε plus anti-CD28 stimulation, then subsided and remained low between 24 to 72 h post-activation (**Figure 2.1d**, upper row). CD8$^+$ T cells had less *Armc5* induction and the levels remained low between 24 to 72 h (**Figure 2.1d**, lower row).

ARMC5 was mainly a cytosolic protein, as it was detected in the cytoplasm of L cells transiently transfected with a mouse ARMC5 expression construct (**Figure 2.1e and f**).
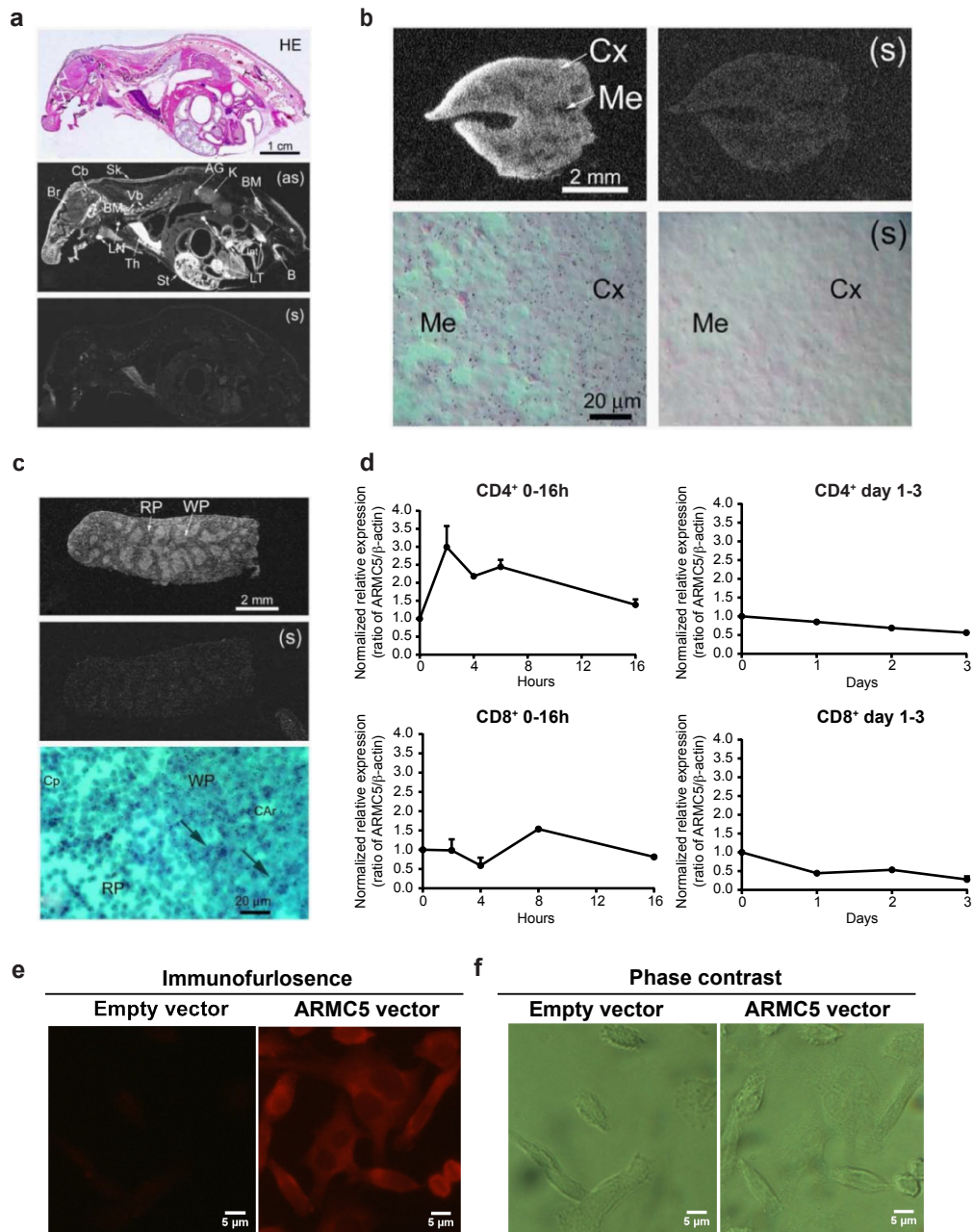
Fig. 2.1. *Armc5* tissue-specific expression *(see next page)*

Fig. 2.1 *(previous page).* *Armc5* mRNA expression in mice is assessed by ISH. (**a**) *Armc5* expression in adult mouse using whole-body sections. Upper panel: H/E staining; middle and bottom panels: dark field X-ray film autography with anti-sense (AS) cRNA or sense (S) cRNA as probes, respectively. Bar = 1 cm. AG: adrenal gland; B: bone; BM: bone marrow; Cb: cerebellum; K: kidney; Lint: large intestine; LT: lymphatic tissue; Sk: skin; ST: stomach; Th: thymus; VB: vertebrae. (**b**) *Armc5* expression in the adult thymus. Upper row: dark field X-ray film autography; lower row: bright field emulsion autoradiography; left column: anti-sense probe; right column: sense probe. Bars = 2 mm and 20 μm. Cx: cortex; Me: medulla. (**c**) *Armc5* expression in the adult spleen. Upper and middle panels: dark field X-ray film autography, with anti-sense and sense probes, respectively; bottom panel: bright field emulsion autoradiography. Bars = 2 mm and 20 μm. WP: white pulp; RP: red pulp; CAr: central artery; Cp: capillary. (**d**) *Armc5* mRNA in mouse spleen CD4$^+$ and CD8$^+$ cells, measured by RT-qPCR. Experiments were performed 3 times. The results of representative experiments are shown. To facilitate comparison, normalized ratios of *Armc5* versus *β-actin* signals (means ± SEM) are presented; the 0 h signal ratio of each experiment is considered as 1. (**e**) ARMC5 subcellular localization in L cells was detected by immunofluorescence. L cells were transfected with HA-tagged mouse ARMC5-expressing construct or an empty vector, as indicated. (**f**) Phase contract micrographs of views in (**e**). The experiments were conducted 3 times, and micrographs of a representative experiment are shown. Scale bar: 5 μm.

2.3.2. Generation of *Armc5* KO mice

We generated *Armc5* KO mice to understand the biological roles of ARMC5 in general and T cell-mediated immune responses in particular. Our targeting strategy is illustrated (**Figure 2.2a**). Germline transmission was confirmed by Southern blotting of tail DNA (Supplementary **Figure 2.S.3**). With the 5'-end probe, the WT allele after EcoRV digestion gave a 9.3 kb band, and the KO allele, a 6.6 kb band (Supplementary **Figure 2.S.3**). With the 3'-end probe, the WT allele after HindIII digestion presented a 12.5 kb band, and the KO allele, an 8.7 kb band (Supplementary **Figure 2.S.3**, lower panel). WT (mice 3 and 7) and heterozygous mice (mice 1, 2, 4, 5 and 6) were thus identified. Mouse 1 in the original 129/sv × C57BL/6J background was backcrossed to different genetic backgrounds for experimentation, as detailed below.

Armc5 deletion of KO mice at the mRNA level in spleen T cells, thymocytes, lymph nodes, brain and adrenal glands was confirmed by RT-qPCR (**Figure 2.2b**).
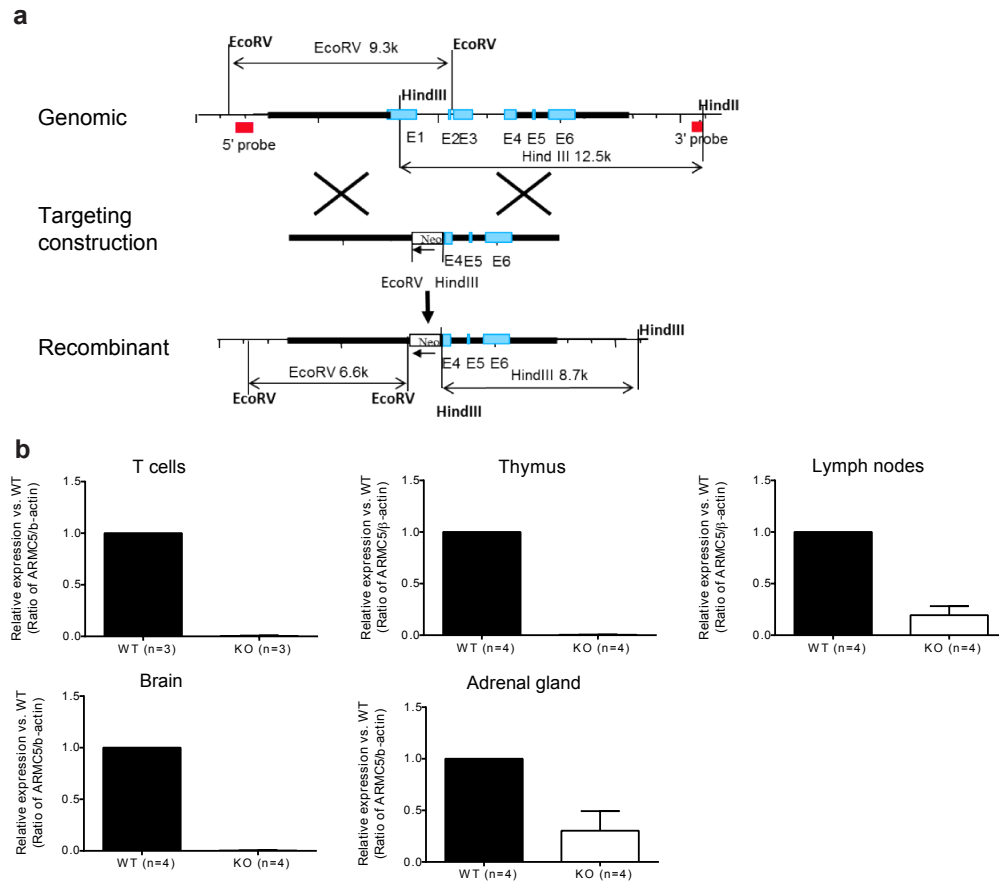
Fig. 2.2. Generation of *Armc5* KO mice (**a**) *Armc5* KO mice were generated by targeted gene deletion. The targeting strategy is depicted. Red squares on 5' and 3' sides of the mouse *Armc5* WT genomic sequence represent sequences serving as probes for genotyping by Southern blotting. (**b**) *Armc5* mRNA deletion in KO mice was confirmed by RT-qPCR. The results are expressed as normalized ratios (means $\pm$ SEM) of *Armc5* versus *β-actin* mRNA signals. The values from WT mice are considered as 1. Experiments were conducted more than 3 times, and representative results are reported.

2.3.3. General phenotype of *Armc5* KO mice

When *Armc5* KO mice were in the C57BL/6J $\times$ 129/sv F1 background, only about 10% live KO pups were delivered in a heterozygous $\times$ heterozygous mating strategy, below the expected 25% Mendelian rate. After F1 mice were backcrossed to C57BL/6 for 5 or more generations, no KO pups could be produced, nor were live KO pups born after the mice were

backcrossed 8 generations to the 129/sv background. This suggested that *Armc5* deletion caused embryonic lethality, with its severity depending on genetic background of the mice: embryonic lethality became more severe with higher degrees of genetic background purity. KO mice in the C57BL/6J × 129/sv F1 background were studied in subsequent experiments.

KO embryos were smaller than WT controls at embryonic day 14 (**Figure 2.3a**). These KO pups were smaller at age 8 – 12 weeks (**Figure 2.3b**). Body weight was significantly lower in KO and WT mice at age 4 and 8 weeks than in their WT littermates (**Figure 2.3c**). Both male and female KO mice weighed only about 60% as much as WT controls.

We examined serum growth hormone levels because of growth retardation in KO mice, but no significant difference was found between them and their WT counterparts (Supplementary **Figure 2.S.4**).

ARMC5 gene mutations have been reported to be linked with PMAH and Cushing's syndrome[1]. However, KO mice presented normal adrenal gland size and histology (Supplementary **Figure 2.S.5**) and serum glucocorticoid levels (Supplementary **Figure 2.S.6**) in young age (less than age 5 months). In old age (> 15 months), grossly, KO mice showed enlarged adrenal glands without apparent nodular structure, and histologically, there is no identifiable nodular hyperplasia (**Figure 2.3d**). Serum glucocorticoid levels were significantly increased in aged KO mice (**Figure 2.3e**), supporting the notion that the adrenal gland hyperplasia is of cortex in nature. It is to be noted that the mice were sacrificed between 12:30 – 1:30 pm, and their blood was harvested for the measurement of glucocorticoids, whose secretion is at the nadir at this time point. The moderate but significant increase of glucocorticoid levels in the KO mice is reminiscent of human PMAH, in which the increase of glucocorticoid levels is not drastic and is caused by the large mass of the adrenal gland, while on a per-cell basis, the secretion is reduced[1].
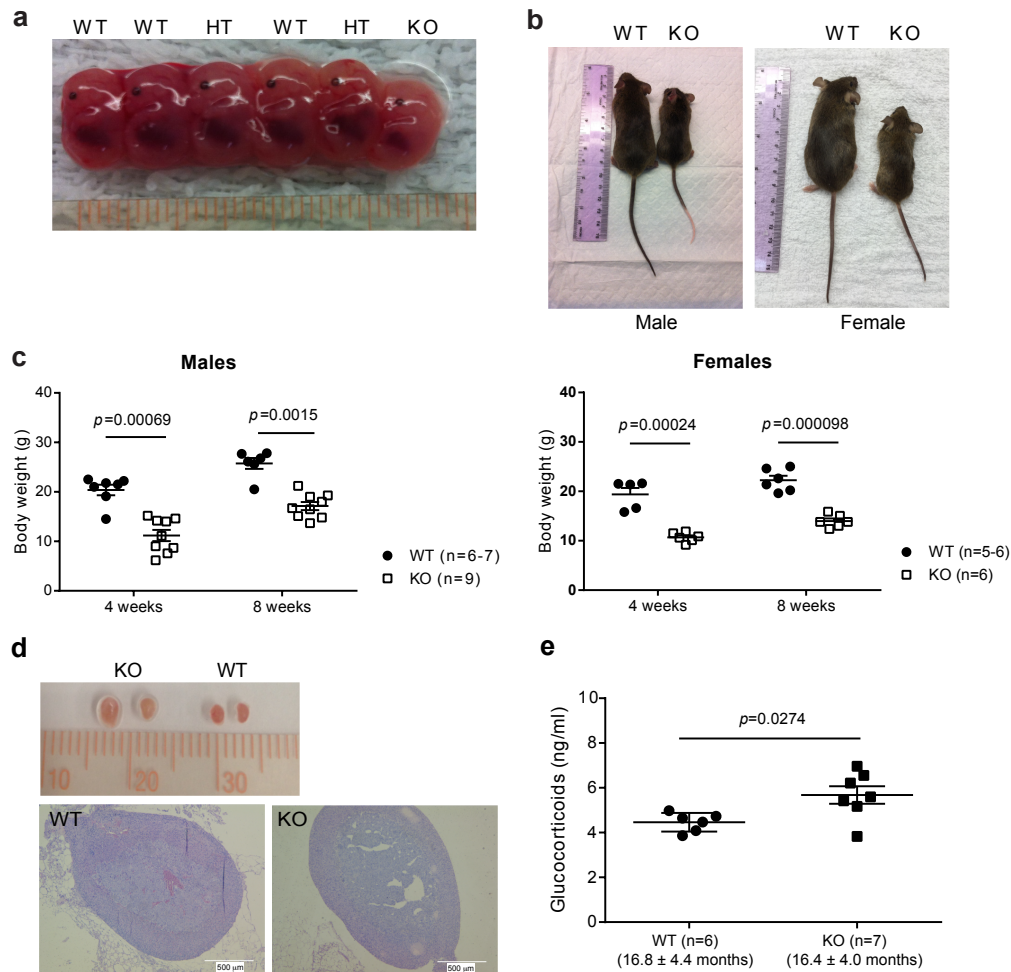
Fig. 2.3. General phenotype of KO mice (**a**) Representative photos of WT, HT, KO fetuses on embryonic day 14. (**b**) Representative photos of adult KO and WT littermates. Left panel: males (8 weeks old); right panel: females (12 weeks old). (**c**) Body weight (means ± SEM) of *Armc5* KO and WT littermates at age 4 and 8 weeks. Mouse numbers (n) per group are indicated. *$p < 0.001$ (two-tailed Student's t test). (**d**) Morphology (upper panel) and histology (lower panel, HE staining) of adrenal glands from old KO mice (19 months old). (**e**) Serum glucocorticoid levels in old KO mice. Means ± SEM of serum glucocorticoids in old KO and WT mice are shown. Age of each group (means ± SEM) and mouse number per group are indicated. Two-tailed Student's t test was used for statistical analysis.

2.3.4. *Armc5* KO phenotype in lymphoid organs and T cells

Thymus (Supplementary **Figure 2.S.7a**) and spleen (Supplementary **Figure 2.S.7b**) weight and cellularity were not significantly different in KO and WT mice. Moreover, thymocyte sub-populations (CD4$^+$CD8$^+$ double-positive, CD4$^+$ single-positive and CD8$^+$ single-positive cells) in the KO and WT thymus were comparable (Supplementary **Figure 2.S.7c**), as were spleen lymphocyte subpopulations (Thy1.2$^+$ T cells versus B220$^+$ B cells; CD4$^+$ versus CD8$^+$ T cells; Supplementary **Figure 2.S.7d**).

Despite seemingly normal T cell development in KO mice, T cell proliferation triggered by anti-CD3ε was compromised in both CD4$^+$ and CD8$^+$ cells (**Figure 2.4a**, left and middle panels; gating strategy: Supplementary **Figure 2.S.2c**). It is to be noted that activation markers CD25 and CD69 shortly after CD3 stimulation were drastically upregulated and were always comparable between WT and KO T cells (Supplementary **Figure 2.S.8**). The proliferation rate of KO B cells was also lower (**Figure 2.4a**, right panel; gating strategy: Supplementary **Figure 2.S.2d**). Cell cycle analysis revealed that G1/S progression was compromised in KO T cells (**Figure 2.4b**; gating strategy: Supplementary **Figure 2.S.2e**). We also demonstrated that KO T cells (gated on CD4$^+$ plus CD8$^+$ cells) presented increased FasL-triggered apoptosis (**Figure 2.4c**; gating strategy: Supplementary **Figure 2.S.2f**).

Naïve KO CD4$^+$ cells cultured under Th1 and Th17 conditions manifested reduced proliferation, as expected (**Figure 2.5a**; gating strategy: Supplementary **Figure 2.S.2g**). The differentiation of naïve CD4$^+$ cells into Th1 and Th17 cells was defective (**Figure 2.5b**; gating strategy: Supplementary **Figure 2.S.2g**), since the percentages of Th1 or Th17 cells were decreased among CD4$^+$ cells, which had already proliferated. The expression of transcription factors T-bet and RORγt -essential for Th1 and Th17 differentiation, respectively- were normal in KO CD4$^+$ cells cultured under Th1 and Th17 differentiation conditions (**Figure 2.5c and d**), when gated on either total CD4$^+$ cells or on those already differentiated cells (IFN-γ^+ or IL-17$^+$ cells), suggesting that the defective differentiation is not caused by a lack of these transcription factors.
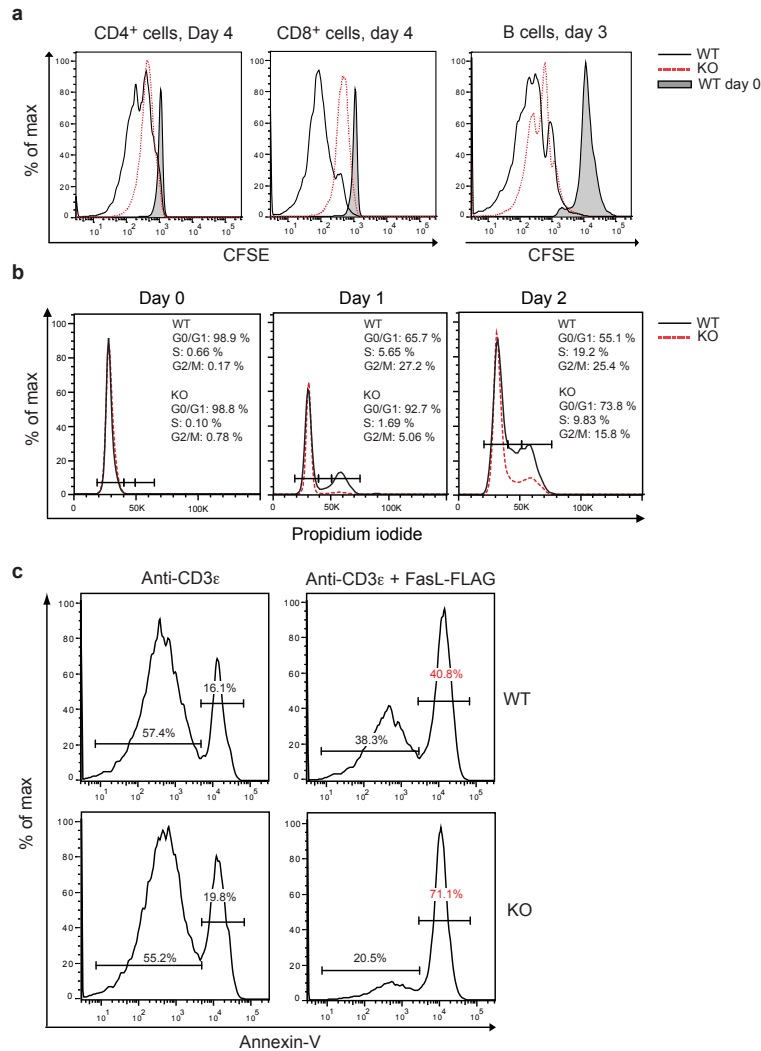
Fig. 2.4. KO T cell proliferation and apoptosis (**a**) Proliferation of spleen CD4$^+$ and CD8$^+$ T cells and B220$^+$ B cells from WT and KO mice according to CFSE staining. CFSE intensity was ascertained by flow cytometry. Experiments were conducted independently $4-6$ times. Representative histograms are shown. (**b**) Cell cycle progression of spleen T cells from WT and KO mice. The percentages of cells in G$_1$, S and G$_2$ phases are indicated. Experiments were conducted independently 3 times. Representative histograms are shown. (**c**) Apoptosis of WT and KO spleen T cells (gated on CD4$^+$ plus CD8$^+$ cells) upon FasL stimulation was determined by their annexin V expression according to flow cytometry. Experiments were conducted independently 3 times. Representative histograms are shown.

As for humoral immune responses, KO serum IgG levels were comparable to those of WT controls (Supplementary **Figure 2.S.9**).

We generated chimeric mice by transplanting KO and WT fetal liver cells in the C57BL/6J $\times$ 129/sv F1 background (CD45.2$^+$ single-positive) into lethally irradiated

C57BL/6J × C57B6.SJL F1 mice (CD45.1$^+$CD45.2$^+$ double-positive). Peripheral white blood cells of the recipients were analyzed by flow cytometry 8 weeks after transplantation, and recipients of similar degrees of KO and WT chimerism were paired for experimentation. Typically, about 80 – 85% of peripheral white blood cells were of donor origin (CD45.2$^+$ single-positive), and 12 – 15%, of recipient origin (CD45.1$^+$CD45.2$^+$ double-positive). In spleen Thy1.2$^+$ total T cells, CD4$^+$ T cells and CD8$^+$ T cells, 60 – 70% were of donor origin, and 30 – 35%, of recipient origin (Supplementary **Figure 2.S.10**). Unlike in *Armc5* KO mice, KO T cells in chimeras were developed in a WT environment, devoid of influence by putatively unknown factors which might exist in the total KO environment and have aberrant effects on T cell development.

We showed that donor-derived KO naïve CD4$^+$ cells were defective in differentiating into Th1 cells (**Figure 2.5e**; gating strategy: (Supplementary **Figure 2.S.2h**), similar to CD4$^+$ cells from unmanipulated, naïve KO mice (**Figure 2.5b**). The KO Th17 cell differentiation in this model was also compromised, although did not reach statistical significance, probably due to an inadequate sample size (**Figure 2.5e**).

2.3.5. Experimental autoimmune encephalomyelitis (EAE) in KO mice

To understand the role of ARMC5 in *in vivo* T cell immune responses, particularly CD4$^+$ cells-mediated immune responses, we induced experimental autoimmune encephalomyelitis (EAE) in *Armc5* WT and KO mice. As shown in **Figure 2.6a**, WT mice started to manifest clinical signs of EAE on day 13.2 ± 1.30 (means ± SEM) after immunization, and their symptoms peaked on day 23. The onset of clinical symptoms in KO mice was delayed by about 7 days, and their maximum disease score was significantly lower than that of WT controls (p < 0.01, two-tailed Student's *t* test) after day 18. Disease incidence was lower in KO mice between days 15 to 18, although it reached 100% in both KO and WT groups after day 28 (**Figure 2.6b**). A trend toward less body weight loss in KO mice was noted

after EAE induction compared to WT controls, but a statistically significant difference was reached only on day 22 (**Figure 2.6c**).

KO mice had significantly fewer cells in their draining LN and fewer infiltrating mononuclear cells in the brain and spinal cords on day 14 after MOG immunization compared to WT controls (**Figure 2.6d**). After *ex vivo* PMA/ionomycin stimulation, the percentage of
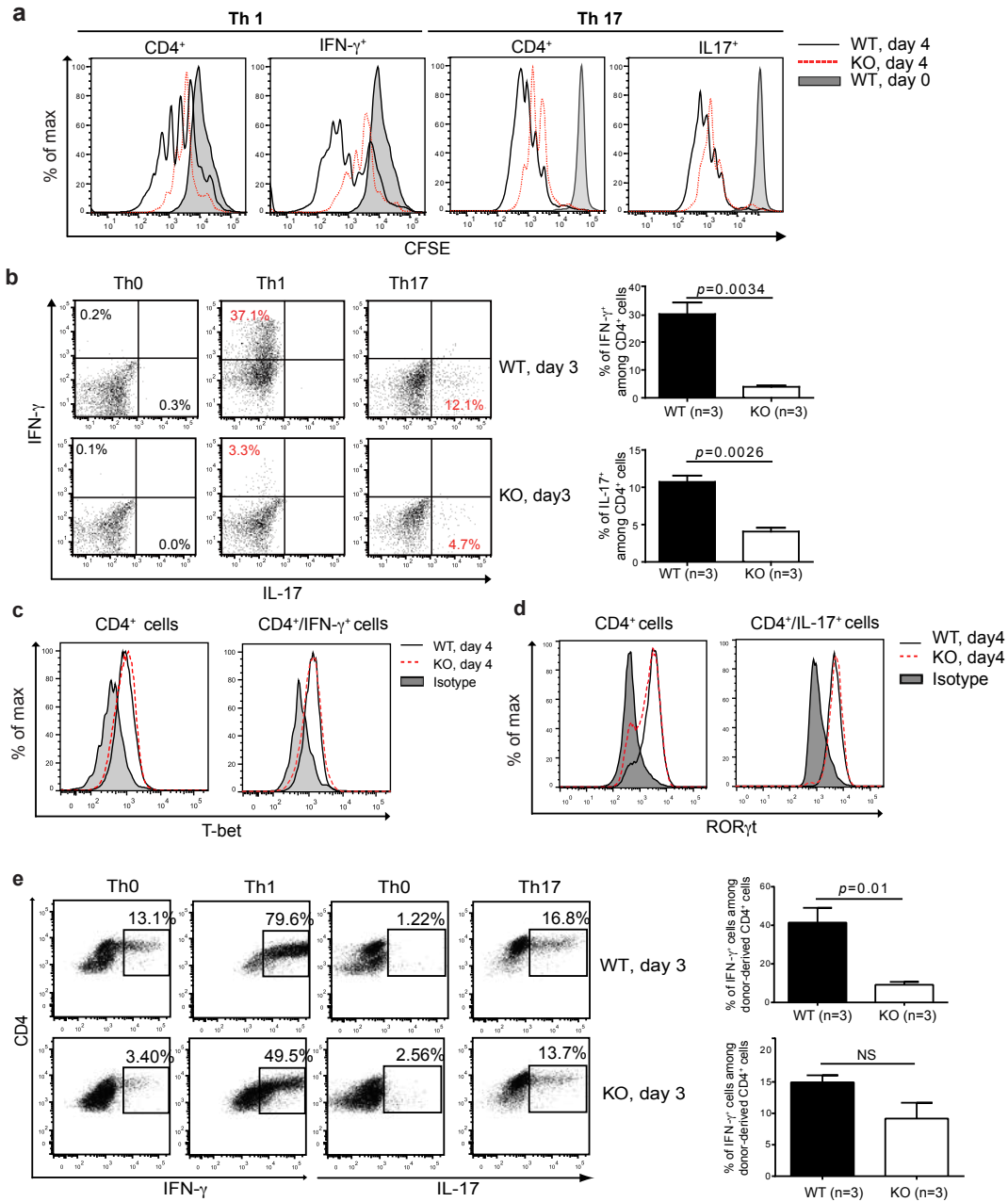
Fig. 2.5. Proliferation and differentiation of naïve KO CD4$^+$ cells into Th1 and Th17 cells *(see next page)*

Fig. 2.5 *(previous page).* (**a**) Proliferation of WT and KO naïve spleen CD4$^+$ cells under Th1 and Th17 conditions was assessed based on CFSE content according to flow cytometry. Experiments were conducted 3 times, and representative histograms are shown. Grey peaks represent the CFSE content of CD4$^+$ cells at day 0. (**b**) These cells' differentiation into Th1 and Th17 cells was also determined by flow cytometry according to intracellular IFN-γ and IL-17 positivity (gated on total CD4$^+$). Representative dot plots are shown in the left panel. Means $\pm$ SEM of data from 3 experiments are presented as bar graphs in the right panel. Mouse numbers (n) per group are indicated. *p*-values are reported in the bar graphs (two-tailed Student's *t* test). (**c and d**) T-bet and RORγt expression in CD4$^+$ cells cultured under Th1 and Th17 conditions or in IFNγ^+ or IL-17$^+$ cells was determined by flow cytometry. Experiments were conducted 3 times. Representative histograms are shown. (**e**) Th1 and Th17 differentiation of naïve spleen CD4$^+$ cells (CD45.2 single-positive) derived from WT and KO donors in chimeric mice was analyzed by flow cytometry based on their intracellular IFN-γ and IL-17 expression. Representative dot plots are shown in the left panel. Means $\pm$ SEM of data from 3 experiments are presented as bar graphs in the right panel. Mouse numbers (n) per group are indicated. *p*-values are reported in the bar graphs (two-tailed Student's *t* test).

IFN-γ^+ CD4$^+$ cells among total CD4$^+$ cells from the LN of KO mice was significantly lower than that of WT mice (6.2% versus 15.4%), although the percentage of IL-17$^+$ cells among CD4$^+$ cells was similar in KO and WT draining LN (**Figure 2.6e**). The percentages of IFN-γ^+ and IL-17$^+$ populations in CD4$^+$ T cells from the central nervous system (CNS) of KO mice were significantly lower after *ex vivo* PMA/ionomycin stimulation than in WT mice (**Figure 2.6f**).

Histologically, spinal cords from KO animals on day 30 after MOG immunization showed less severe mononuclear cell infiltration and demyelination, according to hematoxylin/eosin and Luxol Fast Blue staining, respectively, compared to their WT counterparts (**Figure 2.6g**). Histological data from 4 KO and 5 WT spinal cords are summarized (**Figure 2.6h**). Mononuclear cell infiltration in KO spinal cords was significantly lower than in WT controls. Although demyelination in the former was also lower, it did not reach statistical significance. However, combined pathological scores, which included degrees of both mononuclear cell infiltration and demyelination, were significantly lower in KO mice. We did not observe changes in the percentages of Treg cells in the spleen of naïve KO mice or in the draining LN of KO mice on day 17 during EAE induction, compared to WT controls

(**Figure 2.6i**; gating strategy: Supplementary **Figure 2.S.2i**). Therefore, it is unlikely that Treg cells are implicated in reduced EAE severity in KO mice.

To exclude the possible influence of the *Armc5* KO background on the immune system of KO mice, EAE was also induced in chimeras transplanted with fetal liver cells from *Armc5* WT and KO embryos on day 13 – 15. Overall, KO chimeras still displayed a lower degree of EAE than WT chimeras, but the difference was not as dramatic as in real KO versus WT mice. The onset of clinical symptoms in KO chimeras occurred 2.5 days (mean) later than in WT chimeras. KO chimera clinical scores tended to be lower than those of WT mice, but were only significantly different between day 12 and 14 (**Figure 2.6j**). EAE incidence was significantly lower on days 11, 12 and 14 after immunization (**Figure 2.6k**). A trend of less body weight loss was noted in KO chimeric mice, although no statistical difference was apparent between the KO and WT groups (**Figure 2.6l**). When stimulated *ex vivo* by PMA/ionomycin, the percentage of KO donor-derived IFN-γ^+ CD4$^+$ cells among total KO donor-derived CD4$^+$ cells from the CNS was significantly lower than in WT controls (**Figure 2.6m**; gating strategy: Supplementary **Figure 2.S.2h**), as was the case in real KO mice. However, there was no significant difference between the percentage of KO donor-derived IL-17$^+$ CD4$^+$ cells among total KO donor-derived CD4$^+$ cells and that of WT mice. The reduced degree of difference in EAE manifestation in KO versus WT chimeras, compared to that in real KO versus WT mice, was not unexpected, as KO chimeras contained about 30% recipient-derived T cells, which were fully immuno-competent WT T cells.

2.3.6. Antiviral immune responses in KO mice

CD8$^+$ T cell-mediated immune responses play a critical role against lymphocytic chori-omeningitis virus (LCMV) infection. Therefore, we assessed KO CD8$^+$ T cell functions in LCMV infection. Eight days after mice were infected with LCMV (strain WE), absolute numbers of WT CD8$^+$ T cells but not CD4$^+$ T cells increased significantly (**Figure 2.7a**; gating strategy: Supplementary **Figure 2.S.2j**). LCMV tetramer staining showed that both

the number of gp_{33-41}-, $np_{396-405}$- and $gp_{276-286}$-specific $CD8^+$ T cells per spleen and their percentage among total spleen $CD8^+$ T cells were significantly lower in KO than in WT mice (**Figure 2.7b, c, and d**; gating strategy: Supplementary **Figure 2.S.2k**), suggesting compromised $CD8^+$ cell clonal expansion after viral Ag stimulation in KO mice.
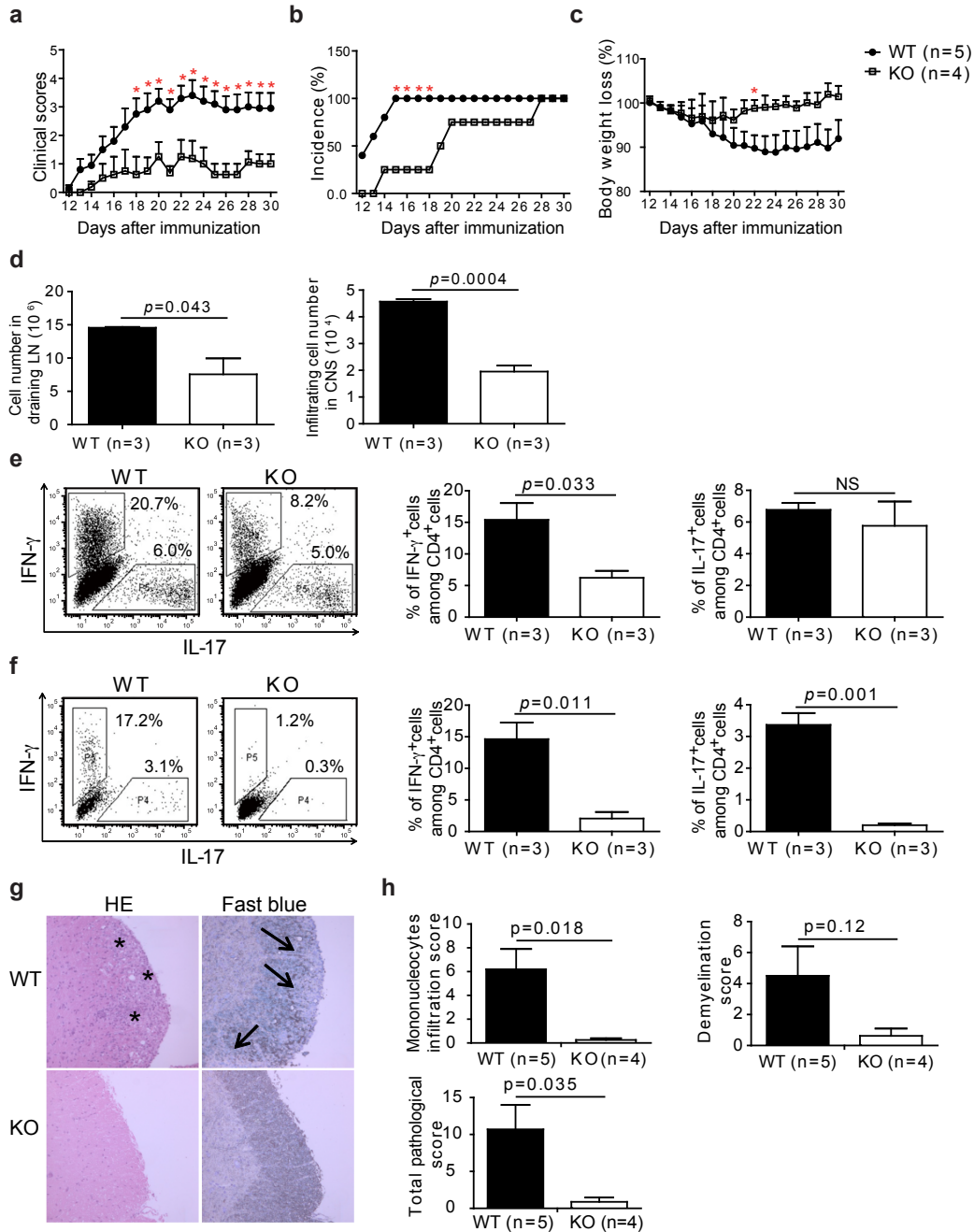
Fig. 2.6. EAE induction in KO mice *(see next page)*

Fig. 2.6 *(previous page)*. (**a**) Means $\pm$ SEM of EAE clinical scores of KO and WT mice. *$p<0.05$ (two-tailed Student's t test). (**b**) EAE incidence in KO and WT mice. *$p<0.05$ (chi-square test). (**c**) Means $\pm$ SEM of body weight of KO and WT mice during EAE induction. Body weight of mice on day 10 post-immunization was considered as 100%. *$p < 0.05$ (two-tailed Student's t test). (**d**) Means $\pm$ SEM of cellularity in draining LN and of cells infiltrating the CNS of mice 14 days after MOG immunization. Mouse numbers (n) and p-values (paired two-tailed Student's t test) are indicated. (**e** and **f**) Cytokine-producing cells among CD4$^+$ cells from draining LN (**e**) and CNS (**f**) on days 13-18 after MOG immunization. Left panels: representative dot plots; right panel: bar graphs (means + SEM) summarizing all the results, with mouse numbers and p-values (two-tailed Student's t test) indicated. (**g**) HE (left column) or Luxol Fast Blue (right column) staining of spinal cords 30 days after MOG immunization. Asterisks indicate cell infiltration. Arrows point to demyelination. (**h**) Means $\pm$ SEM of mononuclear cell infiltration scores, demyelination scores, and total pathological scores, which is the sum of the first 2 scores. Mouse numbers (n) and p-values (two-tailed Student's t test) are indicated.

After infection, CD8$^+$ cells develop into KLRG1^{hi}CD127^{lo} short-lived effector cells (SLEC) and KLRG1^{lo}CD127^{med} memory precursor effector cells (MPEC)[320]. In KO mice, 8 days after LCMV infection, the percentage of SLEC among CD8$^+$ T cells was significantly lower (**Figure 2.7e and f**; gating strategy: Supplementary **Figure 2.S.2k**), indicating defective anti-virus effector cell development. At the same time, MPEC percentage among CD8$^+$ T cells was increased in KO mice. The significance of this finding is not clear at present, although the percentage of CD62L^{lo}CD44^{hi} effector memory cells among total CD8$^+$ cells (Supplementary **Figure 2.S.11a**) and LCMV subdominant epitope (np$_{396\text{-}405}$ and gp$_{276\text{-}286}$)-specific CD8$^+$ T cells (**Figure 2.7g**; gating strategy, Supplementary **Figure 2.S.2k**) in KO mice was reduced.

We next examined the presence of LCMV-specific, cytokine-producing splenic T cells in virus-infected mice. As seen in **Figure 2.7h** (gating strategy: Supplementary **Figure 2.S.2j**), the absolute number of gp$_{33\text{-}41}$-specific TNF-α-positive CD8$^+$ T cells, IFN-γ-positive CD8$^+$ T cells, and IFN-γ/TNF-α double-positive CD8$^+$ T cells per spleen was significantly lower in KO than in WT mice 8 days post-infection. Significantly lower percentages of gp$_{33\text{-}41}$-specific, IFN-γ-positive CD8$^+$ T cells and IFN-γ/TNF-α double-positive CD8$^+$ T cells, but not TNF-α-positive CD8$^+$ T cells, among total spleen CD8$^+$ T cells,
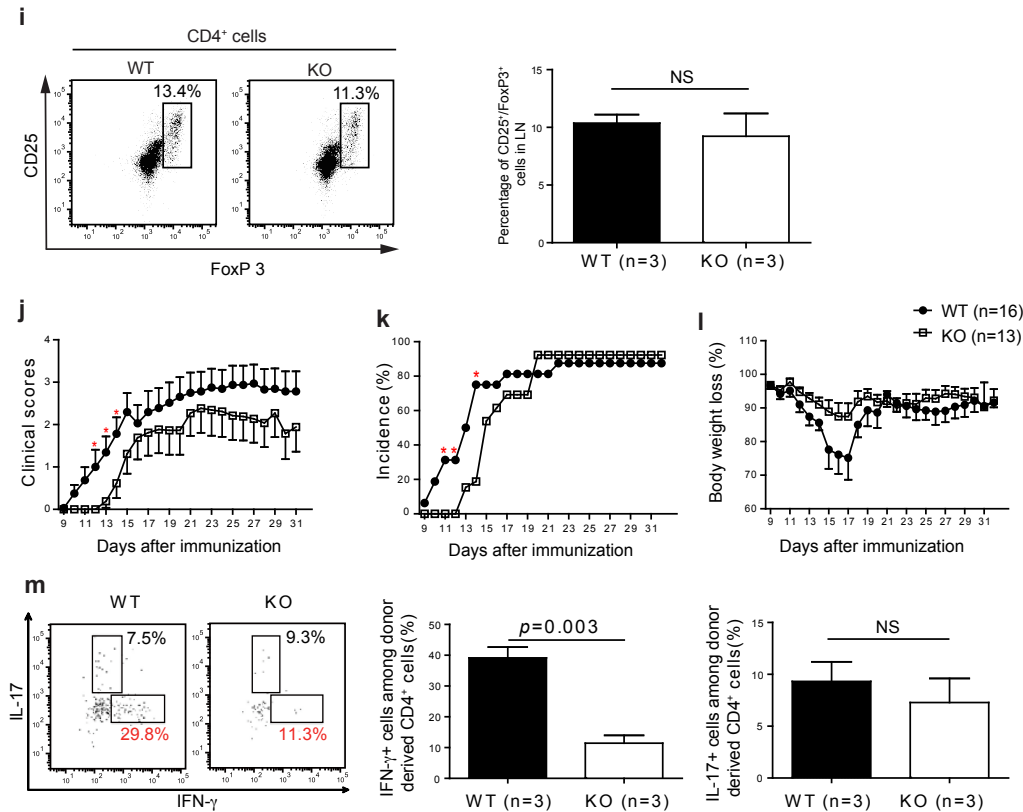
Fig. 2.6. *(continued).* **EAE induction in KO mice** (**i**) Treg cells in naïve KO mice on day 17 during EAE induction. Left panel; representative dot plots; right panel: means + SEM of data from 3 experiments. NS: not significant (two-tailed Student's t test). (**j**) Means $\pm$ SEM of EAE clinical scores of chimeric mice. *$p < 0.05$ (two-tailed Student's t test). (**k**) EAE incidence in chimeric mice. *$p < 0.05$ (chi-square test). (**l**) Means $\pm$ SEM of body weight of chimeric mice, with body weight on day 10 after MOG immunization considered as 100%. No significant difference is found (two-tailed Student's t test). (**m**) Cytokine-producing donor-derived CD4$^+$ cells in the CNS of chimeric mice on day 14 after MOG immunization. Left panel: representative dot plots; right panel: summary (means $\pm$ SD) of all the results, with mouse numbers (n) and p-values (paired two-tailed Student's t test) indicated.

were found in KO spleens (**Figure 2.7i**; representative dot plots shown in Supplementary **Figure 2.S.2j**). Similarly, decreased numbers and percentages of LCMV-specific cytokine-producing cells were observed in the CD4$^+$ cell population, although the reduction was of lower magnitude compared to those in the CD8$^+$ cells (**Figure 2.7j**; representative dot plots shown in Supplementary **Figure 2.S.2k**).

In addition, lower percentages of gp$_{33\text{-}41}$-specific CD107a$^+$GranB$^+$ T cells among total CD8$^+$ T cells were observed in KO spleen (**Figure 2.7k**; representative dot plots shown

in Supplementary **Figure 2.S.11**), implying the presence of fewer functional virus-specific cytotoxic CD8$^+$ T cells in KO mice. Virus titers in the kidneys, liver and spleen were significantly higher in KO mice 8 days post-LCMV infection, suggesting compromised virus clearance (**Figure 2.7l**).
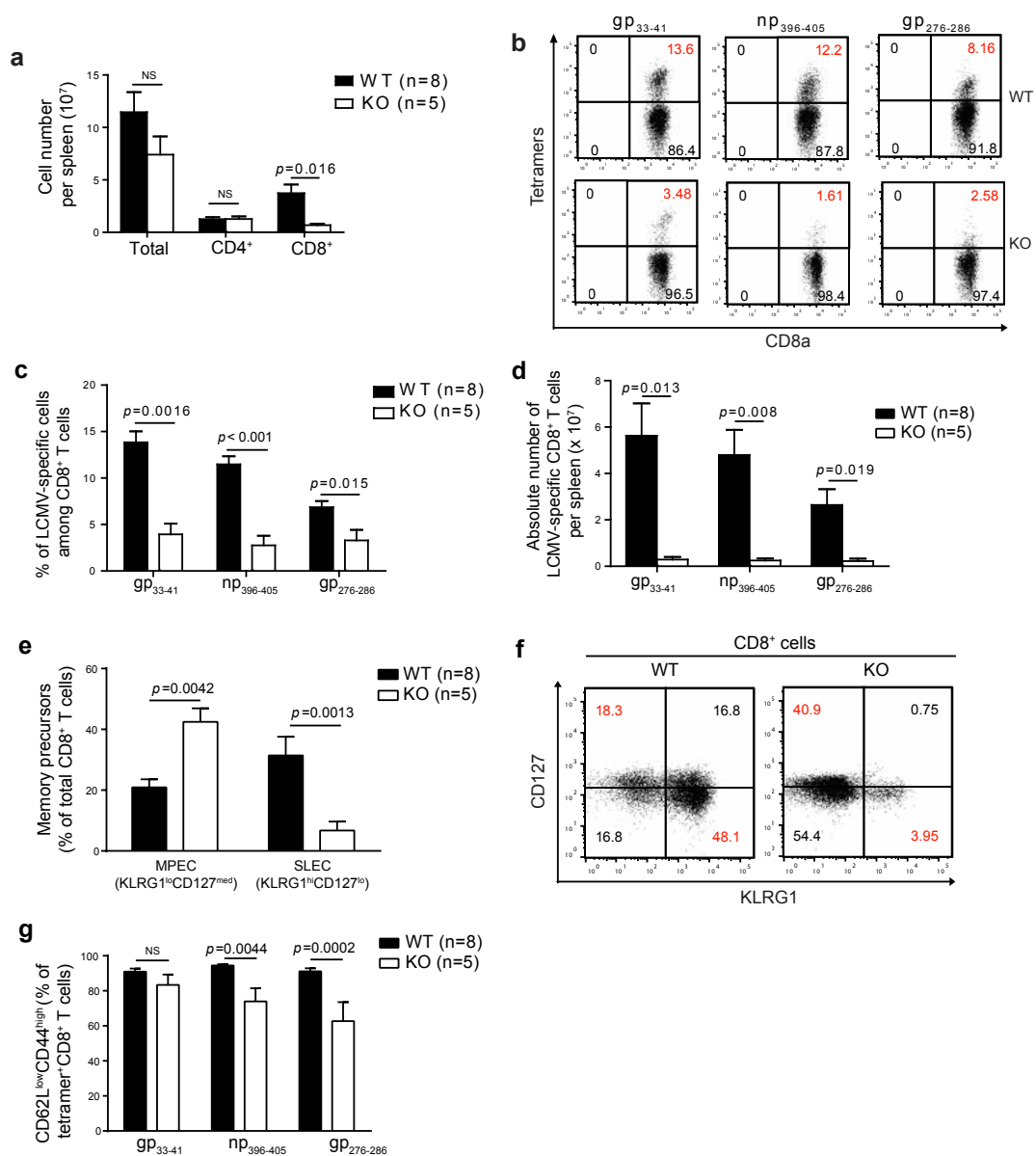
Fig. 2.7. Anti-LCMV immune responses in KO mice *(see next page)*

Fig. 2.7 *(previous page).* (**a**) Spleen CD8$^+$ cell numbers in KO mice on day 8 after LCMV infection as determined by flow cytometry. Mice number (n), means $\pm$ SEM and p-values (two-tailed Student's t test) are indicated. (**b**) Virus-specific spleen CD8$^+$ cells in KO mice on day 8 post-LCMV infection according to flow cytometry. Representative dot plots are shown. (**c**) Means $\pm$ SEM of percentages of gp$_{33\text{-}41}$, np$_{396\text{-}405}$ and gp$_{276\text{-}286}$ tetramer-positive cells among spleen CD8$^+$ cells from all the results are presented. Numbers (n) of mice per group and p-values (two-tailed Student's t test) are indicated. (**d**) Means $\pm$ SEM of absolute numbers of gp$_{33\text{-}41}$, np$_{396\text{-}405}$ and gp$_{276\text{-}286}$ tetramer-positive CD8$^+$ cells in the KO and WT mouse spleens on day 8 post-infection. Numbers (n) of mice per group and p-values (two-tailed Student's t test) are indicated. (**e and f**) Memory and effector CD8$^+$ cell maturation in LCMV-infected WT and KO mice on day 8 post-LCMV infection. KLRG1$^{\text{lo}}$CD127$^{\text{hi}}$ cells are considered as memory precursor effector cells (MPEC), and KLRG1$^{\text{hi}}$CD127$^{\text{lo}}$ cells as short-lived effector cells (SLEC). Means $\pm$ SEM are presented. Numbers (n) of mice per group and p-values (two-tailed Student's t test) are indicated (**e**). Representative dot plots are shown (**f**). (**g**) On day 8 post-infection, total gp$_{33\text{-}41}$ np$_{396\text{-}405}$ and gp$_{276\text{-}286}$ tetramer-positive CD8$^+$ cells in KO and WT mouse spleen were assessed for activation markers. Means $\pm$ SEM are presented. Numbers (n) of mice per group and p-values (two-tailed Student's t test) are indicated.

2.3.7. Identification of ARMC5-binding proteins by Y2H assay

ARMC5 has no enzymatic activity: its functions depend on interaction with molecules involved in different signaling pathways. To identify ARMC5-binding proteins, we conducted Y2H assays with human ARMC5 protein (Glu30 – Ala935) as bait, and a human primary thymocyte cDNA expression library as prey. The binding proteins were given Predicted Biological Confidence (PBC) scores[321], and 16 proteins with scores between A and D ("A" having the highest confidence of binding) are found **Table 2.1**, if their coding sequences are in-frame and have no in-frame stop codons. A complete list of binding proteins identified by Y2H assay and a map showing the interaction regions between ARMC5 and its binding partners are provided in the Supplementary materials section (Supplementary **Table 2.S.2** and Supplementary **Figure 2.S.12**).
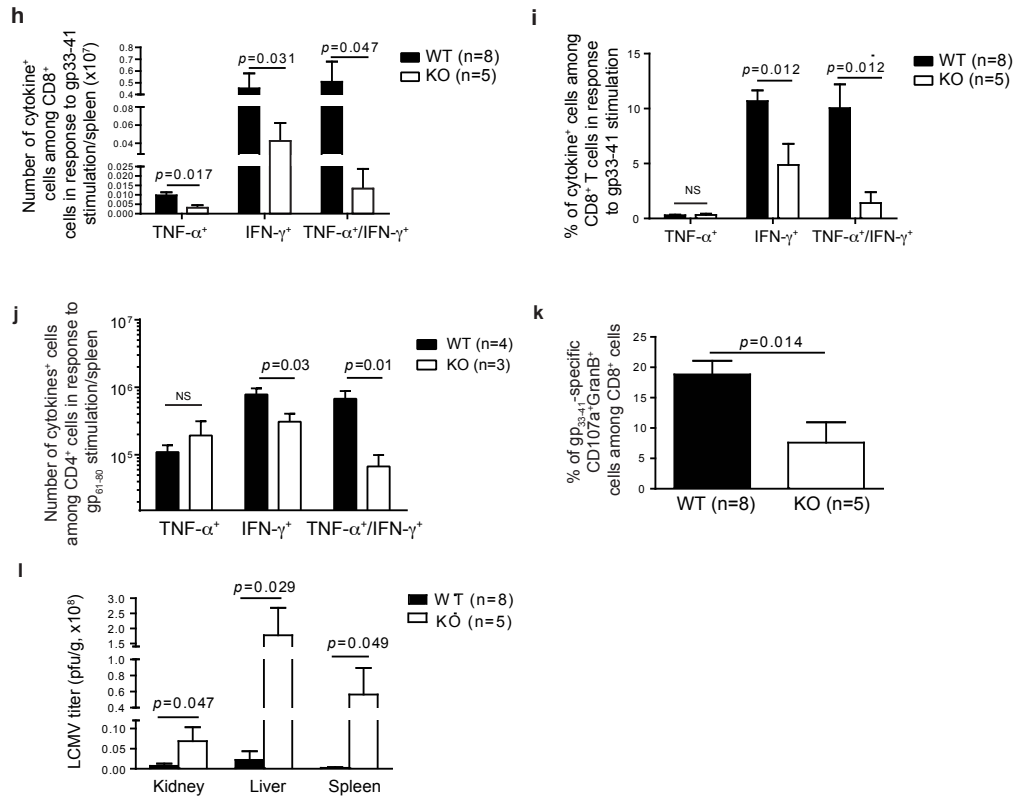
Fig. 2.7. Anti-LCMV immune responses in KO mice (**h**) Absolute number of virus-specific, cytokine-producing $CD8^+$ cells. (**i and j**) Percentages of virus-specific, cytokine-producing cells among $CD8^+$ cells (**i**) and $CD4^+$ cells (**j**) on day 8 post-LCMV infection. Means $\pm$ SEM of data are shown. Mouse numbers (n) per group and *p*-values (two-tailed Student's *t* test) are indicated. (**k**) Means $\pm$ SEM of percentages of gp_{33-41}-specific $CD107a^+GranB^+$ $CD8^+$ T cells on day 8 post-LCMV infection. Mouse numbers (n) per group and *p*-values (two-tailed Student's *t* test) are indicated. (**l**) Means + SEM of viral titers in the kidney, liver and spleen on day 8 post-LCMV infection. Mouse numbers (n) per group and *p*-values (two-tailed Student's *t* test) are indicated.

2.4. Discussion

Our study demonstrated that *Armc5* mRNA was highly expressed in the thymus and adrenal glands. Its deletion led to small body size in mice and compromised T cell proliferation and differentiation. KO mice presented defective induction of EAE and anti-LCMV immune responses. KO mice developed adrenal gland hyperplasia in old age. ARMC5 is a protein without enzymatic activity. Our Y2H assays identified 16 candidate ARMC5-binding

Table 2.1. ARMC5-binding proteins identified by Y2H assay

Gene name	PBC score	Binding clones	Different clones	Major known function
DAPK1	A	13	4	Tumor suppressor, apoptosis, autophagy
ARMC5	B	2	2	Self- dimerization
STK24	B	3	3	Apoptosis, upstream of MAPK, acts on Tao
TTF1	B	3	2	Transcription terminator, apoptosis, tumor risk
POLR2A	B	3	2	DNA-directed RNA polymerase II subunit RPB1
CUL3	C	2	2	E3 component, WNK degradation, BTB domain, cell cycle, cyclin E degradation
CDCA7L	D	1	1	Cell cycle, transcription co-activator, c-Myc interactor, FoxP3-binding
C10orf46 (CACUL1)	D	1	1	CDK2-associated, cell cycle, promotes prolifereation
E2F2	D	1	1	Cell cycle, transcription factor, T cell quiescence
FAM65B	D	1	1	Skeletal muscle development, hearing
FLJ20105 (PICH)	D	1	1	cell division
HUWE1	D	6	1	Ubiquitination and proteasomal degradation, Base-excision repair, neural differentiation and proliferation
KIF11	D	1	1	ATP-dependent microtubule motor activity
PCBP1	D	2	1	Cadherin binding, involved in cell-cell adhesion, Burkitt lymphoma
RPN2	D	1	1	Endopeptidase activity, ubiquitin-dependent protein catabolic process
TCF12	D	1	1	Immune response, regulation of transcription
ZBTB40	D	2	1	Bone mineralization, cellular response to DNA damage stimulus

Note: Y2H assays were performed by Hybrigenics Services (Paris, France). The coding sequence for human *ARMC5* cDNA (aa 30-935) served as bait to screen a random-primed human thymocyte cDNA library. Eighty million yeast clones (8-fold the complexity of the library) were screened. One hundred and sixty-five His$^+$ colonies were selected. The prey fragments of positive clones were amplified by PCR and sequenced at their 5' and 3' junctions. The resulting sequences were used to identify corresponding interacting proteins in the GenBank database via a fully-automated procedure. A Predicted Biological Confidence (PBC) score (from A-F; A being of very high confidence in the interaction and F being experimentally-proven artifacts) was attributed to each interaction. Sixteen proteins with PBC scores between A and D are listed, if their coding sequences are in-frame and have no in-frame stop codons. Binding clones: number of total clones interacting with the bait. Different clones: number of different clones of the same cDNA interacting with the bait. Known functions of the prey proteins are described.

proteins potentially capable of linking ARMC5 to different signaling pathways involved in cell cycling and apoptosis.

Armc5 expression at the mRNA level was upregulated immediately (within 2 h) after CD4$^+$ T cell activation by TCR ligation and less so and at a slower pace in CD8$^+$ cells. Its expression level then declined in the following days (**Figure 2.1d**). *Armc5* expression in CD4$^+$ cells cultured under Th1 or Th17 conditions after 1 day remained low (Supplementary **Figure 2.S.13**), and was not influenced by the presence of different lymphokines, such as IL-2, IL-6, or TGF-β1 (Supplementary **Figure 2.S.14**). *Armc5* mRNA expression in CD8$^+$ cells 8 days after LCMV infection was significantly lower than in naïve CD8$^+$ cells (Supplementary **Figure 2.S.15**). These data suggest that this molecule is probably important in the early stage of TCR-triggered T cell activation to prepare cells for entry into the cell cycle. This notion is supported by cell cycle analysis, which revealed that KO T cells were compromised in G1/S progression (**Figure 2.4b**).

We found reduced numbers of infiltrating T cells as well as Th1 (IFN-γ^+) and Th17 (IL-17$^+$) cells in the CNS of KO EAE mice compared to WT EAE controls. Such decreases were likely responsible for the diminished EAE manifestations in KO mice. Reduced CNS lymphocyte infiltration could be caused by compromised clonal expansion/differentiation of T cells in the periphery, defective migration of such cells into the CNS, reduced expansion/differentiation of these cells in the CNS, decreased apoptosis of cells in the periphery and CNS, or all of the above. Defective KO T cell clonal expansion/differentiation in the periphery was apparent according to our *in vitro* and *in vivo* results (**Figure 2.4** and **Figure 2.7b**), but whether this is also the case in the CNS remains to be studied.

We demonstrated that *Armc5* deletion resulted in compromised TCR-stimulated proliferation of both CD4$^+$ and CD8$^+$ T cells *in vitro* and LCMV-specific CD8$^+$ T cell clonal expansion *in vivo*. Moreover, we observed a significant reduction in SLECs in KO mice following LCMV infection while MPECs were increased. Taken together, these results suggest a function for ARMC5 in promoting T cell growth following TCR engagement possibly by regulating activation threshold levels; this provides a potential explanation for the observed increase in T cell death following FasL engagement (**Figure 2.4c**) in KO mice. It is possible that augmented apoptosis also plays a role in compromised Th1 and Th17 development

from naïve KO CD4$^+$ cells. However, *ARMC5* mutations lead to PMAH in humans and diffuse adrenal gland hyperplasia in mice (**Figure 2.3d**), indicating that it has a default function of repressing adrenal cell proliferation, or a default pro-apoptotic function, or both. Indeed, an *in vitro* study of the human adrenal gland cell line H295R revealed that ARMC5 overexpression culminates in apoptosis[1], supporting its putative default anti-apoptotic function in adrenal glands. Dichotomous functions of ARMC5 in T cells versus adrenal glands indicate its tissue- or context-specificity, likely due to ARMC5's association with different binding partners. In different types of cells, ARMC5 might preferentially bind to a certain partner, depending on its relative abundance in a given cell type or cell status. Consequently, the default function of ARMC5 in certain types of cells or cells with a given status could be either pro- or anti-proliferation, pro- or anti-apoptosis, or neutral. It could explain the different phenotypes seen in T cells versus adrenal glands, in terms of proliferation and apoptosis. It could also explain the obvious dilemma that KO T cell development in the thymus, which involves fast thymocyte proliferation, is normal, but TCR-stimulated T cell proliferation/differentiation and virus-induced T cell clonal expansion are defective in KO mice.

Although B cells were not the focus of this study, we did demonstrate that KO B cells were compromised in proliferation triggered by BCR ligation. Although KO mice had normal serum IgG levels, it is possible that, under strenuous conditions, KO mice might manifest defective humoral immune responses.

Based on the functional results of our ARMC5 study, those from PMAH investigations, and protein association information from Y2H assays, we propose the following speculative model of ARMC5 mechanisms of action. ARMC5 transcription and protein expression are increased when the cells are activated. Induced ARMC5 forms dimers (or multimers) in cytosol. Such dimers are able to interact with different molecules in pathways regulating cell cycling and apoptosis, e.g., CUL3 for cell cycling, and DAPK1 for apoptosis. Therefore, depending on the relative abundance of binding proteins in different cell types and cells in different states, ARMC5 may interact preferentially with one or the other, leading to

opposite functions in regulating cell proliferation and apoptosis. It should be noted that list of associating proteins might expand, pending further verification.

In summary, we demonstrated that ARMC5 has vital functions in fetal development, T cell biology, immune responses and adrenal gland biology. We have created *Armc5* KO mice as the first animal model of a rare human disease: PMAH. Our mechanistic study to identify ARMC5-binding partners has laid the groundwork for further elucidation of ARMC5's mechanisms of action. With a better understanding of these mechanisms, this molecule may be deployed as a therapeutic target in immune and endocrine disorders.

2.5. Materials and Methods

2.5.1. ISH

To localize *Armc5* mRNA, 1526 bp (starting from GATATC to the end) mouse *Armc5* cDNA (GenBank: BC032200, cDNA clone MGC: 36606) in pSPORT1 was employed as template for S and AS riboprobe synthesis, with SP6 and T7 RNA polymerase for both ^{35}S-UTP and ^{35}S-CTP incorporation[322].

Tissues from WT mice were frozen in -35 °C isopentane and kept at -80 °C until they were sectioned. ISH, X-ray and emulsion autoradiography focused on 10 μm thick cryostat-cut sections. Briefly, overnight hybridization at 55 °C was followed by extensive washing and digestion with RNase to eliminate non-specifically bound probes. Anatomical level images of ISH were generated using X-Ray film autoradiography after 4 days' exposure. Microscopical level ISH was produced by dipping sections in NTB-2 photographic emulsion (Kodak). The exposure time was 28 days. The autoradiography labelling was revealed by D19 Developer (Kodak) and fixation with 35% sodium thiosulphate. Slides were left unstained or slightly stained with haematoxylin/eosin[322].

2.5.2. RT-qPCR

Armc5 mRNA in thymocytes, T cells, B cells and tissues from KO and WT mice was measured by RT-qPCR. Total RNA was extracted with TRIzol (Invitrogen, Carlsbad, CA,

USA) and then reverse-transcribed with Superscript II reverse-transcriptase (Invitrogen). Thymocytes were stained with anti-CD4 (1:400, Clone RM4-5, BD Bioscience), anti-CD8 (1:400, Clone 53-6.7, BioLegend), anti-CD25 (1:200, Clone PC61.5, eBioscience) and anti-CD44 (1:200, Clone IM7, BioLegend) Abs. CD4$^+$ cells, CD8$^+$ cells, DP cells, DN cells and DN cells in different stages were sorted by flow cytometry. T cells and B cells were isolated by magnetic beads (EasySep, Stem Cell Technology, Vancouver, BC, Canada). For RT-qPCR measurement of *Armc5* expression during T cell activation, mouse T-activator CD3/CD28 Dynabeads (ThermoFisher Scientific, Burlington, ON, Canada) were used for T cell activation *in vitro*, to avoid introducing Ag-presenting cells into purified CD4$^+$ or CD8$^+$ cells.

Forward and reverse primers were 5'-CAG TTA TGT GGT GAA GCT GGC GAA-3' and 5'-ACC CTC AGA AAT CAG CCA CAA CCT-3', respectively. A 139-bp product was detected with the following amplification program: 95 °C × 15 min, 1 cycle; 95 °C × 10 s, 59 °C × 15 s, 72 °C × 25 s, 35 cycles. *β-actin* mRNA levels were measured as internal controls. Forward and reverse primers were 5'-TCG TAC CAC AGG CAT TGT GAT GGA-3' and 5'-TGA TGT CAC GCA CGA TTT CCC TCT-3', respectively, with the same amplification program as for *Armc5* mRNA. The data were expressed as ratios of *Armc5* versus *β-actin* signals.

2.5.3. ARMC5 overexpression in L cells

L cells (CRL-2648, ATCC) were transiently transfected with pReceiver-Lv120 plasmid expressing mouse *Armc5* with HA tag (EX-Mm23477-LV120, GeneCopoeia, Rockville, MD, USA) for 2 days, and fixed with 4% paraformaldehyde. Subcellular ARMC5 localization in L cells was detected by immunofluorescence with biotinylated rat anti-HA Ab (1:500, 12158167001, Roche, Laval, QC, Canada), followed by FITC-conjugated streptavidin (1:2000, S11223, ThermoFisher, Burlington, ON, Canada). The L cells were not authenticated, and possible mycoplasma contamination was not tested.

2.5.4. Generation of *Armc5* KO mice

A PCR fragment amplified from the *Armc5* cDNA sequence served as probe to isolate genomic BAC DNA clone 7O8 from the RPCI-22 129/sv mouse BAC genomic library. The targeting vector was constructed by recombination and routine cloning methods, with a 15-kb *Armc5* genomic fragment from clone 7O8 as starting material. A 2.7-kb HindIII/EcoRV genomic fragment containing exon 1 – 3 was replaced by a 1.1-kb Neo cassette from pMC1Neo-Poly A flanked by 2 diagnostic restriction sites, EcoRV, and HindIII, as illustrated in **Figure 2.2a**. The final targeting fragment was excised from its cloning vector backbone by NotI/EcoRI digestion and electroporated into R1 embryonic stem (ES) cells for G418 selection. Targeted ES cell clones were injected into C57BL/6J blastocysts. Chimeric male mice were mated with C57BL/6 females to establish mutated *Armc5* allele germline transmission.

Southern blotting with probes corresponding to 5' and 3' sequences outside the targeting region, as illustrated in **Figure 2.2a** (red squares), screened for gene-targeted ES cells and eventually confirmed gene deletion in mouse tail DNA. With the 5' probe, the targeted allele presented a 6.6-kb EcoRV band, and the WT allele, a 9.3-kb EcoRV band. With the 3' probe, the targeted allele presented an 8.7-kb HindIII band, and the WT allele, a 12.5-kb HindIII band (Supplementary **Figure 2.S.3**).

Heterozygous mice were backcrossed to the C57BL/6J background for 8 generations and then crossed with 129/sv mice. WT and KO mice in the C57BL/6J × 129/sv F1 background were studied. All animals were housed under specific pathogen-free conditions and handled in accordance with a protocol approved by the Institutional Animal Protection Committees of the CRCHUM and INRS-IAF.

2.5.5. Serum total IgG measurement

Flat bottom 96-well plates (Costar EIA/RIA, No. 3369, Fisher Scientific) were coated with goat anti-mouse IgG (100 μl/well, 1 μg/ml in PBS) and incubated overnight at 4 °C. After five times of washings with PBS containing 0.05% Tween 20, the plates were blocked with PBS containing 3% BSA and 5% FBS for 1.5 h at room temperature. After 5 washings,

diluted serum samples (1:100,000) and serially-diluted standard mouse IgG (sc-2025, Santa Cruz) were added to the wells (100 μl/well) and incubated at 37 °C for 1 h. The plates were then washed 10 times, and diluted (1:4,000) horse radish peroxidase-conjugated horse anti-mouse IgG (#7076S, Cell Signaling Technology) was added (100 μl/well) to the wells. The plates were incubated for 1 h at 37 °C. After another 10 washings, 1-Step™ Ultra TMB-ELISA Substrate Solution (#34028, Thermo Scientific) was added to the wells (100 μl/well). The plates were incubated in the dark at room temperature for 20 – 30 min, and the reaction was stopped by 2M sulfuric acid (100 μl/well). Optical density at 450 nm of reactants was measured. Samples were assayed in duplicate. Mouse total IgG concentrations were calculated according to a standard curve established by serial dilutions of standard mouse IgG. Assay sensitivity was in the 0.39 and 6.25 ng/ml range.

2.5.6. Enzyme-linked immunosorbent assay (ELISA)

Glucocorticoid levels in WT and KO mouse sera were quantified by ELISA, detecting mouse glucocorticoids according to the manufacturer's instructions (MBS028416, MyBioSource, San Diego, CA, USA).

2.5.7. Flow cytometry

Single cell suspensions from the thymus, spleen and LN were prepared and stained immediately or after culture with Abs against CD3 (1:200, Clone 145-2C11, BD Bioscience), CD4 (1:400, Clone RM4-5, BD Bioscience), CD8$^+$ (1:400, Clone 53-6.7, BioLegend), CD25 (1:200, Clone PC61.5, eBioscience), CD44 (1:200, Clone IM7, BioLegend), CD45.1 (1:200, Clone A20, BD Bioscience), CD45.2 (1:200, Clone 104, BD Bioscience), CD62L (1:200, Clone MEL-14, BD Bioscience), CD107a (1:200, Clone 1D4B, BD Bioscience), CD127 (1:200, A019D5, BioLegend), KLRG1 (1:200, Clone 2F1/KLRG1, BioLegend), Thy1.2 (1:1000, Clone 30-H12, BioLegend), B220 (1:200, Clone RA3-6B2, BD Bioscience), 7AAD (1:25, 51-68981E, BD Bioscience), Annexin-V (1:50, 550474, BD Bioscience). In some experiments, intracellular proteins, such as IFN-γ (1:200, Clone XMG1.2, BD Bioscience), IL-17 (1:200, Clone

TC11-18H10, BD Bioscience), FoxP3 (1:200, Clone 150D, BioLegend), T-bet (1:200, Clone 4B10, BioLegend), RORγt (1:200, Clone B2D, eBioscience) and TNF-α (1:200, Clone MP6-XT22 BD Bioscience) and Granzyme B (1:200, Clone GB11, BioLegend), were detected after the cells were pre-stained with Abs against cell surface Ag, permeabilized with BD Cytofix/Cytoperm™ solution (BD Biosciences), and then stained with Abs against intracellular Ag[323,324].

Flow cytometry was also employed to assess LCMV-specific T cells. The synthetic peptides gp_{33-41}: KAVYNFATC (LCMV-GP, H-2D^b), $np_{396-405}$: FQPQNGQFI (LCMV-NP, H-2D^b) and $gp_{276-286}$: SGVENPGGYCL (LCMV-GP, H-2D^b) were purchased from Sigma-Genosys (Oakville, ON, Canada). PE-gp_{33-41}, PE-$np_{396-405}$, and PE-$gp_{276-286}$ H-2D^b tetrameric complexes were synthesized in-house and applied at 1:100 dilution[324]. These MHC-tetramers served to detect LCMV-specific CD8$^+$ T cells on day 8 post-LCMV infection. Briefly, splenocytes were first stained with PE-gp_{33-41}, PE-$np_{396-405}$ or PE-$gp_{276-286}$ tetramers for 30 min at 37 °C, directly followed by surface staining (CD3, CD8, CD44, and CD62L) and dead cell exclusion (7AAD) for another 20 min at 4 °C. The cells were then fixed with 1% paraformaldehyde, and samples were analyzed by flow cytometry.

For intracellular cytokine staining, 10^6 splenocytes from LCMV-infected mice were maintained for 5 h at 37 °C in RPMI-1640 with 10% FCS and 55 μg/mL β-ME, supplemented with a final concentration of 50 U/ml IL-2, 5 μg/ml Brefeldin A, 2 μM Monensin, 2.5 μg/mL FITC-labeled anti-mouse CD107a and 5 μM gp_{33-41} or gp_{61-80} GLNGPDIYKGVYQFKSVEFD (LCMV-GP, I-Ab) synthetic peptide from Sigma-Genosys. After *ex vivo* incubation, surface staining and cell viability were verified with anti-mouse CD8a, CD4 and CD62L mAbs and 7AAD. The cells were then fixed, permeabilized and stained with anti-mouse TNF-α, IFN-γ and Granzyme B mAbs. Cytokine-producing T cells were analyzed by flow cytometry[325].

2.5.8. Lymphocyte proliferation and apoptosis *in vitro*

Spleen cells were loaded with carboxyfluorescein succinimidyl ester (CFSE: 5 μM for 5 min). After washing, they were stimulated with soluble hamster anti-mouse CD3ε mAb

(clone 145-2C11, 2 μg/ml; BD Biosciences) for T cell proliferation assays. This protocol allows the best long-term T cell proliferation over a 4-day period, for clear demonstration of multiple cell proliferation rounds according to CFSE staining. In B-cell proliferation assays, CFSE-loaded spleen cells were stimulated with goat anti-mouse IgM (5 μg/ml; Jackson ImmunoResearch), IL-4 (10 ng/ml) and goat-anti-mouse CD40 (2 μg/ml, Jackson ImmunoResearch). After 3-4 days, the cells were gated on CD4-, CD8- or B220-positive cells, and their CFSE intensity was ascertained by flow cytometry.

To assess Th1 and Th17 cell proliferation, naïve CD4$^+$ cells were loaded with CFSE, and cultured under Th1 and Th17 conditions for 4 days (detailed below). CD4$^+$ or intracellular IFN-γ^+ or IL-17$^+$ cells were then gated, and their CFSE intensity was assessed by flow cytometry.

For cell cycle analysis, total spleen cells were stimulated with anti-CD3ε mAb, as described above, and stained with anti-Thy1.2 mAb and propidium iodide (PI) (20 μg/ml) on days 0, 1 and 2. Thy1.2$^+$ T cells were gated, and their PI signal strength was measured by flow cytometry.

For T cell apoptosis analysis, spleen cells were stimulated with anti-CD3ε mAb (2 μg/ml) plus crosslinked human FasL-FLAG (0.133 μg/ml; FasL-FLAG was pre-incubated for 24 hours at 4 °C at a 1:1 ratio with 0.133 μg/ml mouse monoclonal Ab against FLAG; the final concentration of crosslinked FasL-FLAG for culture was 0.6 μg/ml[326]) and cultured for 4 h. Cells positive for CD4 or CD8 were gated and analyzed for annexin V expression.

2.5.9. Th1 and Th17 cell differentiation *in vitro*

T cell differentiation *in vitro* was undertaken as follows[323,324]. Naïve CD4$^+$ T cells (CD4$^+$ CD62L$^+$CD44^{low}) were isolated from KO or WT mouse Spleen with EasySep^{TM} mouse naïve CD4$^+$ T cell isolation kits (19765, Stem Cell Technology). Naïve CD4$^+$ T cells from WT and KO mice (0.1 × 10^6 cells/well) were mixed with feeder cells (0.5 × 10^6 cells/well) and cultured in 96-well plates in the presence of soluble anti-CD3ε Ab (2 μg/ml). Feeder cells plus anti CD3ε Ab were used, as in our hands, they achieved the most consistent Th1 and

Th17 differentiation conditions. Cultures were supplemented with recombinant mouse IL-12 (10 ng/ml; 419-ML, R&D Systems, Minneapolis, MN, USA) and anti-IL-4 mAb (5 μg/ml; MAB404, R&D Systems) for the Th1 condition, with recombinant mouse IL-6 (20 ng/ml; 406-ML, R&D Systems), recombinant human TGF-β1 (5 ng/ml; 240-B, R&D Systems) and anti-IL-4 (5 μg/ml) and anti-IFN-γ mAbs (5 μg/ml; MAB485, R&D Systems) for the Th17 condition. The cells were stimulated with PMA (10 μM) and ionomycin (100 μg/ml) in the presence of 5 μg/ml Brefeldin A for the last 4 h of culture, and their intracellular IFN-γ, T-bet, IL-17, and RORγt were analyzed by flow cytometry.

2.5.10. Chimera generation

Eight- to 10-week-old C57BL/6J (CD45.2$^+$) $\times$ C57B6.SJL (CD45.1$^+$) F1 mice were irradiated at 1,100 rads. Twenty-four h later, they received *i.v.* 2×10^6 fetal liver cells from WT or KO mice in the C57BL/6J $\times$ 129/sv (CD45.2$^+$) F1 background. Peripheral white blood cells of recipients were analyzed by flow cytometry 8 weeks after fetal liver cell transplantation. Twelve weeks after transplantation, chimeras with successful implantation of donor-derived white blood cells were studied in *in vitro* T cell function experiments and for EAE induction.

2.5.11. EAE induction and assessment

EAE was induced in 8- to 12-week-old female WT and KO mice[327]. Briefly, mice were immunized with 300 μg MOG$_{35\text{-}55}$ peptide (Biomatik, Wilmington, DE, USA) emulsified in complete Freund's adjuvant, followed by *i.p.* injection of 400 ng pertussis toxin (List Biological Laboratories, Campbell, CA, USA) on days 0 and 2. EAE development was scored daily between days 0 and 35 according to a scale ranging from 0 to 5, as follows: 0, no sign of paralysis, 1, weak tail; 2, paralyzed tail; 3, paralyzed tail and weakness of hind limbs; 4, completely paralyzed hind limbs; 5, moribund. Scores were assigned in 0.5 unit increments when symptoms fell between 2 full scores.

Female chimeras with successful donor cell implantation (verified according to CD45.2 single- positive cells in peripheral blood) 12 weeks after fetal liver transplantation were also used for EAE induction. The same protocol described above was followed, except that 200 μg MOG$_{35\text{-}55}$ peptide for immunization and 200 ng pertussis toxin/injection were administered to each chimeric mouse.

2.5.12. EAE histology

To assess the degree of inflammation and CNS demyelination, EAE mice were euthanized on day 30 and perfused by intra-cardiac injection of PBS. Spinal cord sections were stained with H/E or NovaUltra$^{\text{TM}}$ Luxol Fast Blue Staining Kit (IHC World, Woodstock, MD, USA), according to the manufacturer's instructions. Each SC section was subdivided into 4 regions: 1 anterior, 1 posterior, and 2 lateral. Each region was scored on a scale of 0 to 3 for lymphocyte infiltration and demyelination in a 1-way blinded fashion. Thus, each animal had a potentially maximal score of 12 points for lymphocyte infiltration and demyelination, respectively[328,329]. Total pathological scores were the sum of these 2 parameters.

2.5.13. Isolation of mononuclear cells from the spinal cord and brain

Peripheral blood was removed from the spinal cord and brain by intra-cardiac perfusion through the left ventricle with heparinized ice-cold PBS. The spinal cord and brain were harvested, ground and then passed through a 70-μm mesh screen. The cells were centrifuged through a 40%-60%-90% discontinuous Percoll gradient. Mononuclear cells at the 40% to 60% Percoll interface were collected and stained for phenotype analysis.

2.5.14. Differentiation/characterization of mouse Th1 and Th17 cells

Cells from draining LN and mononuclear cells from the spinal cord and brain were further stimulated with PMA (5 nM) and ionomycin (500 ng/ml) for 4 h in the presence of

Golgi Stop (554724, BD Biosciences), before being harvested. They were stained with Abs against cell surface Ag, fixed with Cytofix/Cytoperm solution (555028, BD Biosciences), and then stained with mAbs against intracellular IFN-γ (1:200, Clone XMG1.2, BD Bioscience) and IL-17 (1:200, Clone TC11-18H10, BD Bioscience). Stained cells were analyzed by flow cytometry.

2.5.15. LCMV infection

LCMV-WE was obtained from Dr. R.M. Zinkernagel (University of Zurich, Zurich, Switzerland). Viral stock was propagated *in vitro*, and viral titers were quantified by focus-forming assay[325]. Mice were infected by the *i.v.* route with 200 focus-forming units (ffu) of LCMV-WE. They were sacrificed 8 days post-infection, and their spleens were harvested for primary immune response analysis.

CD8$^+$ T cells were isolated from the spleen of naïve or infected WT mice, with EasySep mouse CD8$^+$ T cell isolation kits (19853, Stem Cell Technology). RNA from isolated cells was extracted with TRIzol (Invitrogen), followed by RT-qPCR.

2.5.16. Y2H assay

Y2H screening was performed by Hybrigenics Services (Paris, France). The coding sequence for human *ARMC5* cDNA (aa30-935) (GenBank accession number GI: 157426855) was PCR-amplified and cloned into pB29 as an N-terminal fusion protein to LexA (N-ARMC5-LexA-C). The construct was verified by sequencing the entire insert and served as bait to screen a random-primed human thymocyte cDNA library constructed in the pP6 plasmid. pB29 and pP6 vectors were derived from the original pBTM116[330,331] and pGADGH[332] plasmids, respectively.

Eighty million yeast clones (8-fold the complexity of the library) were screened via a mating approach with YHGX13 (Y187 ade2-10: loxP-kanMX-loxP, matα) and L40Gal4 (mata) yeast strains[333]. One hundred and sixty-five His$^+$ colonies were selected on medium lacking tryptophan, leucine, and histidine, and supplemented with 5.0 mM of 3-aminotriazole

to quench bait auto-activation. Prey fragments of positive clones were amplified by PCR and sequenced at their 5' and 3' junctions. The resulting sequences were considered to identify corresponding interacting proteins in the GenBank database via a fully-automated procedure. A Predicted Biological Confidence score was attributed to each interaction[321].

2.5.17. Statistics and general methods

For *in vivo* animal studies, the sample size was determined by estimation, based on our experience and literature. No formal randomization was used, but littermates or age and sex matched WT and KO mice were used. In general, two-tailed Student's t tests were used. *Chi*-square test was used to compare the difference between two proportions. One-way ANOVA followed with Bonferroni's multiple comparisons test was used in data from more than three groups. For the histology experiments, one-way blind examination was performed.

2.6. Data availability

The data that support the findings of this study are available in the text and the supplementary data files, or from the corresponding author upon request.

2.7. Acknowledgments

2.8. Supplementary Figures

2.8.1. Supplementary Figure 1
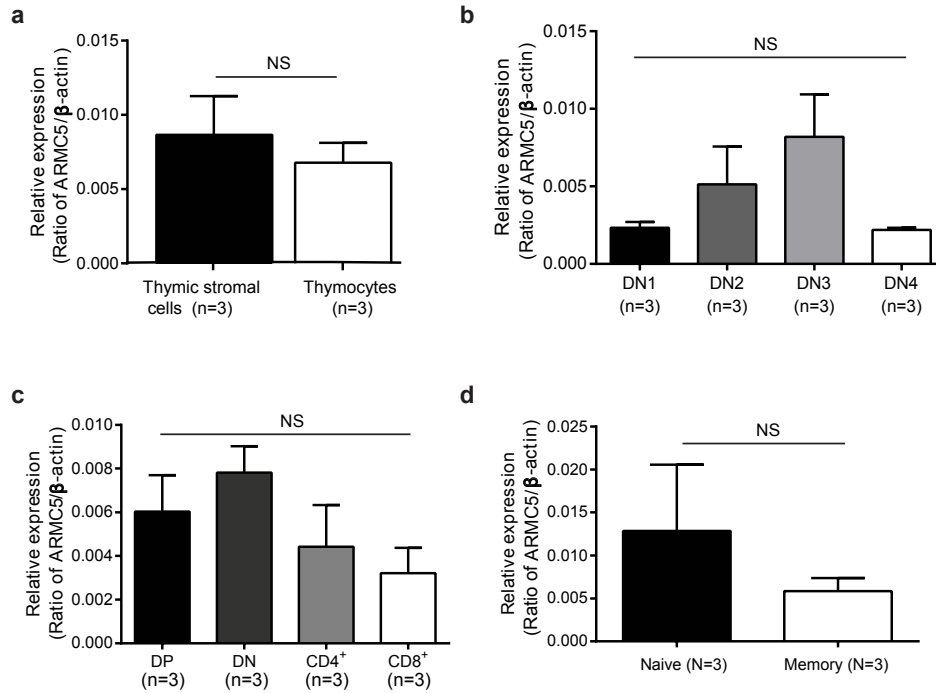
Fig. 2.S.1. *Armc5* **mRNA expression in different thymocyte and T-cell subpopulations.** **(a)**. *Armc5* mRNA expression in thymocytes and thymic stroma cells Thymocytes were flushed out from the thymus of WT mice, and the remainder was considered to be thymic stroma cells. **(b-c)**. *Armc5* expression in thymocyte subpopulations (B: DN1-4; C: CD4 SP, CD8 SP; CD4CD8 DP, an DN) were sorted by flow cytometry. **(d)**. *Armc5* mRNA expression in naïve versus memory T cells CD62L$^+$CD44^{lo} naïve T cells and CD62L$^+$CD44^{int-hi} memory T cells were sorted by flow cytometry from WT spleen cells. RNA was extracted from different subpopulations of thymocytes and T cells, or from T cells cultured under different conditions. *Armc5* mRNA expression levels in these cells were measured by RT-qPCR, with β-actin mRNA levels as internal controls. The numbers (n) of experiments performed are indicated. Pooled results of multiple experiments are expressed as means $\pm$ SEM of ratios of *Armc5* versus β-actin signals, unless specified otherwise. Two-tailed Student's *t*-test was used in data between two groups. One-way ANOVA followed with Bonferroni's multiple comparisons test was used in data among four groups. NS, no significance.

2.8.2. Supplementary Figure 2
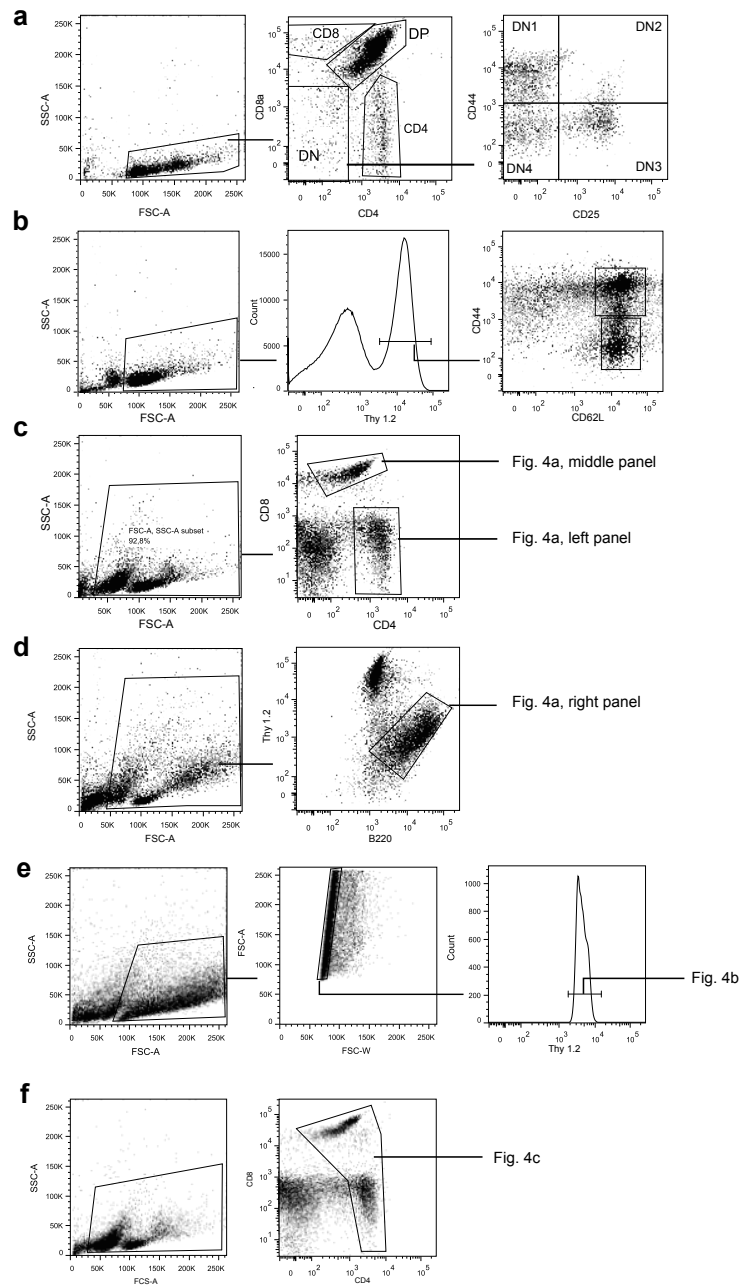
Fig. 2.S.2. Gating strategies (a). Sequential gating strategy for sorting SP, DP and DN (DN1-4) populations in thymocytes. (b). Sequential gating strategy for sorting memory cells (CD62L$^+$CD44^{int-hi}) and naïve cells (CD62L$^+$CD44^{lo}) in Thy1.2$^+$ lymph nodes cells. (c). Sequential gating strategy for CD4$^+$ or CD8$^+$ cells in spleen cells. (d). Sequential gating strategy for B220$^+$ cells in spleen cells. (e). Sequential gating strategy for the cell cycle analysis of T cells. (f). Sequential gating strategy for apoptosis analysis of CD4$^+$ and CD8$^+$ cells.
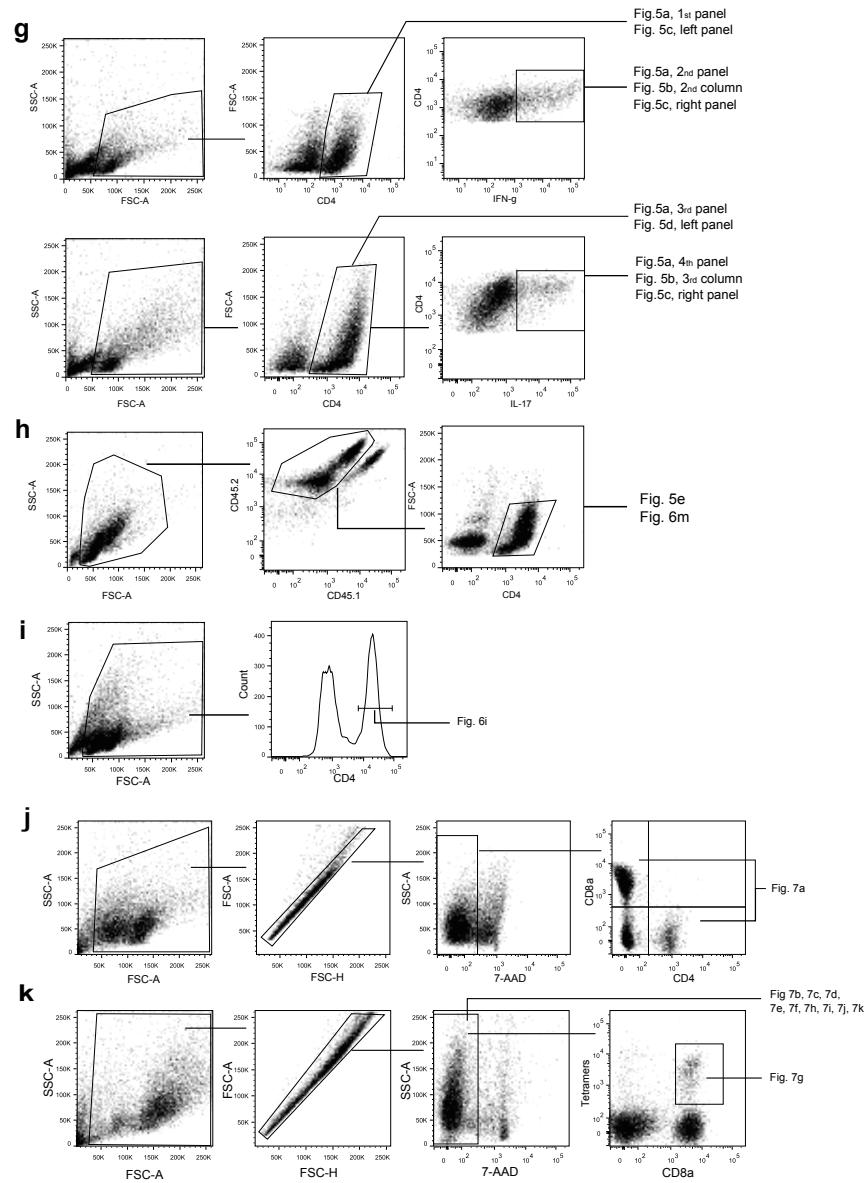
Fig. 2.S.2. (Continued). Gating strategies (g). Sequential gating strategy for CD4$^+$/IL17$^+$ or CD4$^+$/IFN-γ^+ cells. (h.) Sequential gating strategy for CD4$^+$ cells derived from WT and KO donor cells in chimeric mice. (i). Sequential gating strategy for CD4$^+$ cells. (j). Sequential gating strategy for CD4$^+$ or CD8$^+$ cells in LCMV infection mice model. (k). Sequential gating strategy for CD8$^+$ and tetramers$^+$ cells in LCMV infection mice model.

2.8.3. Supplementary Figure 3
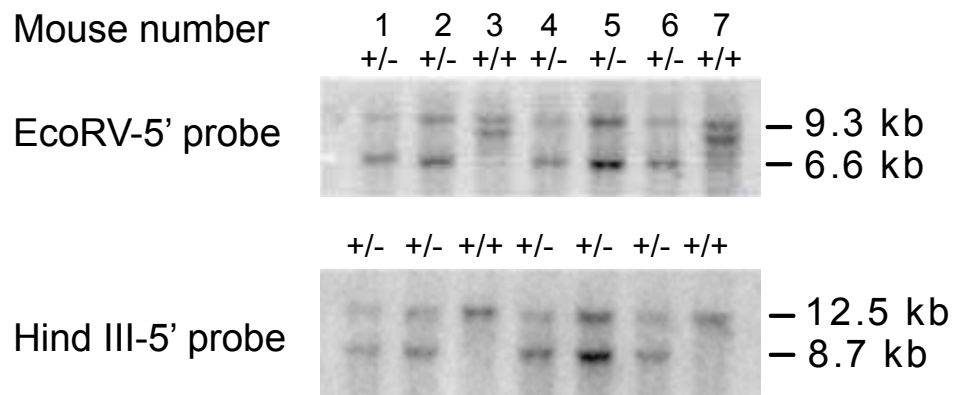
Fig. 2.S.3. Genotyping of *Armc5* mutant mice Tail DNA was digested with EcoRV and analyzed by Southern blotting (top panel) with the 5' probe whose location is indicated Figure 2.2a. A 9.3-kb band representing the WT allele and a 6.6-kb band representing the recombinant allele are shown. Similarly, tail DNA was digested with HindIII and analyzed with the 3' probe (bottom panel). A 12.5-kb band representing the WT allele and an 8.7-kb band representing the recombinant allele are indicated.

2.8.4. Supplementary Figure 4
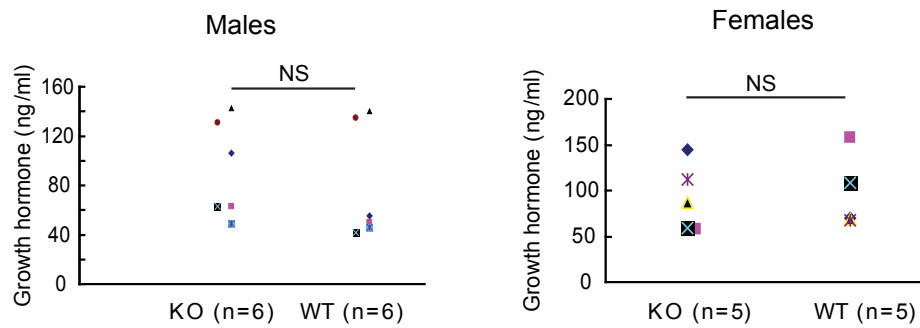
Fig. 2.S.4. Serum growth hormone levels in KO mice Serum growth hormone levels in 8 to 12-week-old KO and WT mice were measured by ELISA. The results are reported as scatter plots, with each symbol representing actual values. Mouse numbers (n) in each group are indicated. $p > 0.05$ (two-tailed Student's t test). NS: not significant.

2.8.5. Supplementary Figure 5

WT   KO
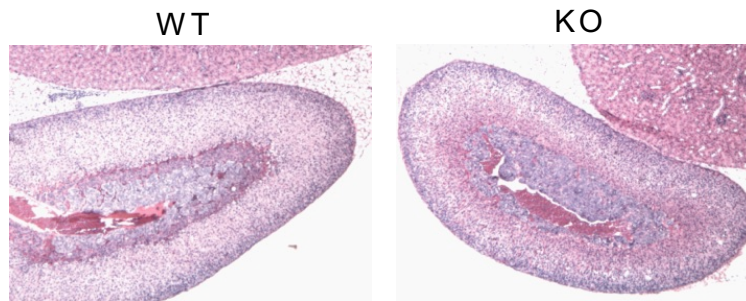
Fig. 2.S.5. Adrenal gland histology of young KO mice Adrenal glands from WT and KO mice (8 – 12 weeks old) were sectioned and stained with H/E. Representative micrographs from a KO (12-week-old male) mouse and its WT male littermate. No histological abnormalities were found in the KO adrenal gland.

2.8.6. Supplementary Figure 6
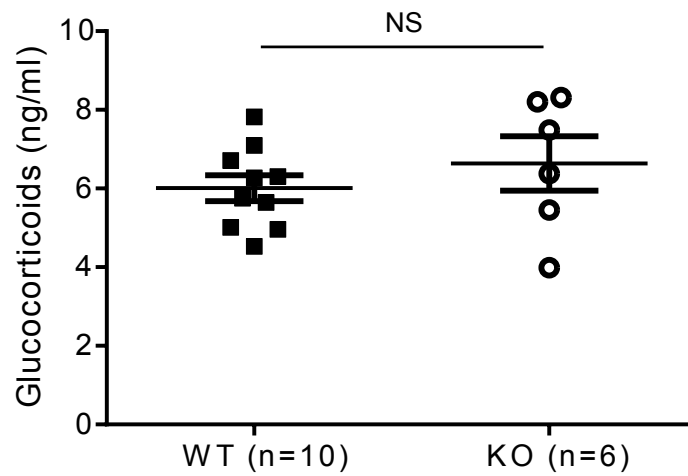
Fig. 2.S.6. Serum glucocorticoid levels in young WT and KO mice The mice were bled between 12:30 – 1:30 pm. Serum levels (means $\pm$ SEM) of glucocorticoids in young KO and WT mice are shown. Two-tailed Student's t test was used for statistical analysis. NS: not significant.

2.8.7. Supplementary Figure 7
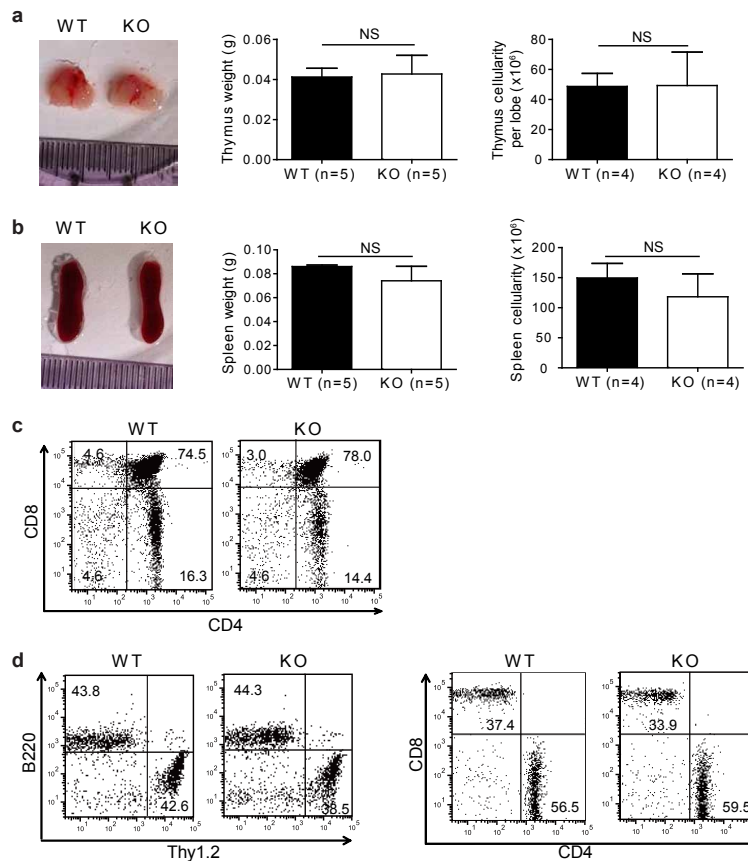
Fig. 2.S.7. Thymus and spleen weight, cellularity and cell subpopulations in KO mice (a). thymus size, weight, and cellularity in WT and KO mice. Left panel: Representative photo of the KO and WT thymus from 8-week-old littermates. Right panels: thymus weight and cellularity of KO and WT from 8 to 12-week-old male littermates. Mouse numbers (n) in each group (n) are indicated. $p > 0.05$ (2-tailed Student's t test). NS: not significant. **(b)**. spleen size, weight, and cellularity in WT and KO mice Left panel: Representative photo of KO and WT spleen from 8-week-old littermates. Right panels: weight and cellularity of KO and WT spleen from 8-12-week-old male littermates. Mouse numbers (n) in each group (n) are indicated. $p > 0.05$ (2-tailed Student's t test). **(c)**. T-cell subpopulations in KO thymus in WT and KO mice Thymocytes from adult KO and WT mice (8 – 12 weeks old) were analyzed by flow cytometry for percentages of $CD4^+$, $CD8^+$, and $CD4^+CD8^+$ subpopulations. Experiments were conducted more than 3 times. Representative dot plots are reported. **(d)**. Cell subpopulations in WT and KO Spleen cells from adult KO and WT mice (8 – 12 weeks old) were analyzed by flow cytometry for percentages of $Thy1.2^+$ T cells versus $B220^+$ B cells (left panel), and $CD4^+$ versus $CD8^+$ cells (right panel). Experiments were conducted more than 3 times, and representative dot plots are shown.

2.8.8. Supplementary Figure 8
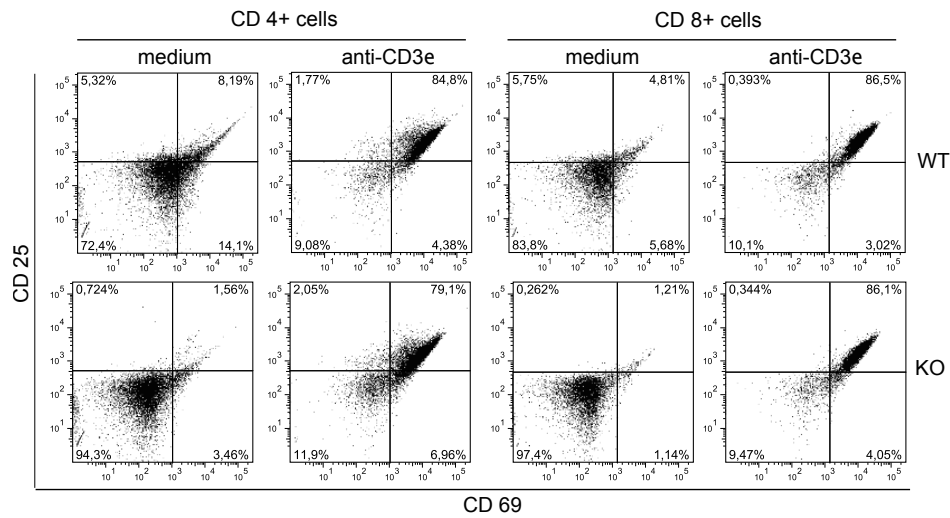
Fig. 2.S.8. CD25 and CD69 expression in CD4$^+$ and CD8$^+$ cells after anti-CD3ε stimulation Spleen cells were stimulated with anti-CD3ε mAb(2 μg/ml) for 16 hours. The cells were gated on CD4-positive and CD8-positive. Experiments were conducted independently 3 times. Representative dot plots are shown.

2.8.9. Supplementary Figure 9
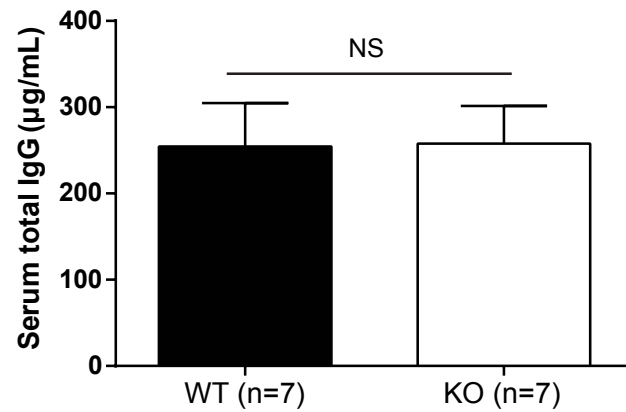
Fig. 2.S.9. Serum IgG levels in WT and KO mice Total IgG levels in WT and KO mouse serum were measured by ELISA, and means $\pm$ SEM are presented. Mouse numbers (n) in each group are indicated. No statistically significant difference between WT and KO IgG levels was observed (2-tailed Student's t test). NS: not significant.

2.8.10. Supplementary Figure 10
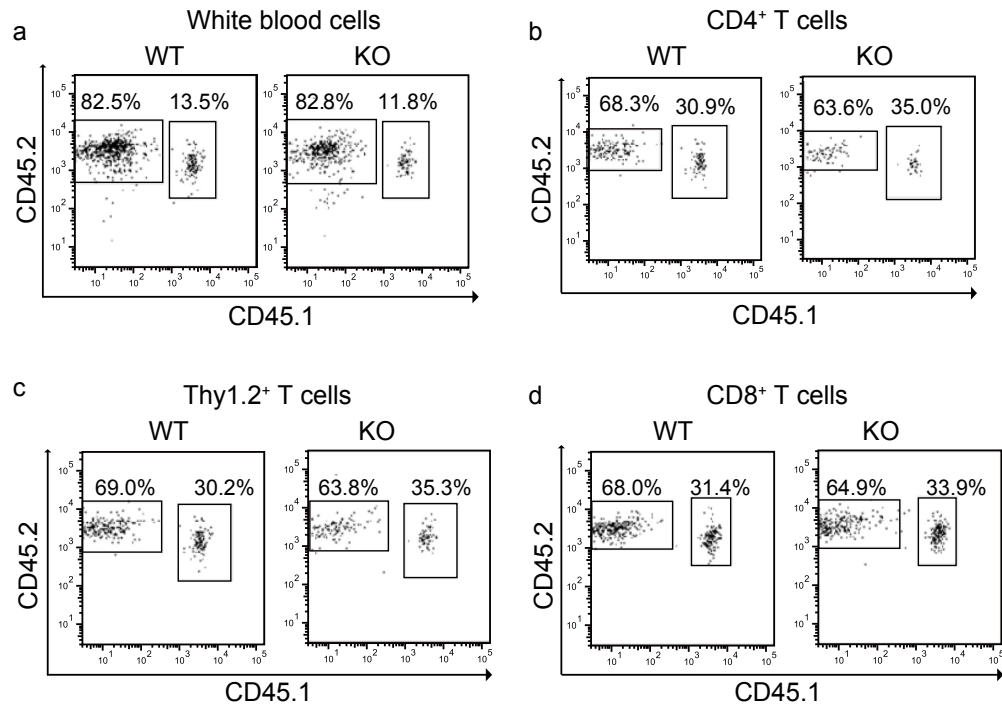
Fig. 2.S.10. Implantation of donor cells in blood and spleen of chimeric mice
Eight weeks after KO and WT fetal liver transplantation, peripheral blood cells from recipient mice were examined by flow cytometry. Percentages of donor-derived (CD45.2 single-positive cells) versus recipient-derived (CD45.1/CD45.2 double-positive cells; panel a), total T cells (Thy1.2$^+$ cells; panel c), CD4$^+$ cells (panel b) and CD8$^+$ cells (panel d) in the spleen were measured by flow cytometry. Experiments were conducted more than 4 times, and representative dot plots are shown.

2.8.11. Supplementary Figure 11
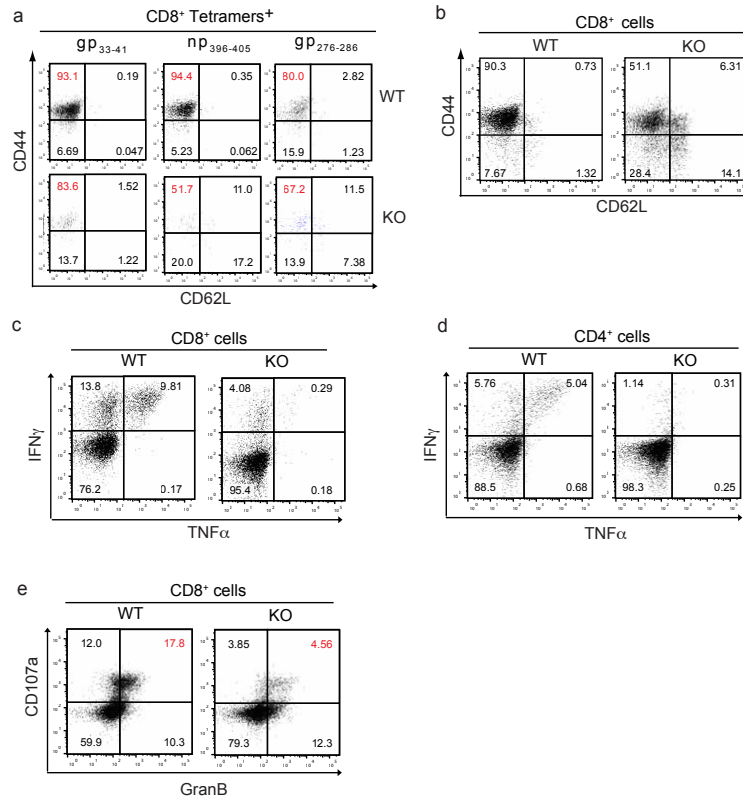
Fig. 2.S.11. Multiple parameters in spleen T cells on day 8 post-LCMV infection WT and KO spleen T cells were analyzed for different parameters by flow cytometry on day 8 post-LCMV infection. Representative flow cytometric dot plots are shown (pooled results of the experiments are summarized in bar graphs and presented in Figure 7 in the text proper). $(\mathbf{a} - \mathbf{b})$. Activation of LCMV-specific spleen CD8$^+$ T cells On the left panel, gp$_{33\text{-}41}$, np$_{396\text{-}405}$ and gp$_{276\text{-}286}$ tetramer-positive CD8$^+$ cells in the KO and WT mouse spleens were assessed for activation markers (CD62L^{lo}CD44^{hi}), and the percentages of this effector memory cell subpopulation are indicated. On the right panel, the dot plot shows the percentages of this population gated on total CD8$^+$ cells. $(\mathbf{c} - \mathbf{d})$. Expression of IFN-$\gamma$ and/or TNF-α in gp$_{33\text{-}41}$-stimulated CD8$^+$ and gp$_{61\text{-}80}$-stimulated CD4$^+$ cells Dot plots show intracellular IFN-γ and TNF-α expression in WT and KO CD8$^+$ and CD4$^+$ cells stimulated by gp$_{34\text{-}41}$ (for CD8$^+$ cells) and gp$_{61\text{-}80}$ (for CD4$^+$ cells) (both at 5 μM), respectively, for 5 h, in the presence of 50 U/ml IL-2, 5 μg/ml Brefeldin A and 2 μM Monensin. $(\mathbf{e})$. gp$_{33\text{-}41}$-specific CD107a$^+$GranB$^+$ CD8$^+$ T cells on day 8 post-LCMV infection Spleen cells from KO and WT mice on day 8 post-LCMV infection were stimulated *ex vivo* with gp$_{33\text{-}41}$ peptide (5 μM) for 5 h in the presence of 50 U/ml IL-2, 5 μg/ml Brefeldin A, 2 μM Monensin and 2.5 μg/ml FITC-labeled anti mouse CD107a. Percentages of CD107a$^+$ and GranB$^+$ cells among CD8$^+$ cells were quantified by flow cytometry.

2.8.12. Supplementary Figure 12

Fig. 2.S.12. Binding regions between ARMC5 and its associating molecules
The cDNA coding sequences representing the protein binding regions between ARMC5 and its preys are illustrated.

DomSight: HTH_RP1_hgx4009v1 vs. Human Thymocytes (CD4+, CD8+) RP1 (19 Jun 2015)
(Bait plasmid(s): hgx4009v1_pB29)

MTF2 — Homo sapiens, GID: 256419011, Gene ID: 22823
491 SID
429 449 Coiled domain
2 53 Zinc finger, PHD-type - SM00249
2 53 Zinc finger, PHD-finger - PF00628
101 151 Zinc finger, PHD-type - SM00249
442 490 Polycomb-like MTF2 factor 2, C-terminal - PF14061

NFE2L1 — Homo sapiens, GID: 189181670, Gene ID: 4779
185 341 SID
672 713 Coiled domain
652 716 Basic-leucine zipper domain - SM00338
656 714 Basic-leucine zipper domain - PF00170

PCBP1 — Homo sapiens, GID: 222352150, Gene ID: 5093
60 283 SID
12 80 K Homology domain - SM00322
15 75 K Homology domain, type 1 - PF00013
96 167 K Homology domain - SM00322
99 162 K Homology domain, type 1 - PF00013
278 348 K Homology domain - SM00322
282 343 K Homology domain, type 1 - PF00013

POLR2A — Homo sapiens, GID: 306482654, Gene ID: 5430
1429 1499 SID
699 719 Coiled domain
949 969 Coiled domain
1262 1282 Coiled domain
16 354 RNA polymerase Rpb1, domain 1 - PF04997
246 549 RNA polymerase, N-terminal - SM00663
356 521 RNA polymerase, alpha subunit - PF00623
525 691 RNA polymerase Rpb1, domain 3 - PF04983
718 823 RNA polymerase Rpb1, domain 4 - PF05000
830 1424 RNA polymerase Rpb1, domain 5 - PF04998
896 1079 RNA polymerase Rpb1, domain 6 - PF04992
1164 1299 RNA polymerase Rpb1, domain 7 - PF04990
1500 1580 SID
1594 1606 RNA polymerase II, heptapeptide repeat, - PF05001
1610 1622 RNA polymerase II, heptapeptide repeat, - PF05001
1616 1629 RNA polymerase II, heptapeptide repeat, - PF05001
1623 1636 RNA polymerase II, heptapeptide repeat, - PF05001
1630 1643 RNA polymerase II, heptapeptide repeat, - PF05001
1644 1657 RNA polymerase II, heptapeptide repeat, - PF05001
1658 1671 RNA polymerase II, heptapeptide repeat, - PF05001
1672 1685 RNA polymerase II, heptapeptide repeat, - PF05001
1686 1699 RNA polymerase II, heptapeptide repeat, - PF05001
1700 1713 RNA polymerase II, heptapeptide repeat, - PF05001
1714 1727 RNA polymerase II, heptapeptide repeat, - PF05001
1728 1741 RNA polymerase II, heptapeptide repeat, - PF05001
1742 1755 RNA polymerase II, heptapeptide repeat, - PF05001
1757 1769 RNA polymerase II, heptapeptide repeat, - PF05001
1770 1783 RNA polymerase II, heptapeptide repeat, - PF05001
1784 1797 RNA polymerase II, heptapeptide repeat, - PF05001
1798 1811 RNA polymerase II, heptapeptide repeat, - PF05001
1826 1839 RNA polymerase II, heptapeptide repeat, - PF05001
1841 1853 RNA polymerase II, heptapeptide repeat, - PF05001
1847 1860 RNA polymerase II, heptapeptide repeat, - PF05001
1855 1867 RNA polymerase II, heptapeptide repeat, - PF05001
1861 1874 RNA polymerase II, heptapeptide repeat, - PF05001
1868 1881 RNA polymerase II, heptapeptide repeat, - PF05001
1875 1888 RNA polymerase II, heptapeptide repeat, - PF05001
1882 1895 RNA polymerase II, heptapeptide repeat, - PF05001
1889 1902 RNA polymerase II, heptapeptide repeat, - PF05001
1896 1909 RNA polymerase II, heptapeptide repeat, - PF05001
1904 1916 RNA polymerase II, heptapeptide repeat, - PF05001
1910 1923 RNA polymerase II, heptapeptide repeat, - PF05001
1924 1936 RNA polymerase II, heptapeptide repeat, - PF05001
1931 1947 RNA polymerase II, heptapeptide repeat, - PF05001
1941 1954 RNA polymerase II, heptapeptide repeat, - PF05001
1948 1960 RNA polymerase II, heptapeptide repeat, - PF05001

RPN2 — Homo sapiens, GID: 209413736, Gene ID: 6185
88 196 SID
7 29 Transmembrane domain
539 561 Transmembrane domain
573 595 Transmembrane domain
605 622 Transmembrane domain
1 22 Signal peptide domain
3 628 Ribophorin II - PF05817

SERTAD3 — Homo sapiens, GID: 5718684, Gene ID: 29946
22 196 SID
33 70 SERTA - PF06031

STK24 — Homo sapiens, GID: 110349738, Gene ID: 8428
4 140 SID
50 77 Coiled domain
24 274 Protein kinase, catalytic domain - PF00069
24 274 Serine/threonine / dual specificity pro - SM00220

TAF1 — Homo sapiens, GID: 20357585, Gene ID: 6872
1144 1313 SID
1076 1096 Coiled domain
1140 1160 Coiled domain
1234 1265 Coiled domain
1495 1499 Coiled domain
26 86 TAFII-230 TBP-binding - PF09247
586 1048 Transcription initiation factor TFIID su - PF12157
1399 1499 Bromodomain - SM00297
1411 1490 Bromodomain - PF00439
1500 1522 Coiled domain
1507 Bromodomain - SM00297
1525 1630 Bromodomain - SM00297
1534 1616 Bromodomain - PF00439

TCF12 — Homo sapiens, GID: 46370078, Gene ID: 6938
277 490 SID
578 631 Myc-type, basic helix-loop-helix (bHLH) - PF00010
583 636 Myc-type, basic helix-loop-helix (bHLH) - SM00353

TOX4 — Homo sapiens, GID: 39077116, Gene ID: 9878
157 442 SID
273 293 Coiled domain
222 292 High mobility group (HMG) box domain - SM00398
223 291 High mobility group (HMG) box domain - PF00505

TTF1 — Homo sapiens, GID: 328927003, Gene ID: 7270
430 640 SID
616 663 SANT/Myb - SM00717
621 677 Myb-like DNA-binding domain - PF13921
665 747 SANT/Myb domain - SM00717

ZBTB38 — Homo sapiens, GID: 148276989, Gene ID: 253461
366 622 SID
23 127 BTB/POZ - PF00651
33 131 BTB/POZ-like - SM00225
342 364 Zinc finger, C2H2-like - SM00355
357 382 Zinc-finger double domain - PF13465
371 398 Zinc finger, C2H2-like - SM00355
460 482 Zinc finger, C2H2 - PF00096
460 482 Zinc finger, C2H2-like - SM00355
488 510 Zinc finger, C2H2-like - SM00355
504 525 Zinc-finger double domain - PF13465
516 539 Zinc finger, C2H2-like - SM00355
1010 1032 Zinc finger, C2H2-like - SM00355
1025 1048 Zinc-finger double domain - PF13465
1038 1060 Zinc finger, C2H2-like - SM00355
1052 1076 Zinc-finger double domain - PF13465
1066 1088 Zinc finger, C2H2-like - SM00355
1081 1104 Zinc-finger double domain - PF13465
1094 1116 Zinc finger, C2H2-like - SM00355
1125 1147 Zinc finger, C2H2-like - SM00355

DomSight: HTH_RP1_hgx4009v1 vs. Human Thymocytes (CD4+, CD8+) RP1 (19 Jun 2015)
(Bait plasmid(s): hgx4009v1_pB29)

Legend

Bait fragment
SID fragment
Pfam or SMART domain
Transmembrane domain (TMHMM)
Coiled-coil domain (Ncoils)
Signal peptide (SignalPHMM)

SID: Selected Interaction Domain
It is the amino acid sequence shared by all prey
fragments matching the same reference protein.
SIDs have been found in numerous cases to correspond
to an identified structural or functional domain.
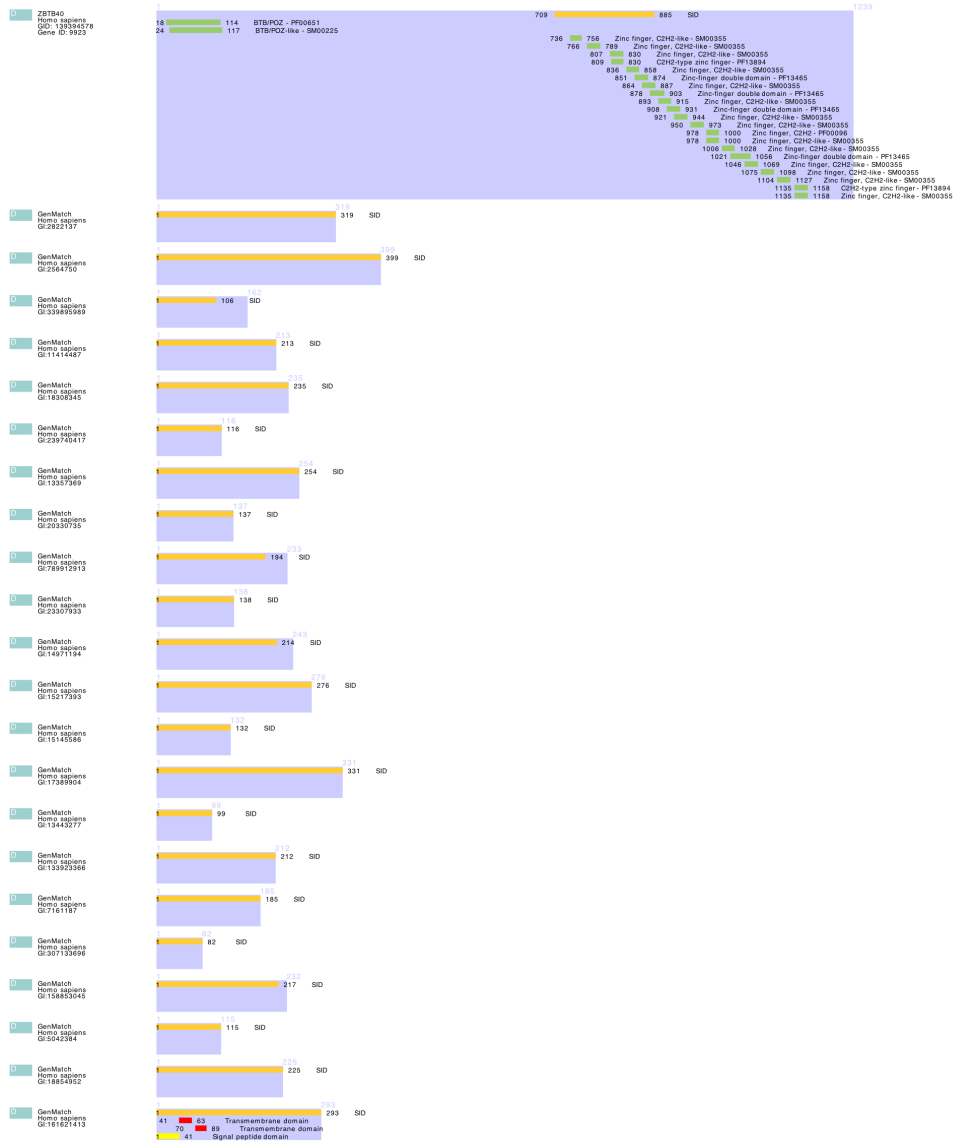
2.8.13. Supplementary Figure 13
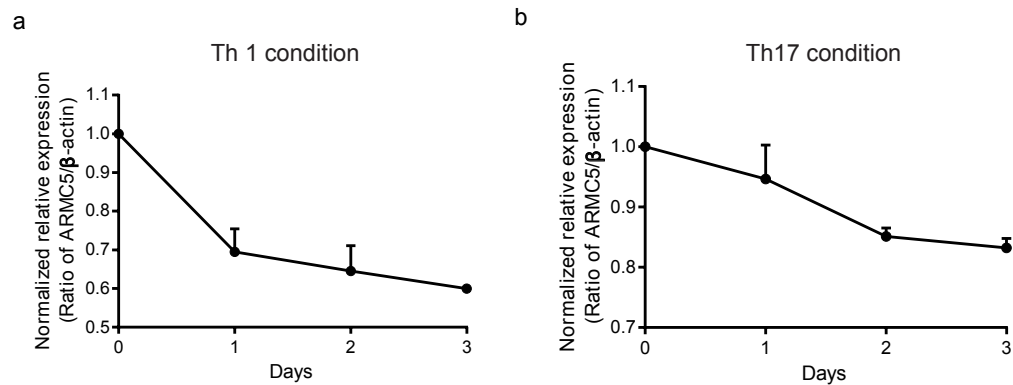
Fig. 2.S.13. ***Armc5* mRNA expression in CD4$^+$ cells cultured under Th1 and Th17 conditions** WT naïve CD4$^+$ cells were cultured under Th1 (panel a) or Th17 (panel b) conditions and harvested at 24, 48 and 72 h. *Armc5* mRNA expression was measured by RT-qPCR. Experiments were conducted more than 3 times and the normalized ratios of *Armc5* versus *β-actin* signals (means ± SEM) of representative experiments are shown. The signal ratios at 0 h are designated as 1.

2.8.14. Supplementary Figure 14
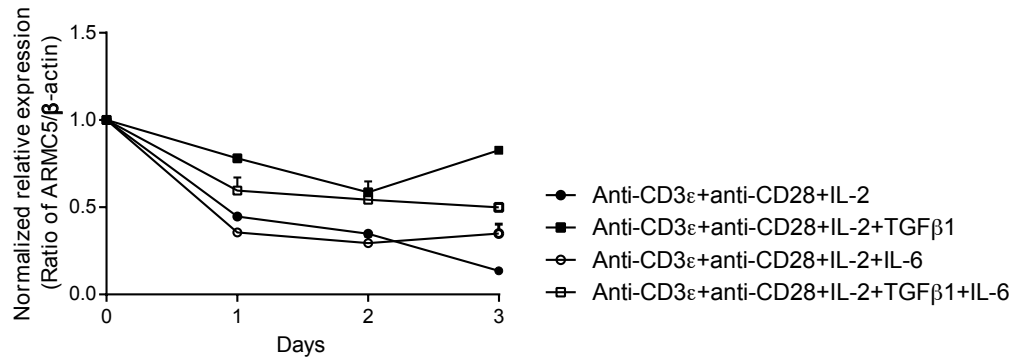
Fig. 2.S.14. *Armc5* **mRNA expression in CD4$^+$ cells cultured in the presence of different lymphokines** WT naïve CD4$^+$ cells were cultured in wells coated with anti-CD3ε and anti-CD28 (0.5μg/ml and 1μg/ml during coating) in the presence of IL-2 (2 μg/ml). In addition, IL-6 (20 ng/ml) or TGF-β1 (5 ng/ml), or both was added to culture. The cells were harvested at 24, 48 and 72 h, and their *Armc5* mRNA expression was measured by RT-qPCR. Experiments were conducted more than 3 times and the normalized ratios of *Armc5* versus β-*actin* signals (means $\pm$ SEM) of representative experiments are shown. The signal ratios at 0 h are designated as 1.

2.8.15. Supplementary Figure 15
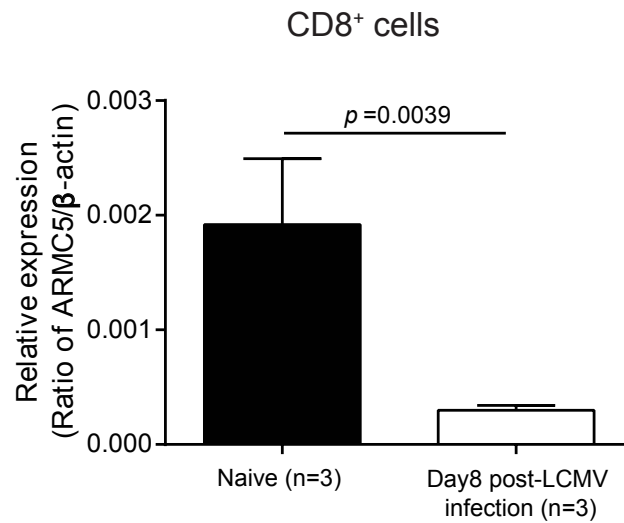
CD8⁺ cells

Fig. 2.S.15. *Armc5* **mRNA expression in CD8⁺ T cells on day 8 post-LCMV infection** CD8⁺ cells were isolated from the spleens of naïve or LCMV-infected (day 8 post-infection) WT mice, with EasySep^TM mouse CD8⁺ T-cell isolation kits. *Armc5* mRNA levels were measured by RT-qPCR. Means ± SEM of ratios of *Armc5* signals versus *β-actin* signals from 3 pairs of mice are shown.

2.9. Supplementary Tables

2.9.1. Supplementary Table 1

Table 2.S.1. Summary of adrenal glands hyperplasia in WT and KO mice

Group	Age (months) (Mean ± SD)	Bilateral Hyperplasia	Unilateral Hyperplasia
WT (n=5)	20 ± 1.7	0/5*	1/5*
KO (n=5)	18 ± 3.2	3/5*	2/5*

* number of positive mouse/number of total mice

2.9.2. Supplementary Table 2

Table 2.S.2. A complete list of binding identified by Y2H assay *(Continued in next pages)*

Results Summary

ULTImate Y2H SCREEN
Homo sapiens - ARMC5
vs Human Thymocytes (CD4+, CD8+) RP1

Fri, Jun 19, 2015 - 09:15 AM

Screen Parameters

Nature	cDNA
Reference Bait Fragment	Homo sapiens - ARMC5 (aa 30-935) ; hgx4009v1
Prey Library	Human Thymocytes (CD4+, CD8+) RP1
Vector(s)	pB29 (N-bait-LexA-C fusion)
Processed Clones	165 (pB29_A)
Analyzed Interactions	80 millions (pB29_A)
3AT Concentration	5.0 mM (pB29_A)

Global PBS®

	Global PBS (for Interactions represented in the Screen)	Nb	%
A	Very high confidence in the interaction	1	2.0%
B	High confidence in the interaction	5	10.2%
C	Good confidence in the interaction	1	2.0%
D	Moderate confidence in the interaction This category is the most difficult to interpret because it mixes two classes of interactions : - False-positive interactions - Interactions hardly detectable by the Y2H technique (low representation of the mRNA in the library, prey folding, prey toxicity in yeast)	38	77.6%
E	Interactions involving highly connected (or relatively highly connected) prey domains, warning of non-specific interaction. The total number of screens performed on each organism is taken into account to set this connectivity threshold: 20 interactions to different bait proteins in our entire database for Human, 10 for Mouse, Drosophila and Arabidopsis and 6 for all other organisms. They can be classified in different categories: - Prey proteins that are known to be highly connected due to their biological function - Proteins with a prey interacting domain that contains a known protein interaction motif or a biochemically promiscuous motif	4	8.2%
F	Experimentally proven technical artifacts	0	0.0%
	Non Appliable		
N/A	The PBS is a score that is automatically computed through algorithms and cannot be attributed for the following reasons : - All the fragments of the same reference CDS are antisense - The 5p sequence is missing - All the fragments of the same reference CDS are either all OOF1 or all OOF2 - All the fragments of the same reference CDS lie in the 5' or 3' UTR		

Table 2. (Continued) A complete list of binding identified by Y2H assay

Prey Fragment Analysis

Symbols	Means
✳	The fragment contains the full length CDS
↰	Fragment is fully in 5' UTR
↱	Fragment is fully in 3' UTR
✖	Fragment contains at least one In Frame STOP codon
[NR]	Fragment was found to be non relevant (poor quality, high N density)
IF OOF1 OOF2	With regard to the theoretical frame of each corresponding CDS (GeneBank),fragments are cloned in frame (IF) if they are in the same frame as Gal4AD.In general, polypeptides synthesized from OOF fragments are not considered of biological interest, unless found together with another frame. However, some of the proteins expressed from an OOF fragment can be translated in the correct frame,due to the existence of natural frame-shift events during translation in yeast
??	Unidentified frame when : - The clone sequence is antisense - The 5p sequence is missing
N	Antisense
Start...Stop	Position of the 5p and 3p prey fragment ends, relative to the position of the ATG start codon (A=0)

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop (nt)	Frame	Sense	%Id 5p	%Id 3p	PBS
pB29_A-71	5p/3p	Homo sapiens - APBA2	2251..1808	??	N	100.0	100.0	N/A
pB29_A-44	5p/3p	Homo sapiens - ARMC5	-490..106	IF		99.8	99.7	B
pB29_A-52	5p/3p	Homo sapiens - ARMC5	-388..113	IF		99.0	95.9	B
pB29_A-7	5p/3p	Homo sapiens - BIRC2	519..-186	??	N	87.4	95.5	N/A
pB29_A-67	5p/3p	Homo sapiens - C10orf46	81..811	IF		96.0	91.2	D
pB29_A-142	5p/3p	Homo sapiens - CAMK2D	202..-447	??	N	100.0	98.5	N/A
pB29_A-72	5p/3p	Homo sapiens - CDCA7L var1	373..1380 ✖	IF		100.0	95.7	D
pB29_A-18	5p/3p	Homo sapiens - CEP152	-15..1424	OOF2		99.8	97.2	N/A
pB29_A-138	3p	Homo sapiens - CIZ1	1138..936	??	N		100.0	N/A
pB29_A-39	5p/3p	Homo sapiens - COMMD4	-24..608 ✳ ✖	OOF1		100.0	98.3	N/A
pB29_A-131	5p/3p	Homo sapiens - CUL3	3..1130	IF		97.0	97.2	C
pB29_A-36	5p/3p	Homo sapiens - CUL3	54..1556	IF		99.2	96.7	C
pB29_A-5	5p/3p	Homo sapiens - CYBA	451..-27	??	N	100.0	100.0	N/A
pB29_A-97	5p/3p	Homo sapiens - DAPK1	2958..3804	IF		97.2	97.0	A
pB29_A-132	5p/3p	Homo sapiens - DAPK1	2958..3804	IF		97.5	96.9	A
pB29_A-13	5p/3p	Homo sapiens - DAPK1	2958..3804	IF		95.7	97.0	A
pB29_A-166	5p/3p	Homo sapiens - DAPK1	2958..3804	IF		95.0	94.0	A
pB29_A-20	5p/3p	Homo sapiens - DAPK1	3111..3959	IF		98.0	93.3	A
pB29_A-160	5p/3p	Homo sapiens - DAPK1	3111..3959	IF		96.6	95.5	A
pB29_A-175	5p/3p	Homo sapiens - DAPK1	3192..4035	IF		96.1	95.5	A
pB29_A-73	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		99.8	99.8	A
pB29_A-96	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		99.8	99.8	A
pB29_A-59	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		99.8	99.8	A
pB29_A-181	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		99.8	99.2	A
pB29_A-147	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		98.6	97.1	A
pB29_A-155	5p/3p	Homo sapiens - DAPK1	3213..3713	IF		99.8	99.8	A
pB29_A-24	5p/3p	Homo sapiens - E2F2	45..815	IF		98.2	94.9	D
pB29_A-104	5p	Homo sapiens - EEF1A1	800	OOF2		97.6		N/A
pB29_A-46	5p/3p	Homo sapiens - EIF2AK1	2435..1744	??	N	99.1	96.7	N/A

Table 2. (Continued) A complete list of binding identified by Y2H assay

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop (nt)		Frame	Sense	%Id 5p	%Id 3p	PBS
pB29_A-107	5p/3p	Homo sapiens - EXT2	-78..241		OOF1		100.0	100.0	N/A
pB29_A-37	5p/3p	Homo sapiens - FAM65B	1365..2293		IF		97.2	95.6	D
pB29_A-137	5p/3p	Homo sapiens - FLJ10706	-155..569	✗	OOF2		97.2	98.6	N/A
pB29_A-15	5p/3p	Homo sapiens - FLJ20105	3..1110		IF		92.3	94.4	D
pB29_A-64	5p/3p	Homo sapiens - FTSJD2	-150..837		OOF1		97.1	97.3	N/A
pB29_A-75	3p	Homo sapiens - GOLPH3	..523		??			96.2	N/A
pB29_A-9	5p/3p	Homo sapiens - GSE1	2994..2481		??	N	99.8	99.8	N/A
pB29_A-124	5p/3p	Homo sapiens - H4FJ	63..336	✗	IF		100.0	100.0	D
pB29_A-25	5p/3p	Homo sapiens - HDLBP	3210..3931	✗	IF		95.6	96.6	D
pB29_A-171	5p/3p	Homo sapiens - HK1	2080..1541		??	N	99.8	100.0	N/A
pB29_A-189	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	652..1275	✗	OOF1		98.4	86.4	N/A
pB29_A-38	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	703..976		OOF1		100.0	100.0	N/A
pB29_A-6	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	712..1282	✗	OOF1		100.0	70.3	N/A
pB29_A-91	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	712..1282	✗	OOF1		100.0	69.8	N/A
pB29_A-78	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	712..1282	✗	OOF1		100.0	71.4	N/A
pB29_A-90	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	717..1484	✗	OOF1		99.8	97.1	N/A
pB29_A-116	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	721..1261	✗	OOF1		100.0	90.1	N/A
pB29_A-92	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	721..1261	✗	OOF1		100.0	88.4	N/A
pB29_A-186	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	721..1261	✗	OOF1		100.0	96.0	N/A
pB29_A-19	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	745..1280	✗	OOF1		100.0	70.9	N/A
pB29_A-84	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	745..1280	✗	OOF1		100.0	73.0	N/A
pB29_A-110	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	745..1280	✗	OOF1		99.5	68.2	N/A
pB29_A-54	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	766..1538	✗	OOF1		96.5	94.2	N/A
pB29_A-183	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	766..1538	✗	OOF1		96.3	92.9	N/A
pB29_A-30	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	766..1538	✗	OOF1		97.2	94.1	N/A
pB29_A-35	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	766..1538	✗	OOF1		97.8	93.7	N/A
pB29_A-32	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	778..1222	✗	OOF1		100.0	100.0	N/A
pB29_A-79	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	787..1233	✗	OOF1		81.3	81.3	N/A
pB29_A-11	5p	Homo sapiens - HNRNPA2B1 variant B1	790		OOF1		99.4		N/A
pB29_A-70	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	802..1227	✗	OOF1		73.1	72.0	N/A
pB29_A-125	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	802..1227	✗	OOF1		73.1	73.0	N/A
pB29_A-133	5p/3p	Homo sapiens - HNRNPA2B1 variant B1	802..1227	✗	OOF1		73.1	72.8	N/A
pB29_A-111	3p	Homo sapiens - HUWE1	..6533		??			90.4	D
pB29_A-145	5p/3p	Homo sapiens - HUWE1	5849..6533		IF		99.7	97.7	D
pB29_A-184	5p/3p	Homo sapiens - HUWE1	5849..6533		IF		99.8	96.8	D
pB29_A-108	5p/3p	Homo sapiens - HUWE1	5849..6533		IF		99.8	98.1	D

Table 2. (Continued) A complete list of binding identified by Y2H assay

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop (nt)	Frame	Sense	%Id 5p	%Id 3p	PBS
pB29_A-48	5p/3p	Homo sapiens - HUWE1	5849..6533		IF		99.8	97.0	D
pB29_A-61	5p/3p	Homo sapiens - HUWE1	5849..6533		IF		99.8	96.9	D
pB29_A-144	5p/3p	Homo sapiens - IMMT	2013..2596 ✖	IF		99.5	99.7	E
pB29_A-33	5p	Homo sapiens - KDM2A	963	IF		97.6		E
pB29_A-121	5p/3p	Homo sapiens - KDM2A	999..2416	IF		97.6	96.9	E
pB29_A-114	5p/3p	Homo sapiens - KDM2A	999..2416	IF		98.6	96.9	E
pB29_A-14	5p/3p	Homo sapiens - KDM2A	1100..1868	IF		99.8	96.0	E
pB29_A-27	5p/3p	Homo sapiens - KDM2A	1100..1868	IF		99.7	95.0	E
pB29_A-28	5p/3p	Homo sapiens - KDM2A	1100..1868	IF		100.0	92.5	E
pB29_A-179	5p/3p	Homo sapiens - KDM2A	1379..2433	IF		100.0	96.5	E
pB29_A-22	5p/3p	Homo sapiens - KDM2A	1379..2433	OOF1		100.0	95.7	E
pB29_A-53	5p/3p	Homo sapiens - KDM2A	1379..2433	IF		100.0	95.2	E
pB29_A-162	5p/3p	Homo sapiens - KIF11	1851..2811	IF		94.0	95.3	D
pB29_A-55	5p/3p	Homo sapiens - MARCH6	753..10	??	N	97.3	97.1	N/A
pB29_A-163	5p	Homo sapiens - MTF2	87	IF		92.7		D
pB29_A-66	5p	Homo sapiens - MTF2	87	IF		96.5		D
pB29_A-29	5p/3p	Homo sapiens - NBEA	410..1181	OOF2		99.8	84.0	N/A
pB29_A-57	5p/3p	Homo sapiens - NDUFS2	1549..1099	??	N	93.0	95.7	N/A
pB29_A-168	5p/3p	Homo sapiens - NFE2L1	551..1120	OOF2		99.6	95.4	B
pB29_A-1	5p/3p	Homo sapiens - NFE2L1	553..1025	OOF1		100.0	98.7	B
pB29_A-158	5p/3p	Homo sapiens - NUP153	2788..3269	OOF1		100.0	100.0	N/A
pB29_A-134	5p/3p	Homo sapiens - OSBPL5	1927..1451	??	N	100.0	100.0	N/A
pB29_A-120	5p/3p	Homo sapiens - OSBPL5	1927..1451	??	N	100.0	100.0	N/A
pB29_A-130	5p/3p	Homo sapiens - OSBPL5	1927..1451	??	N	99.6	99.0	N/A
pB29_A-85	5p/3p	Homo sapiens - PCBP1	177..850	IF		98.8	95.5	D
pB29_A-76	5p/3p	Homo sapiens - PCBP1	177..850	IF		98.4	97.0	D
pB29_A-68	5p	Homo sapiens - PIK3R6	1872..1512	??	N	97.5		N/A
pB29_A-112	3p	Homo sapiens - POLR2A	..4850	??			91.3	B
pB29_A-187	5p/3p	Homo sapiens - POLR2A	4220..4740	OOF2		98.7	99.2	B
pB29_A-109	5p/3p	Homo sapiens - POLR2A	4285..4850	OOF1		100.0	98.3	B
pB29_A-101	5p/3p	Homo sapiens - RAB11FIP3	2389..2080	??	N	99.7	99.7	N/A
pB29_A-93	5p/3p	Homo sapiens - RAB11FIP3	2549..2066	??	N	99.8	96.0	N/A
pB29_A-177	5p/3p	Homo sapiens - RAB11FIP3	2549..2066	??	N	99.8	99.2	N/A
pB29_A-17	5p	Homo sapiens - RAB11FIP3	2549 ⬈	??	N	94.3		N/A
pB29_A-47	5p/3p	Homo sapiens - RPN2	261..589	IF		100.0	100.0	D
pB29_A-87	5p/3p	Homo sapiens - SERTAD3	62..965 ✖	IF		99.1	95.8	D
pB29_A-140	5p/3p	Homo sapiens - SLC11A2	3250..3658 ⬈ ✖	OOF1		99.8	99.7	N/A
pB29_A-94	5p/3p	Homo sapiens - SNX14	2888..2131	??	N	98.5	94.4	N/A
pB29_A-113	5p/3p	Homo sapiens - SON	5709..5128	??	N	99.5	99.3	N/A
pB29_A-169	5p/3p	Homo sapiens - SON	5709..5128	??	N	98.8	97.9	N/A
pB29_A-122	5p/3p	Homo sapiens - SON	5709..5128	??	N	99.3	98.6	N/A
pB29_A-173	5p/3p	Homo sapiens - SON	5709..5128	??	N	99.3	99.0	N/A
pB29_A-141	5p/3p	Homo sapiens - SP3	1253..2188	OOF2		98.5	97.2	N/A
pB29_A-148	5p/3p	Homo sapiens - SP3	1253..2188	OOF2		97.5	96.4	N/A
pB29_A-188	5p/3p	Homo sapiens - SP3	1253..2188	OOF2		96.4	95.4	N/A
pB29_A-167	5p/3p	Homo sapiens - SP3	1283..2599 ✖	OOF2		96.9	94.6	N/A
pB29_A-176	5p/3p	Homo sapiens - SP3	1370..2623 ✖	OOF2		97.8	95.9	N/A
pB29_A-69	5p/3p	Homo sapiens - SP4	1190..2190	OOF2		94.5	95.7	N/A
pB29_A-3	5p/3p	Homo sapiens - STK24	-55..420	IF		100.0	100.0	B
pB29_A-63	3p	Homo sapiens - STK24	..584	??			96.0	B

Table 2. (Continued) A complete list of binding identified by Y2H assay

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop (nt)		Frame	Sense	%Id 5p	%Id 3p	PBS
pB29_A-190	5p	Homo sapiens - STK24	9		IF		94.6		B
pB29_A-146	5p/3p	Homo sapiens - TAF1	3000..4185		IF		97.0	96.9	E
pB29_A-88	5p/3p	Homo sapiens - TAF1	3060..3939		IF		94.8	96.8	E
pB29_A-119	5p/3p	Homo sapiens - TAF1	3060..3939		IF		94.4	95.8	E
pB29_A-2	5p/3p	Homo sapiens - TAF1	3429..4255		IF		96.6	97.9	E
pB29_A-12	5p/3p	Homo sapiens - TAF1	3429..4255		IF		95.2	97.9	E
pB29_A-4	5p/3p	Homo sapiens - TAF1	3429..4255		IF		96.7	97.6	E
pB29_A-62	5p/3p	Homo sapiens - TAF1	3429..4255		IF		96.1	97.0	E
pB29_A-105	5p/3p	Homo sapiens - TAF1	3429..4255		IF		94.4	97.7	E
pB29_A-118	5p	Homo sapiens - TAF1	3429		IF		90.2		E
pB29_A-123	5p/3p	Homo sapiens - TAF1	3429..4255		IF		91.7	100.0	E
pB29_A-152	5p/3p	Homo sapiens - TAF1	3429..4255		IF		95.9	97.8	E
pB29_A-159	5p/3p	Homo sapiens - TAF1	3429..4255		IF		95.7	96.7	E
pB29_A-164	5p/3p	Homo sapiens - TAF1	3429..4255		IF		95.1	96.9	E
pB29_A-170	5p/3p	Homo sapiens - TAF1	3429..4255		IF		92.1	95.8	E
pB29_A-117	5p/3p	Homo sapiens - TCF12	828..1470		IF		98.1	94.7	D
pB29_A-115	5p/3p	Homo sapiens - TOX4	467..1327		IF		99.8	93.4	E
pB29_A-58	5p/3p	Homo sapiens - TTF1	1146..1920		IF		99.4	97.8	B
pB29_A-127	5p/3p	Homo sapiens - TTF1	1287..2331		IF		97.5	96.4	B
pB29_A-31	5p/3p	Homo sapiens - TTF1	1287..2331		IF		96.1	95.1	B
pB29_A-157	5p/3p	Homo sapiens - TUBA1B	43..694		OOF1		99.1	96.3	N/A
pB29_A-45	5p/3p	Homo sapiens - TUBA1B	43..694		OOF1		99.7	94.1	N/A
pB29_A-23	5p/3p	Homo sapiens - TUBB	1179..914		??	N	100.0	100.0	N/A
pB29_A-149	5p/3p	Homo sapiens - WDTC1	1369..2036	✖	OOF1		99.3	97.2	N/A
pB29_A-26	5p/3p	Homo sapiens - ZBTB38	1095..1866		IF		98.3	95.6	E
pB29_A-106	5p/3p	Homo sapiens - ZBTB40	2124..2655		IF		100.0	100.0	D
pB29_A-89	5p/3p	Homo sapiens - ZBTB40	2124..2655		IF		98.9	92.9	D
pB29_A-172	5p/3p	Homo sapiens - transcription factor Sp1 isoform a	1454..2135		OOF2		98.4	97.5	N/A
pB29_A-86	5p/3p	Homo sapiens - GenMatch GI:2822137	-1..957	✖	IF		99.6	96.8	D
pB29_A-83	5p/3p	Homo sapiens - GenMatch GI:2564750	-1..1199	✖	IF		97.5	97.0	D
pB29_A-80	5p/3p	Homo sapiens - GenMatch GI:339895989	-1..319	✖	IF		100.0	100.0	D
pB29_A-178	5p	Homo sapiens - GenMatch GI:11414487	-1..640	✖	IF		100.0		D
pB29_A-126	5p/3p	Homo sapiens - GenMatch GI:18308345	-1..1076	✖	IF		100.0	74.5	D
pB29_A-51	5p/3p	Homo sapiens - GenMatch GI:239740417	-2..348	✖	IF		100.0	100.0	D
pB29_A-81	5p	Homo sapiens - GenMatch GI:13357369	-1		IF		100.0		D
pB29_A-49	5p/3p	Homo sapiens - GenMatch GI:20330735	-1..413	✖	IF		99.8	97.8	D
pB29_A-43	5p/3p	Homo sapiens - GenMatch GI:789912913	-1..583	✖	IF		100.0	100.0	D
pB29_A-65	5p	Homo sapiens - GenMatch GI:23307933	-2		IF		100.0		D
pB29_A-56	5p/3p	Homo sapiens - GenMatch GI:14971194	-1..642	✖	IF		100.0	96.4	D
pB29_A-182	5p/3p	Homo sapiens - GenMatch GI:15217393	-1..828	✖	IF		97.6	94.4	D
pB29_A-180	5p/3p	Homo sapiens - GenMatch GI:15145586	-1..397	✖	IF		100.0	100.0	D

Table 2. (Continued) A complete list of binding identified by Y2H assay

Clone Name	Type Seq	Gene Name (Best Match)	Start..Stop (nt)	Frame	Sense	%Id 5p	%Id 3p	PBS
pB29_A-174	5p/3p	Homo sapiens - GenMatch GI:17389904	-1..995	✖	IF		96.6	95.9	D
pB29_A-161	5p/3p	Homo sapiens - GenMatch GI:13443277	-1..298	✖	IF		100.0	100.0	D
pB29_A-154	5p/3p	Homo sapiens - GenMatch GI:133923366	-1..636	✖	IF		99.5	99.7	D
pB29_A-151	5p/3p	Homo sapiens - GenMatch GI:7161187	-1..556	✖	IF		99.1	100.0	D
pB29_A-129	5p/3p	Homo sapiens - GenMatch GI:307133696	-1..246	✖	IF		100.0	100.0	D
pB29_A-128	5p/3p	Homo sapiens - GenMatch GI:158853045	-1..653	✖	IF		99.6	98.8	D
pB29_A-10	5p/3p	Homo sapiens - GenMatch GI:5042384	-1..346	✖	IF		99.7	100.0	D
pB29_A-102	5p/3p	Homo sapiens - GenMatch GI:18854952	-1..675	✖	IF		99.4	99.6	D
pB29_A-95	5p/3p	Homo sapiens - GenMatch GI:161621413	-1..879	✖	IF		92.3	93.5	D

Chapter 3

ARMC5 is part of an RPB1-specific ubiquitin ligase implicated in adrenal gland hyperplasia

[1]Linjiang Lao, [1,2]Isabelle Bourdeau,[3-6]Lucia Gagliardi, [1]Xiao He,[1]Wei Shi, [7]Bingbing Hao,[7]Minjia Tan, [1]Yan Hu,[1]Junzheng Peng, [3,4]David J. Torpy,[3,5,8,9]Hamish S. Scott,[1,2]Andre Lacroix,[1,*]Hongyu Luo, and [1,10,*]Jiangping Wu

From the [1] Centre de recherché, [2] Endocrinology Division and [10] Nephrology Division, Centre hospitalier de l'Université de Montréal (CHUM), Montreal, Quebec, Canada; [3] Adelaide Medical School, University of Adelaide, Adelaide, Australia; [4]Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, Australia; [5] Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia; [6] Endocrine and Diabetes Unit, The Queen Elizabeth Hospital, Adelaide, Australia; [7] Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai, China; [8] Centre for Cancer Biology, an alliance between SA Pathology and the University of South Australia, Adelaide, Australia; [9] UniSA Clinical and Health Sciences, University of South Australia, Adelaide, Australia

[*]Address correspondence to Dr. Hongyu Luo, CRCHUM, 900 Saint-Denis Street, Room R12.426, Montreal, Quebec, Canada H2X 0A9. Telephone: (514) 890-8000 Extension 25394, Fax: (514) 412-7944, e-mail: hongyu.luo@umontreal.ca; or Dr. Jiangping Wu, CRCHUM, 900 Saint-Denis Street, Room R12.428, Montreal, Quebec, Canada H2X 0A9. Telephone: (514) 890-8000 Extension 25164, Fax: (514) 412-7944, e-mail: jianping.wu@umontreal.ca

This work is under revision in *Nucleic Acids Research*.

3.1. Summary

ARMC5 is a protein implicated in several pathological conditions, but its function is unknown. Ubiquitin ligase (E3) specific for RPB1 has the capability to control the degradation of RPB1, and consequently, the Pol II pool size. We demonstrated that ARMC5 physically interacted with CUL3 and RPB1. *Armc5* deletion caused significant RPB1 accumulation in normal organs, accompanied by reduced RPB1 ubiquitination. ARMC5, CUL3, and RBX1 formed an active E3 for RPB1. The compromised RPB1 degradation caused by *Armc5* deletion did not lead to increased Pol II stalling. RPB1 was highly accumulated in the adrenal gland from primary bilateral macronodular adrenal gland hyperplasia (PBMAH) patients with *ARMC5* mutations. The mutant ARMC5 in these patients had compromised binding with RPB1. *In toto*, we discovered that ARMC5 is part of a novel RPB1-specific E3 largely responsible for RPB1 ubiquitination and degradation in unperturbed normal tissues and cells. Its deletion/mutation results in an enlarged Pol II pool, which likely dysregulates downstream effector genes, leading to pathogenesis in humans and KO mice.

Keywords: ARMC5, ubiquitin ligase, Cullin3, RPB1 degradation, RNA polymerase II pool size, primary bilateral macronodular adrenal gland hyperplasia, STAR, glucocorticoid biogenesis

3.2. Introduction

DNA-dependent RNA polymerase II (Pol II) is responsible for the synthesis of all mRNA and some small RNA[334]. It has 12 subunits, and RPB1 is its largest subunit[335]. Protein biosynthesis and degradation are processes to determine the protein abundance in a cell. Thus, it is logical to assume that RPB1 degradation is part of the equation for maintaining the homeostasis of the Pol II pool size. The effect of an abnormal Pol II pool size is under-appreciated. It is assumed that since Pol II is implicated in the transcription of all genes, its pool size will affect the transcription of all expressed ones. The validity of this assumption is not confirmed.

During transcription, if template DNA is damaged or the cells are under stress, Pol II will stall until the damage is repaired, or the stress is relieved[336,337,338,339]. In case the stalling becomes long-lasting, it is believed that Pol II will be channeled to proteasomes for degradation to resume transcription[340,341,342,338,337,263,343,264].

Proteins need to be ubiquitinated before being degraded by proteasomes. Such ubiquitination depends on a cascade of three enzymes, i.e., E1 (Ub-activating enzyme), E2 (Ub-conjugating enzyme), and E3 (Ub ligase)[344]. There are two E1s (UBA1 and UBA6) in humans, but they do not have substrate specificity[71]. There are a total of 40 known E2s that have limited specificity[345]. E3 decides the substrate specificity. Each protein has its specific E3, sometimes more than one, although each E3 can have several substrates[344]. There are three families of E3s: Really Interesting New Gene (RING)-type E3s (single or multiple subunits), Homologous to the E6-AP Carboxyl Terminus (HECT)-type E3s, and RING-between-RINGS (RBR)-type E3s[346]. The RING-type E3s are the largest family. A multiple subunit RING E3 contains a RING-finger protein (e.g., ROC1/RBX1), a cullin (CUL) protein (CUL1, 2, 3, 4A, 4B, 5 and 7), and a substrate recognition unit[347]. CUL3 interacts with a RING-finger protein, RBX1. CUL3 has a BTB-interacting domain, and it recruits a BTB domain-containing protein as its substrate recognition unit to form an active E3[348].

Given the critical roles of RPB1 degradation, cell biologists are highly interested in identifying RPB1-specific E3s[264,263]. Several such E3s have been reported in yeasts and mammalian cells[275,268,286,194,285,263,343,279]. However, most of these E3s only have proven functions in cells treated with irradiation or DNA-damaging agents. A few of them showed E3 activities in unperturbed cell lines according to knockdown studies, which have not been extended to tissues or organs[194,279]. An RPB1-specific E3 that is active under physiological conditions without massive DNA damage in normal tissues and organs must exist, and cell biologists are keen to find such an E3.

ARMC5 is a protein containing an armadillo (ARM) domain, which comprises multiple ARM repeats in its N-terminus and a BTB domain towards its C-terminus. Human and

mouse ARMC5 proteins share ∼90% amino acid (aa) sequence homology and have similar tertiary structures[316,315]. Mouse ARMC5 is 926 aa in length (NP_666317.2). Human ARMC5 has several isoforms due to different splicing, and in one case, due to two additional exons at the 5'-end of the gene[349]. The longest human ARMC5 isoform has 1030 aa (NP_001275696.1), and the most abundant one in terms of tissue distribution and expression levels is 935 aa long (NP_001098717.1)[349].

Primary bilateral macronodular adrenal gland hypertrophy (PBMAH) is a rare disease with an incidence of 2 – 4 per million people[350]. PBMAH is usually diagnosed at a late stage of life, at 40 – 50 years of age. These patients have variable degrees of endogenous hypercortisolism, which can lead to overt Cushing's syndrome[351]. During our investigation of the function of ARMC5 in gene knockout mice, several groups identified *ARMC5* mutations in about 21 – 26% of PBMAH patients[1,352,19,3]. The biosynthesis of cortisol per adrenocortical cell is relatively inefficient due to partial deficiency of several steroidogenic enzymes[353,1,354,355] in these PBMAH patients. Still, due to the massively enlarged nodular adrenal glands, they have varying degrees of biochemical hypercortisolism[350,1]. No information on other functions of ARMC5 was available until our recent publication[8]. We reported that *Armc5* bi-allelic KO mice were small in body size and presented compromised T-cell proliferation and T-cell immune responses[8]. Aged KO mice showed adrenal gland hypertrophy accompanied by moderately augmented blood cortisol levels[8], similar to that observed in PBMAH patients, suggesting that *Armc5* deletion alone is sufficient to cause a PBMAH-like condition. In corroborating our findings, increased blood cortisol levels in the 30% of aged mice with monoallelic *Armc5* deletion were reported by Berthon et al.[10].

ARMC5's mechanisms of action are unknown. ARMC5 contains no conserved enzymatic motifs or domains, and is thus unlikely an enzyme *per se*. Its function must depend on its interaction with other molecules. To identify the partners of ARMC5, we conducted a yeast 2-hybrid assay (Y2H), using human ARMC5 as bait. RPB1, CUL3, and ARMC5 itself were among the top hits[8].

In the present study, we demonstrated that ARMC5 physically interacted with CUL3 and RPB1 and was the substrate recognition subunit of a novel multiple-unit RING-finger E3. This E3 is largely responsible for RPB1 ubiquitination in unperturbed normal tissues and cells tested. *Armc5* deletion led to failed RPB1 degradation. It is generally believed that RPB1 degradation is needed to resolve stalled Pol II, and if the stalling persists, there will be reduced transcription. However, we did not observe augmented Pol II stalling nor a generally reduced transcription in *Armc5* KO cells in spite of the failed RPB1 degradation. Among 1486 differentially expressed genes in KO adrenal glands, most of them were up-regulated, presumably due to an enlarged Pol II pool size. We further showed that the adrenal gland nodules from PBMAH patients carrying *ARMC5* mutations presented highly elevated RPB1 protein levels, demonstrating the relevance of our findings to human PBMAH.

3.3. Results

3.3.1. ARMC5 physically interacted with CUL3 and RPB1

Our Y2H assay has revealed that CUL3, RPB1, and ARMC5 itself are potential binding partners of ARMC5[8]. Several additional methods were used to confirm such association. HEK293 cells were transfected with a plasmid expressing human ARMC5-HA, and ARMC5 was affinity precipitated (AP) with anti-HA Ab (**Figure 3.1a**). The precipitates were analyzed with LC-MS/MS. The AP-LC-MS/MS experiments were conducted in three biological replications. Protein hits satisfying both following conditions in any of the 1 – 3 biological replicates were listed in **Figure 3.1b**. 1) The protein had equal or more than three peptides corresponding to its sequence in the test sample; 2) the number of the peptides in the test sample was more than 2-fold larger than that in the controls. Two proteins (i.e., RPB1 and CUL3) identified in Y2H[8], were also found to associate with ARMC5 in this LC-MS/MS analysis. CUL3 binds with a RING-finger protein RBX1 and forms a multiple-unit RING-finger E3, using a protein containing a BTB domain as its substrate recognition subunit[348]. We hypothesized that ARMC5, which harbors a BTB domain at its C-terminus, was the substrate recognition subunit of a novel RPB1-specific multiple-unit RING-finger

E3. Consistent with this hypothesis, some components of the ubiquitination system, such as an E1 (UBE) and ubiquitin (UBC) (shown in blue in the list) were also found in ARMC5 co-precipitates. Detailed information about these hits, including the number of peptides representing a given protein and the fold change of the number of these peptides in the test versus control samples, is provided in **Table 3.S.1**, in which a laxer criterion was employed. The protein had equal or more than two peptides corresponding to its sequence, and with a 2-fold higher number of peptides in the test sample versus the controls were included. This allows a more comprehensive appreciation of the LC-MS/MS results.

Immunoprecipitation was employed to further prove the interaction among ARMC5, CUL3, and RPB1. CUL3-Myc were found in ARMC5-HA precipitates from HEK293 cells transfected with plasmids expressing ARMC5-HA and CUL3-Myc (**Figure 3.1c**), and endogenous RPB1 was present in ARMC5-HA precipitates from HEK293 cells transfected plasmid expressing ARMC5-HA (**Figure 3.1d**). This confirmed that ARMC5 interacted with CUL3 and RPB1, respectively. We previously showed in the Y2H analysis that ARMC5 bait interacts with ARMC5 prey[8], suggesting that ARMC5 can dimerize with itself. Indeed, ARMC5-FLAG was found in ARMC5-HA precipitates from HEK293 cells transfected with plasmids expressing ARMC5-FLAG and ARMC5-HA (**Figure 3.1e**), indicating that ARMC5 partnered with itself and formed homodimers, at least. Using HEK293 cells transfected with plasmids expressing both ARMC5-HA and CUL3-Myc, we first precipitated CUL3-Myc from the cell lysates and then further precipitated ARMC5-HA from the first-round precipitates. CUL3-Myc, ARMC5-HA, and endogenous RPB1 were all found in the 2nd precipitation (**Figure 3.1f**), indicating that ARMC5, CUL3, and RPB1 formed a tri-molecule complex. To alleviate the concern of artifacts caused by excessive protein expression in HEK293 cells, we transfected adrenal gland cortex carcinoma SW-13 cells with plasmids expressing ARMC5-HA. We detected both endogenous CUL3 and RPB1 in the anti-HA precipitates (**Figure 3.1g**), proving that ARMC5 was associated with endogenous CUL3 and RPB1 in these cells. In these experiments, ARMC5 was always detected as two bands of 130 kD and 100 kD in size in immunoblotting due to proteolysis. Due to the poor specificity of

all anti-ARMC5 Abs currently available, we were not able to confirm the interactions among endogenous ARMC5, RPB1, and CUL3.
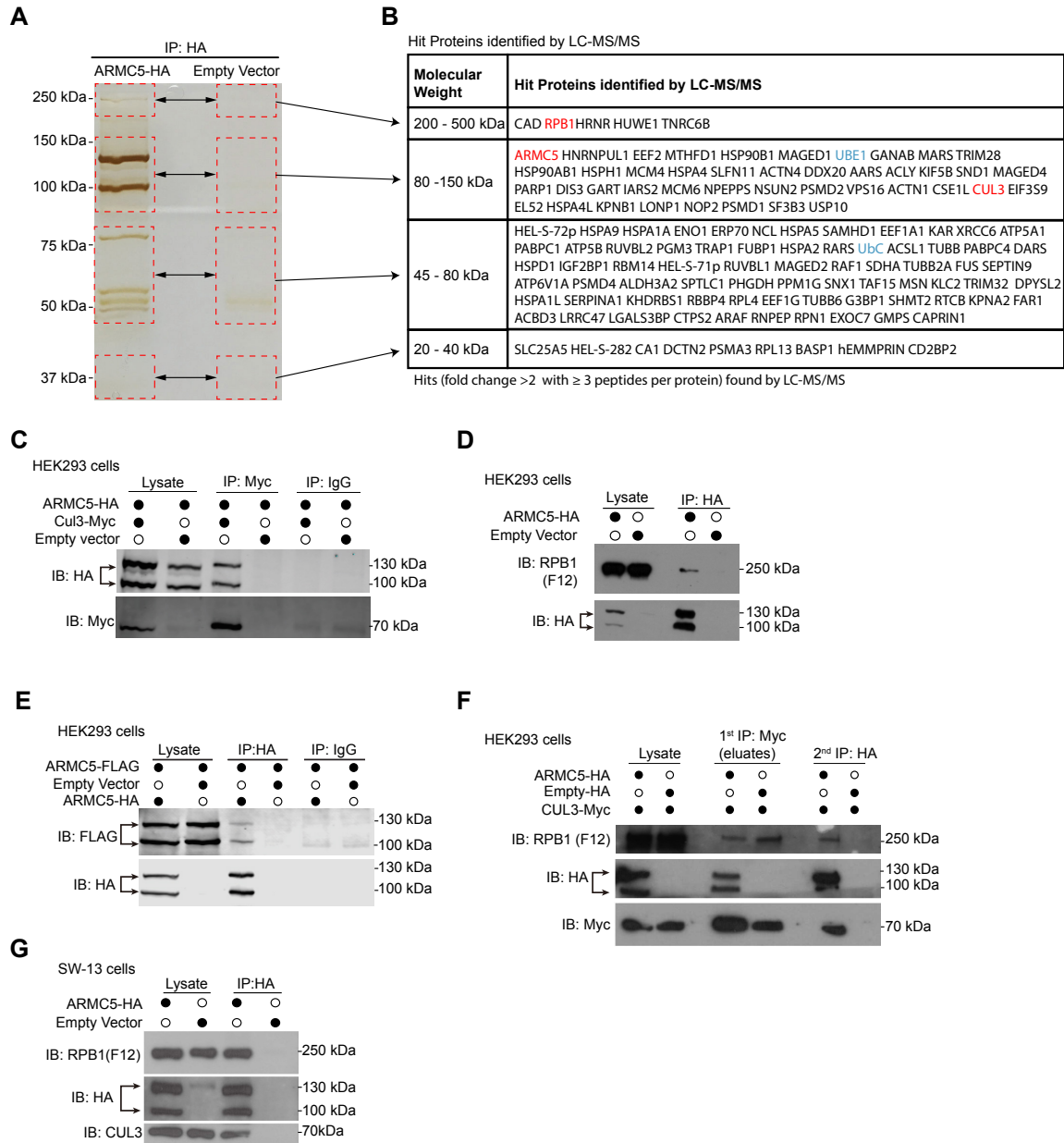
Fig. 3.1. ARMC5 forms a complex with CUL3, RPB1, and itself *(see next page)*

Fig. 3.1 *(previous page).* **a**. Silver staining of ARMC5 precipitates. Transfected ARMC5-HA in HEK293 cells were precipitated with anti-HA Ab. The regions (rectangles with dashed lines) with visible bands in the test sample and the corresponding positions in the empty vector-transfected lane were excised and were analyzed by LC-MS/MS. Three independent experiments were conducted, and a representative gel with silver staining is shown. **b**. Proteins found in the ARMC5 precipitates according to the LC-MS/MS analysis. Proteins met with the following two conditions in any of the biological replicates (200 – 500 kDa: duplicates; 80 – 150 kDa: duplicates; 45 – 80 kDa: triplicates; 30 kDa: once) were listed. 1) The protein had equal or more than three peptides corresponding to its sequence in the ARMC5-HA transfected sample; 2) the number of the peptides in the ARMC5-HA-transfected sample was more than 2-fold larger than that in the empty vector control. The gel pieces from which the proteins were derived were indicated. **c**. ARMC5 interacted with CUL3. HEK293 cells were transfected with plasmids expressing ARMC5-HA and CUL3-Myc. Cell lysates were precipitated with anti-HA Ab and immunoblotted with anti-Myc. **d**. ARMC5 interacted with RPB1. HEK293 cells were transfected with plasmids expressing ARMC5-HA. Cell lysates were precipitated with anti-HA Ab and immunoblotted with anti-RPB1 N-terminal Ab (clone F12). **e**. ARMC5 interacted with itself. HEK293 cells were transfected plasmids expressing ARMC5-HA and ARMC5-FLAG. Cell lysates were precipitated with anti-HA Ab and immunoblotted with anti-FLAG Ab. **f**. ARMC5, CUL3, and RPB1 formed tri-molecule complexes. HEK293 cells were transfected with plasmids expressing ARMC5-HA and CUL3-Myc. Cell lysates were first precipitated with anti-Myc Ab and eluted with Myc peptides. The precipitates were then re-precipitated with anti-HA Ab. The secondary precipitates were blotted with anti-RPB1 N-terminus Ab (clone F12). **g**. ARMC5 interacted with endogenous CUL3 and RPB1 in adrenal gland cortical carcinoma SW-13 cells. SW-13 cells were transfected with plasmids expressing ARMC5-HA. Cell lysates were precipitated with anti-HA. The precipitates were immunoblotted with anti-CUL3 or anti-RPB1 N-terminus Ab (clone F12). In all the experiments, the lysates were also immunoblotted with Abs against HA, MYC, FLAG to demonstrate the effectiveness of transfection. Empty vectors were used in transfection as controls. IgG was employed in immunoprecipitation as a control. All experiments were conducted more than three times, and representative results are shown.

3.3.2. Identification of regions of interaction in ARMC5, CUL3, and RPB1 molecules

Human ARMC5 contains an ARM domain at its N-terminus (aa143 – 444) and a BTB domain at its C-terminus (aa748 – 816) (**Figure 3.2a**). CUL3 has 3 Cullin repeats in its N-terminus (aa30 – 377), followed by a Cullin homology domain (aa378 – 675). In its

C-terminus, there is a neddylation site spanning from aa 695 to 762. We created deletion mutations of both molecules to identify regions of their interaction. A CUL3 deletion mutant containing only the cullin repeats plus the 30-aa N-terminal segment (CUL3(aa1 – 376)-Myc; **Figure 3.2b**) could still bind to full-length ARMC5-HA (**Figure 3.2b**). Conversely, a CUL3 mutant with deletion of the cullin repeats plus its following 9-aa (CUL3(Δaa31 – 385)-HA) was no longer bound to full-length ARMC5-FLAG, although the full-length CUL3-HA still did in the same experiment (**Figure 3.2c**). On the other hand, CUL3 with the cullin homology domain deleted (CUL3 (Δaa377 – 675)-Myc) or with the C-terminal neddylation site deleted (CUL3(Δaa695 – 762)-Myc) still associated well with ARMC5 (**Figure 3.2d**). These deletion studies demonstrated that CUL3 used its cullin repeats in its N-terminus to interact with ARMC5. A short ARMC5 C-terminal segment containing the BTB domain (ARMC5(aa748 – 935)-HA) was sufficient to precipitate down CUL3-Myc (**Figure 3.2e**). CUL3-Myc could precipitate the full-length ARMC5-HA but not the mutant ARMC5-HA (ARMC5(Δ748 – 816)-HA) with the BTB domain deleted (**Figure 3.2f**). This result revealed that the BTB domain in ARMC5 was necessary and sufficient for ARMC5 to associate with the cullin repeats of CUL3.

We next investigated the interaction between ARMC5 and RPB1. We obtained an RPB1 mutant with its C-terminal repeats deleted (FLAG-RPB1-ΔCTD; Addgene). FLAG-RPB1-ΔCTD could precipitate ARM5-HA as efficiently as the full-length FLAG-RPB1 (**Figure 3.2g**), suggesting that the RPB1 sequence upstream of the CTD was essential for ARMC5 binding. ARMC5 mutants with deletions of the following regions were generated: the N-terminal sequence (ARMC5(Δaa2 – 142)-HA) before the ARM domain; the ARM domain (ARMC5(Δaa143 – 444)-HA); the sequence between the ARM domain and BTB domain (ARMC5(Δaa445 – 747)-HA); the BTB domain (ARMC5(Δaa748 – 816)-HA); and the C-terminal sequence after the BTB domain (ARMC5(Δaa817 – 935)-HA). These mutants, as well as full-length ARMC5-HA, were expressed in HEK293 cells. ARMC5(Δaa2 – 142)-HA and ARMC5(Δaa143 – 444)-HA could not pull down endogenous RPB1, while the full-length ARMC5-HA and the rest of the mutants could (**Figure 3.2h**). This indicates
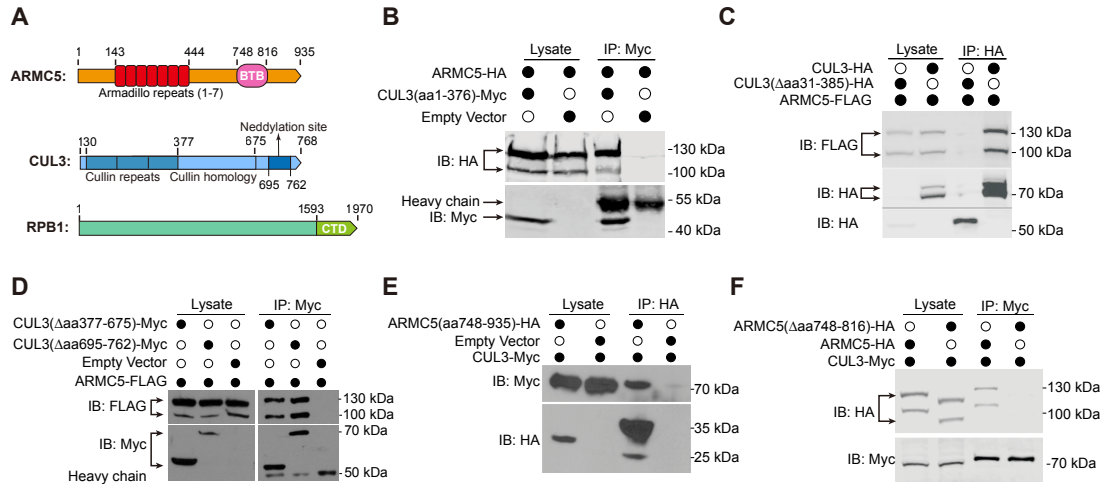
Fig. 3.2. Identification of the regions of interaction in ARMC5, CUL3, and RPB1 molecules a. Schematics of key domains of human ARMC5, CUL3, and RPB1 molecules. **b**. The cullin repeats of CUL3 interacted with ARMC5. HEK293 cells were transfected with plasmids expressing CUL3 N-terminal cullin repeats (CUL3(aa1 – 376)-Myc) and full-length ARMC5-HA. The lysates were precipitated with anti-Myc and blotted with anti-HA and anti-Myc Abs. **c**. CUL3 with cullin repeats deleted no longer interacted with ARMC5. HEK293 cells were transfected plasmids expressing either full-length CUL3 or CUL3 with culling repeats deleted (CUL3(Δaa31 – 385)-HA) and full-length ARMC5-HA. The lysates were precipitated with anti-HA Ab and blotted with anti-FLAG and anti-HA Abs. **d**. The CUL3 cullin homology domain and its C-terminal fragment were not necessary for interaction with ARMC5. HEK293 cells were transfected plasmids expressing CUL3 with the cullin homology domain deleted (CUL3(Δaa377 – 676)-Myc) or with its C-terminal sequence, including the neddylation site deleted (CUL3(Δaa695 – 762)-Myc), and full-length ARMC5-FLAG. The lysates were precipitated with anti-Myc Ab and blotted with anti-FLAG and anti-Myc Abs. **e**. The ARMC5 C-terminal sequence containing the BTB domain was sufficient to interact with CUL3. HEK293 cells were transfected plasmids expressing CUL3-Myc and the ARMC5 C-terminal sequence containing the BTB domain plus a 119-aa region (ARMC5(aa748 – 935)-HA). The lysates were precipitated with anti-HA Ab and blotted with anti-HA or anti-Myc Abs. **f.** ARMC5 without BTB domain no longer bound to CUL3. HEK293 cells were transfected with plasmids expressing CUL3-Myc and BTB domain-deleted ARMC5-HA (ARMC5(Δaa748 – 816)-HA). The lysates were precipitated with anti-Myc and blotted with anti-HA or anti-Myc Abs.

that the ARM domain and the sequence before it at the N-terminus are essential for RPB1 binding. We noticed that ARMC5 mutants with deletions of the sequence between the ARM domain and BTB domain (ARMC5(Δaa445 – 747)-HA) and the sequence after the BTB domain (ARMC5(Δaa817 – 935)-HA) were less effective in pulling down RPB1, compared

to the full-length ARMC5-HA. It is possible that these regions also contributed to RPB1 binding, albeit to a lesser extent.

The essential regions needed for ARMC5 and ARMC5 homologous interaction were assessed by ARMC5 deletion mutants ARMC5(Δaa2 – 142)-HA, ARMC5(Δaa143 – 444)-HA, ARMC5(Δaa445 – 747)-HA, and ARMC5(Δ aa748 – 816)-HA (**Figure 3.2i**). The full-length ARMC5-FLAG could precipitate well the full-length ARMC5-HA and the other HA-tagged deletion mutants, except that it could only weakly precipitate the mutant with the aa143 – 444 deletion, which corresponded to the ARM domain.

The interacting regions between RPB1 and ARMC5, CUL3 and ARMC5, and ARMC5 and ARMC5 are depicted in **Figure 3.2j**. For the interaction regions between ARMC5 and RPB1, the lower intensity of the gray shade indicates a lesser degree of contribution to the binding between these two molecules.

CUL3 is known to interact with a RING-finger protein RBX1, the enzymatic component of the multiple subunit RING-finger E3s, as previously reported[356]. This was confirmed by us in HEK293 cells (data not shown). This RPB1-ARMC5-CUL3-RBX1 complex has the necessary features of an RPB1-specific multiple subunit RING-finger E3, with ARMC5 as the substrate recognition unit. We later demonstrated that an E2 UBE2E1 participated in the function of this E3. A 2D cartoon in **Figure 3.2k** illustrates the proposed structure of the dimeric RPB1-ARMC5-CUL3-RBX1-UBE2E1-UBC complex based on our results and literature. We extracted X-ray crystallographic 3D information of the components in the complex, i.e., RPB1, CUL3, RBX1, UBE2E1, and UBC, from the Protein Database[357,358,359,360]. The 3D structure of ARMC5 has not been determined, but was predicted by AlphaFold[361]. A 3D model of this E3 complex was constructed (**Figure 3.2l**) providing a better visual perspective. Although the 3D structure of each component is reliable, we caution that the contour of the complex is quite speculative.

3.3.3. *Armc5* KO led to an accumulation of RPB1 in normal organs under a physiological condition and in adrenal glands from PBMAH patients

RPB1 is mainly a nuclear protein. Its E3 should also have a nuclear presence. Our earlier report showed that when ARMC5 was overexpressed in HEK293, ARMC5 signals were found mainly in the cytosol[8]. While this remained to be true, we found that in the presence of a nuclear export blocker leptomycin B, ARMC5 was easily detectable in the nuclei (**Figure 3.3a**). This suggested that ARMC5 did enter the nuclei, but at the same time, there was an active shuttling of this molecule between the cytosol and nuclei. We also
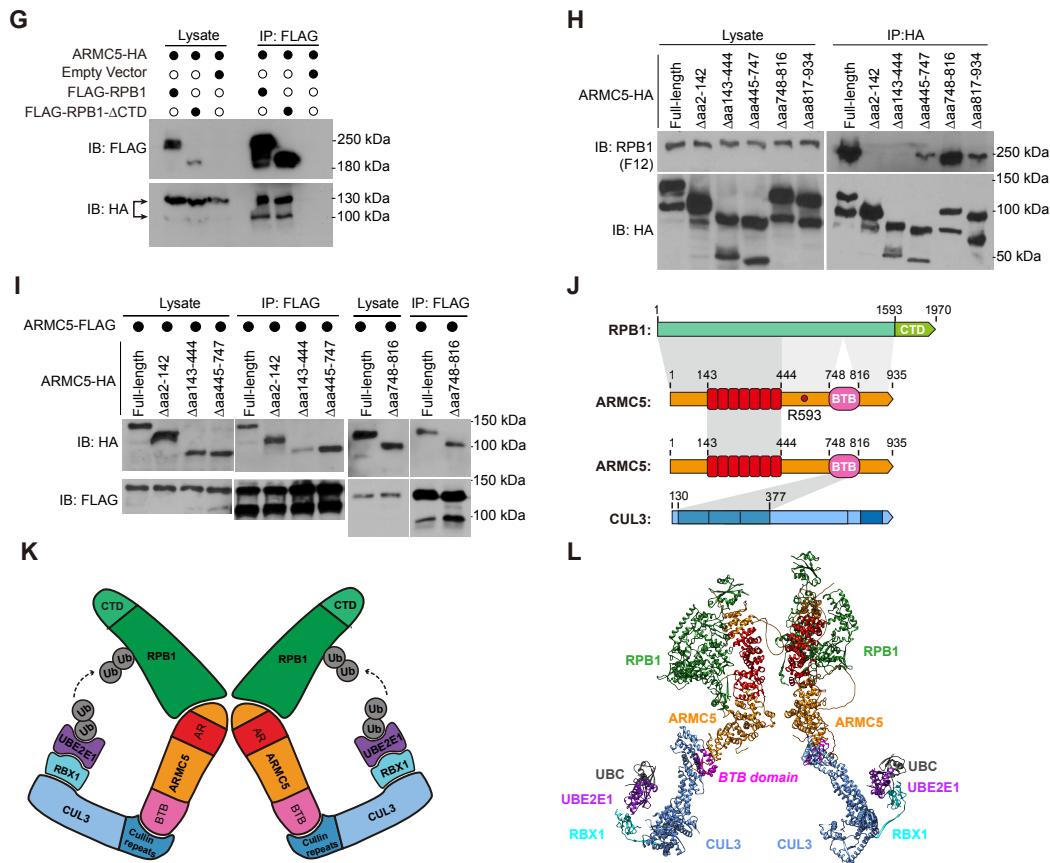
Fig. 3.2. (Continued.) Identification of the regions of interaction in ARMC5, CUL3, and RPB1 molecules *(see next page)*

Fig. 3.2 *(previous page).* **g**. The CTD of RPB1 was not essential for the association between RPB1 and ARMC5. HEK293 cells were transfected plasmids expressing ARMC5-HA, and full-length RPB1 (FLAG-RPB1) or RPB1 with its CTD deleted (FLAG-RPB1-ΔCTD). The lysates were precipitated with anti-FLAG Ab and blotted with anti-HA or anti-FLAG Abs. **h**. The N-terminal sequence (aa2 – 142) before the ARM domain (aa143 – 444) and the ARM domain of ARMC5 were both needed for RPB1 binding. The sequence after the ARM domain and before the BTB domain and the sequence after the BTB domain also contributed to RPB1 binding, but to a lesser degree. HEK293 cells were transfected plasmids expressing HA-tagged ARMC5 deletion mutants as described in the inset table. The lysates were precipitated with anti-HA Ab and blotted with anti-RPB1 (clone F12) or anti-HA Abs. **i**. ARMC5 interacted with ARMC5 through their ARM domains (positions aa143 – 444). FLAG-tagged ARMC5 was transfected into HEK293 cells along with HA-tagged full-length ARMC5 or deletion mutants. The lysates were immunoprecipitated with anti-FLAG Ab and blotted with anti-HA or anti-FLAG Abs. **j**. A schematic showing the regions that contributed to the interaction among RPB1, ARMC5, and CUL3, with the gray shades between the molecules representing the interaction regions. The lighter shade between RPB1 and ARMC5 indicates a lesser contribution of the regions to the association between these two molecules. The position of R593, which is mutated in Adelaide PBMAH patients, is labeled. **k**. A 2D schematic of the novel dimeric RPB1-specific E3. **l**. A 3D model of the novel dimeric RPB1-specific E3. In all the experiments, empty vectors were used in transfection as controls. The lysates were also immunoblotted to confirm that the transfected proteins were present. All the experiments were conducted more than three times, and representative results are shown.

transfected ARMC5-HA-expressing plasmid into human adrenal gland cortex carcinoma SW-13 cells (**Figure 3.3b**). In these cells, ARMC5 was detected both in the cytosol and nuclei in the absence of leptomycin B. However, the presence of leptomycin B enhanced ARMC5 signals in the nuclei. This indicated that ARMC5 was always present in the nucleus, but in different cell types, there were different ARMC5 shuttling dynamics between the cytosol and nucleus, resulting in different degrees of distribution of ARMC5 molecules in these two cellular compartments. Whether we could detect ARMC5 in the nuclei depends on the equilibrium between the import and export. In SW-13 cells, the import is likely faster than export so that we could detect it without an export inhibitor. In HEK293 cells, probably the export is faster than the import, so we cannot observe ARMC5 in the nucleus unless the export is blocked.

RPB1 is heavily modified by phosphorylation on its S2 and S5 residues in its C-terminal domain (CTD), which contains multiple 7-aa long repeats[362]. RPB1 with different phosphorylation statuses could be detected by different Abs[363]. Total RPB1 (detected by anti-RPB1 N-terminus mAb clone F12, (**Figure 3.3c**), RPB1 with CTD S5 phosphorylation (detected by mAb clone D9N5I, **Figure 3.3d**), RPB1 with CTD S2 phosphorylation (detected by mAb clone E1Z3G, **Figure 3.3e**), RPB1 with both high and low phosphorylation of its CTD (detected by mAb 4H8, **Figure 3.3f**), and un-phosphorylated RPB1 (detected by mAb clone 8WG16, **Figure 3.3g**) were all increased according to immunoblotting in the *Armc5* KO lymphoid organs (the spleen and lymph nodes) and adrenal glands, compared to their WT counterparts. Immunofluorescence staining showed that the nuclear RPB1 level in *Armc5* KO mouse embryonic fibroblasts (MEFs) was also augmented (**Figure 3.3h**). Such accumulation of RPB1 with different CTD phosphorylation was also observed in all other mouse OK organs tested (i.e., the thymus, liver, kidney, lung, brain, heart, stomach, colon, and small intestine) (**Figure 3.S.1**).

We fractionated the nuclei and cytosolic RPB1 of the KO and WT adrenal glands. The RPB1 levels in both fractions from the KO tissue were elevated (**Figure 3.3i**), suggesting that this E3 is active in both these cellular compartments.

ARMC5 germline mutations predispose patients to PBMAH. We assessed the RPB1 expression in the adrenal glands of PBMAH patients of two cohorts, one from Adelaide, Australia, and one from Montreal, Canada. The *ARMC5* mutations of patients in these cohorts are depicted in **Figure 3.3j**. In the Adelaide cohort, the 3 PBMAH patients were siblings, and all carried the same missense germline C→T point mutation at Chr16:g.31476121, resulting in an R593W mutation in the ARMC5 protein sequence. The R593W mutation was at a region between the ARM domain and BTB domain. Five adrenal gland adenomas and two normal adrenal glands that were not known to have *ARMC5* mutations were used as controls.

In the Montreal cohort, two PBMAH patients were a father and a daughter, both carrying a heterozygous germline variant in the *ARMC5* gene c.327_328insC (p.A110Rfs*9). This

mutation caused a frameshift starting from the N-terminal region before the ARM domain and resulted in an early truncation and a lack of functional ARMC5 protein (**Figure 3.3j**). Another PBMAH patient, E191, from the cohort harbored a heterozygous germline *ARMC5* deletion of exons 5 – 8. This deletion started in the middle of the ARM domain and resulted in a truncation of all downstream *ARMC5* sequences and hence a lack of functional ARMC5. Adrenal gland adenomas from three patients were also used as controls, which were not known to have *ARMC5* mutations. Two PBMAH samples without *ARMC5* mutations were used as additional controls.

The clinical details of the patients are described in **Table 3.1** and **Table 3.2**. Additional clinical and laboratory findings of these patients could be found in our previous publications[2,364,3,365]. RPB1 protein levels in the resected PBMAH nodular adrenal tissues with *ARMC5* mutations of the Adelaide cohort (**Figure 3.3k**) and Montreal cohort (**Figure 3.3l**) were all highly elevated, compared to those in the adrenal adenomas and normal adrenal glands. More interestingly, only the PBMAH samples with *ARMC5* mutation presented elevated RPB1 levels, while two PBMAH samples without *ARMC5* mutation showed RPB1 levels similar to control adenomas (**Figure 3.3m**). This clearly indicated that RPB1 accumulation was not a general feature for all PBMAH, and *ARMC5* mutation was the cause. *RPB1* mRNA expression was similar between PBMAH tissues and controls (adenomas and normal adrenal glands) (**Figure 3.3n**), indicating that the upregulation of the RPB1 protein occurred at the post-transcriptional level, likely due to compromised degradation since we found that ARMC5 was the substrate recognition subunit of the novel RPB1-specific E3. We assessed how the *ARMC5* mutations found in the patients affected ARMC5's binding to RPB1. Compared to WT ARMC5, ARMC5 with R593W mutation, which was found in the Adelaide PBMAH cohort, presented significantly reduced association with the endogenous RPB1 in HEK293 cells (**Figure 3.3o**). It is to be mentioned that according to our deletion studies, R593 was located in a region that contributed to the association between RPB1 and ARMC5 (**Figure 3.2i**). Thus, this mutation led to a functional consequence related to the interaction between ARMC5 and for RPB1. These

results demonstrated the relevance of our *in vitro* and *in vivo* findings in the KO mice to human pathophysiology and indicated that the novel E3 was indeed essential for maintaining RPB1 homeostasis, hence Pol II homeostasis, in humans in the absence of artificially induced massive DNA damage.
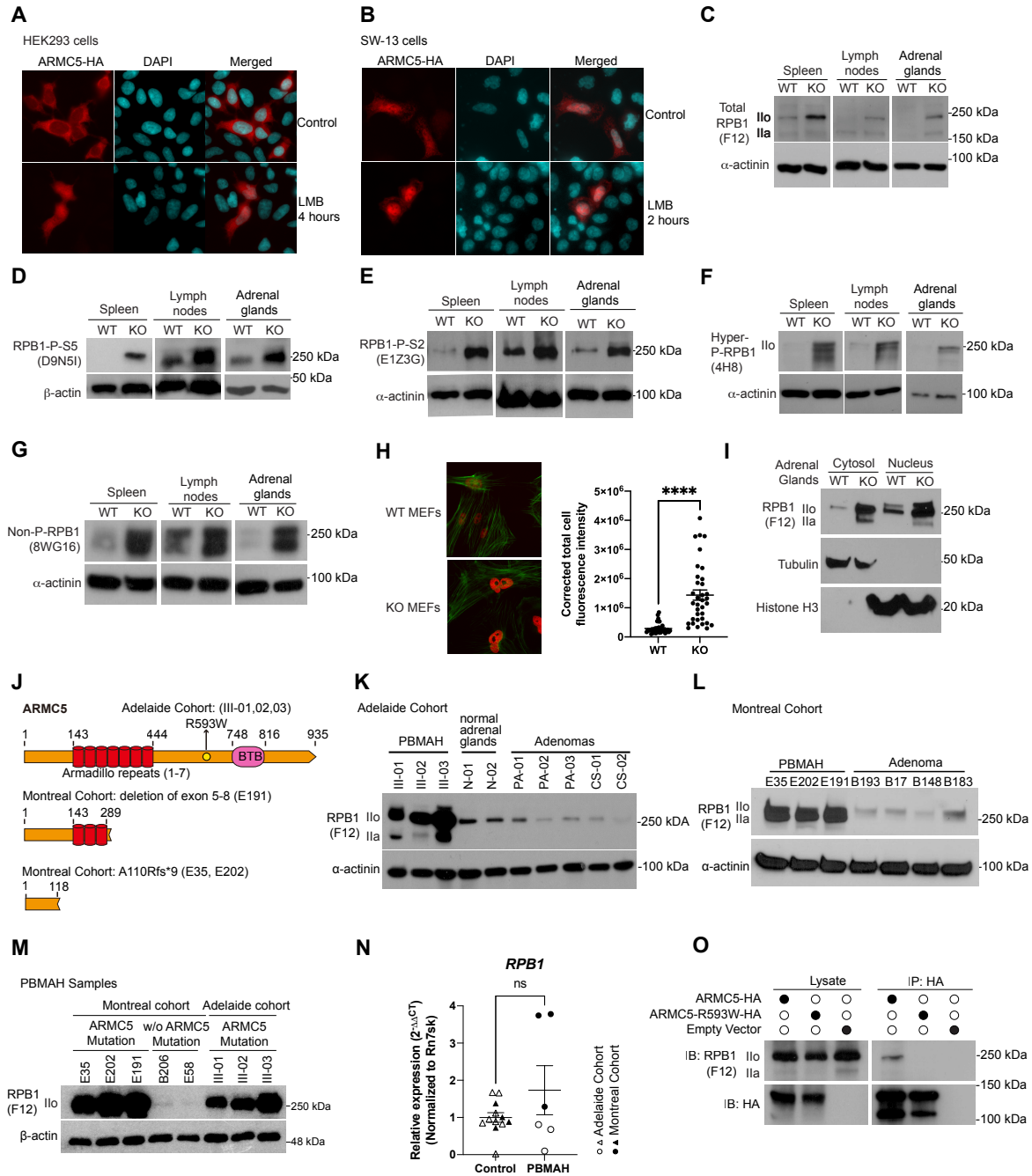
Fig. 3.3. ARMC5 KO or mutation led to RPB1 accumulation *(see next page)*

Fig. 3.3 *(previous page)*. **a** and **b**. ARMC5 was presented in both cytosol and nuclei of HEK293 cells (**a**) and human adrenal carcinoma SW-13 cells (**b**). Both types of cells were transfected with plasmids expressing ARMC5-HA. The cells were harvested after 36 hours and were stained with anti-HA Ab (pseudo-red) and DAPI (pseudo-cyan). In some cultures, nuclear export inhibitor leptomycin B (LMB; 20 nM) was present for the last 2 or 4 hours of culture, as indicated. **c-g**. Accumulation of RPB1 in KO tissues. The spleen, lymph node, and adrenal gland protein of KO and WT mice were assessed by immunoblotting for total RPB1 (**c**; mAb clone F12 against the N-terminal sequence of RPB1), RPB1 with phosphorylated S5 in CTD (**d**; mAb clone D9N5I), RPB1 with phosphorylated S2 in CTD (**e**; mAb clone E1Z3G), hyper- and hypo-phosphorylated RPB1 (**f**; mAb clone 4H8), and non-phosphorylated RPB1 (**g**: mAb clone 8WG16). β-actin or α-actinin was blotted as a loading control. **h**. Elevated RPB1 protein (Red) expression in the nuclei of KO MEFs, according to immunofluorescence using rabbit anti-RPB1 Ab (D8L4Y) followed by Alexa Fluor™ 555 goat anti-rabbit IgG. Filamentous actin (Green) was stained with Alexa Fluor™ 488 Phalloidin. Representative micrographs are shown on the left. A bar graph on the right shows the means $\pm$ SD of corrected total cell fluorescent intensity (CTCF), which is derived from RPB1 signals of 35 WT and 35 KO MEFs from three independent experiments. ****: $p < 0.0001$ (unpaired two-way Student's *t*-test). **i**. Augmented RPB1 levels in both the cytosolic and nuclei fraction of KO adrenal glands. Cytosolic tubulin and nuclear histone H3 were used as fraction purity and loading controls. **j**. Schematics of ARMC5 mutations in the Adelaide and Montreal cohorts. Patient ID numbers are indicated in the parentheses. **k** and **l**. Elevated RPB1 protein expression in the adrenal gland macronodules from Adelaide (**k**) and Montreal (**l**) PBMAH cohorts with germline ARMC5 mutations. Adrenal adenomas or normal adrenal glands were employed as controls as indicated. Immunoblotting was performed using mAb (clone F12) against total RPB1 protein. **m**. Normal RPB1 protein levels in PBMAH adrenal glands without *ARMC5* mutations. The RPB1 protein level of adrenal gland macronodules from two PBMAH patients without *ARMC5* mutation (Montreal cohort) were compared to that of six adrenal gland macronodules from PBMAH patients with germline *ARMC5* mutations (Montreal and Adelaide cohorts). **n**. *RPB1* mRNA levels of the Adelaide and Montreal PBMAH adrenal gland samples with *ARMC5* mutations were similar to those of the controls (adrenal gland adenomas and normal adrenal glands). **o**. The ARMC5 R539W mutation found in the Adelaide cohort resulted in its reduced RPB1 association. HEK293 cells were transfected with plasmids expressing WT ARMC5-HA or ARMC5-R539W-HA. Their association with endogenous RPB1 was detected by immunoblotting and immunoprecipitation.

3.3.4. ARMC5-CUL3 was an RPB1-specific E3 according to *in vivo* and *in vitro* ubiquitination

The increased RPB1 protein levels in *Armc5* KO tissues raised the possibility that ARMC5-CUL3-RBX1 was an E3 responsible for RPB1 ubiquitination, which is a necessary step to channel RPB1 to the proteasome for degradation. We analyzed the ubiquitination of the endogenous RPB1 in the KO spleen and lymph nodes (**Figure 3.4a**) and MEFs

(**Figure 3.4b**). Although the RPB1 protein levels in these KO tissues were all increased, their K48-linked RPB1 ubiquitination was reduced, indicating that ARMC5 was essential for such RPB1 ubiquitination. When WT MEFs were cultured in the presence of a proteasome inhibitor MG132, their RPB1 ubiquitination was drastically augmented (**Figure 3.4b**), suggesting that ubiquitinated RPB1 was usually channeled to the proteasome for degradation. The RPB1 ubiquitination in the KO MEFs was only marginally increased in the presence of MG132, compared to that without the inhibitor, suggesting that in the absence of this putative RPB1-specific E3, RPB1 ubiquitination was very limited, even after the degradation blockage. The slight increase of RPB1 ubiquitination KO MEFs in the presence of MG132 suggested the existence of other RPB1-specific E3(s), which could ubiquitinate RPB1 but to a much lesser extent.

The RPB1 level in the WT adrenal glands was extremely low. This made the detection of RPB1 ubiquitination in this WT tissue impossible. To overcome this technical problem, we designed a strategy by using 5-fold more input protein of the WT tissue than the KO tissue during the immunoprecipitation, along with a limited amount of anti-RPB1 Ab. This approach allowed us to compare the ubiquitination of a similar amount of precipitated RPB1 protein in the WT and KO tissues. This method's validity was first confirmed in lymph nodes because the WT lymph nodes had a reasonable RPB1 signal to be detected for ubiquitination without the equal molar comparison (**Figure 3.4a**). The result revealed that on an equal molar RPB1 basis, RPB1 from the KO lymph nodes had drastically lower total ubiquitination as well as K48-linked ubiquitination (**Figure 3.4c**), and this result was compatible with that of the equal protein input method. This equal molar RPB1 input method was then applied to the adrenal glands. Significantly reduced total and K48-linked ubiquitination of RPB1 in the KO adrenal glands were revealed (**Figure 3.4d**).

The gold standard to prove E3 activity is the *in vitro* ubiquitination assay, in which a substrate is ubiquitinated *in vitro* by a reconstituted ubiquitination enzyme cascade of E1, E3, and E3. We profiled 10 E2s for this putative new E3. UBE2E1 was found to be the optimal one (data not shown). The reconstituted ubiquitination enzymatic cascade comprising

ARMC5, CUL3, RBX1, E1, E2 (UBE2E1), ATP, and WT ubiquitin effectively ubiquitinated RPB1 (**Figure 3.4e**). A mutant ARMC5 with BTB domain deletion (ARMC5-ΔBTB), which rendered the mutant incapable of binding CUL3, failed to ubiquitinate RPB1 in this system. The faint smears of ubiquitin signals in control lanes were due to the pre-existing ubiquitination of recombination FLAG-RPB1, as such smears were not present in the lane where FLAG-RPB1 was omitted. This result proved that ARMC5 was part of a novel RPB1-specific multi-subunit RING-finger E3, and ARMC5 depended on its BTB domain to interact with CUL3 to form a functional E3 complex.

3.3.5. *Armc5* KO resulted in increased transcription of a large number of genes in the adrenal glands

We conducted RNA sequencing (RNA-seq) of the WT and KO adrenal glands. The readcounts were normalized against *Rn7sk* RNA, which was transcribed by Pol III and was not subjected to a putative general influence by abnormal levels of Pol II. Indeed, *Rn7sk* levels in KO and WT adrenal samples had no significant difference (**Figure 3.S.2**). A threshold for gene-level significance of $< 5\%$ FDR was applied to the paired comparison of RNA-seq results from 3 KO and 3 WT adrenal glands. After filtering out nominal genes that were generated by GenPipe but were not presented in the mouse reference genome (GRCm38 release 97), we obtained 1,486 genes with significantly different expressions between KO and WT adrenal glands. These genes were listed in **Table 3.S.2**, along with their FDRs, fold changes, and raw readcounts. Eighty genes in this list with the lowest FDRs are shown in a heatmap (**Figure 3.5a**), in which the color of each gene represents the SD beyond the mean expression of the gene in all the samples tested (i.e., 3 WT and 3 KO adrenal glands). A volcano plot illustrates the fold change and FDR of these 1,486 genes, with several prominently changed ones annotated (**Figure 3.5b**). *Armc5* was among the downregulated genes, as expected (**Figure 3.5b** and **Figure 3.S.3a**). It still had some signal due to transcripts appearing in the undeleted gene body (**Figure 3.S.3b**), although these transcripts would not produce any functional proteins due to frameshifts or early stops. From the volcano
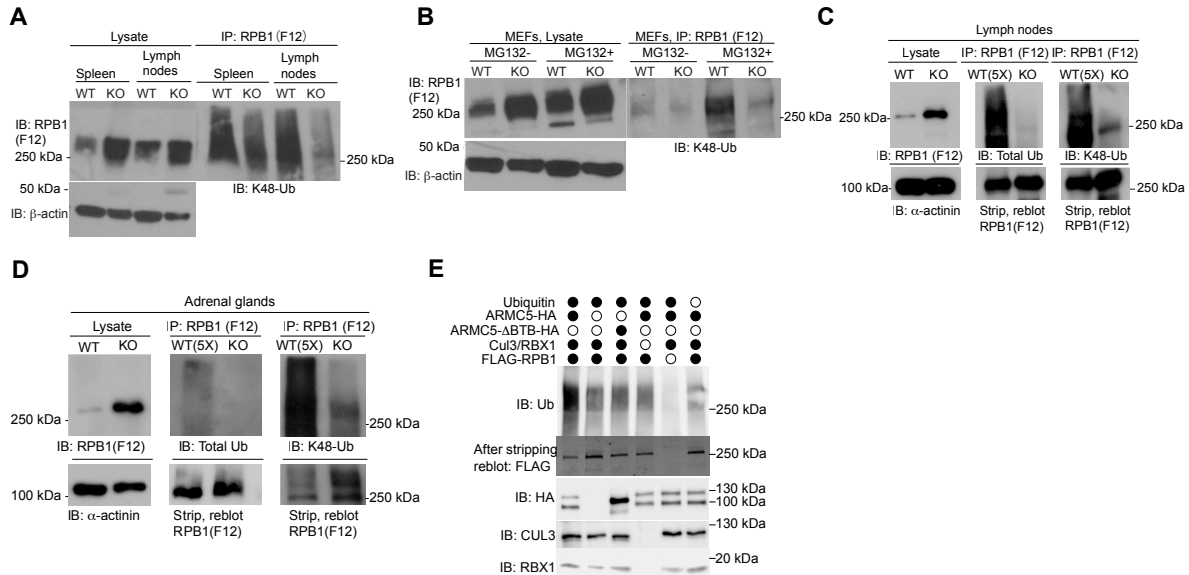
Fig. 3.4. ARMC5-CUL3-RBX1 as an RPB1-specific E3 based on in vivo and in vitro ubiquitination *a and b*. Reduced K48-linked RPB1 in the KO spleen and lymph nodes (**a**) and MEFs (**b**). MEFs were cultured in the absence or presence of proteasome inhibitor MG132 (10 μM) for the last four hours. β-actin was blotted for lysate loading control. **c** and **d**. Reduced total and K48-linked RPB1 ubiquitination in KO lymph nodes (**c**) and adrenal glands (**d**). Tissue proteins were precipitated with anti-total RPB1 mAb (F12) and immunoblotted with Abs against K48-linked ubiquitin or total ubiquitin, as indicated. α-actinin was blotted for lysate loading control. In **c** and **d**, 5-fold (5×) more WT lysates than the KO counterpart were used as input for immunoprecipitation to detect the weak WT RPB1 ubiquitination signals, using a limited amount of anti-RPB1 Ab during the immunoprecipitation. A similar amount of RPB1 protein in the WT and KO precipitates was shown by immunoblotting. **e**. ARMC5-CUL3-RBX1 was a novel RPB1-specific multiple subunit RING-finger E3 according to *in vitro* ubiquitination assays. Different recombinant proteins were added to the *in vitro* ubiquitination assay in the presence of E1, E2 (UBE2E1), and ATP. The reaction product was immunoprecipitated with anti-FLAG Ab followed by magnetic protein G beads. The immunoprecipitates were blotted with anti-Ub Ab to detect RPB1 ubiquitination. The flow-through of the immunoprecipitation was blotted to confirm the presence of RPB1, ARMC5, CUL3, and RBX1 using Abs against these molecules. All the experiments were conducted three times or more, and representative results are shown.

plot, we could appreciate that there were more upregulated than downregulated genes. This was better depicted in a bar graph (**Figure 3.5c**). Among the 1,486 genes with FDR < 0.05, most of them (1389 genes; 93.5%) were upregulated. Only a small fraction (97 genes; 6.5%) was downregulated. Were the upregulated genes preferably short? The gene length

of the upregulated, downregulated, and unchanged genes showed no significant difference (**Figure 3.5d**). We arbitrarily divided the upregulated genes into short (< 30 kb), medium-sized (≥ 30 kb and < 100 kb), and long (≥ 100 kb) ones. Most upregulated genes were short. However, their percentage (50.2%) among all the upregulated genes was similar to that of short genes in the genome (55%) (**Figure 3.5e**).

PBMAH patients have inappropriately regulated and increased cortisol levels. The large mass of the nodular glands can result in Cushing's syndrome. However, the glucocorticoid biogenesis per cell in the hypertrophic gland cortex is actually reduced[366]. STAR is a rate-limiting enzyme in steroidogenesis, regulating cholesterol transfer in the mitochondria[367]. RNA-seq revealed that the *Star* mRNA level was significantly reduced in the KO adrenal gland (**Table 3.S.2**). This was confirmed by RT-qPCR (**Figure 3.5f**), and reduced STAR protein level in the KO adrenal glands was demonstrated by immunoblotting (**Figure 3.5g**). The compromised STAR expression due to *Armc5* KO likely contributes to the compromised per cell cortisol biogenesis in cells.

Among the 1,486 differentially expressed genes in the *Armc5* WT and KO adrenal gland, we selected some of them according to their functions related to tumorigenesis and subjected them to further RT-qPCR confirmation. Some results are presented in **Figure 3.S.4**. Four of the confirmed genes were known as either tumor suppressors (e.g., *Pcdh8* and *Tfcp2l1*) or oncogenes (e.g., *Mafa* and *Taf4b*), and their validation by RT-qPCR are shown in **Figure 3.5h**. The upregulation of two oncogenes at the mRNA level (e.g., *MAFA* and *TAF4b*) was also validated in the hyperplastic adrenal glands of PBMAH patient samples with *ARMC5* mutations (**Figure 3.5i**). All RT-PCR confirmed genes were marked as brown color in **Figure 3.5b**.

3.3.6. The higher Pol II density in genes of KO adrenal gland cells was not a sign of stalling and did not cause a general decrease in transcription

The accumulation of RPB1 in KO cells raised the question whether it was part of stalled Pol II due to failed degradation. We conducted an RPB1 ChIP-seq of the adrenal gland, which was analyzed along with RNA-seq data to address this question. RPB1 signals were used customarily as a surrogate marker of Pol II[263], as it is the catalytic and largest Pol II
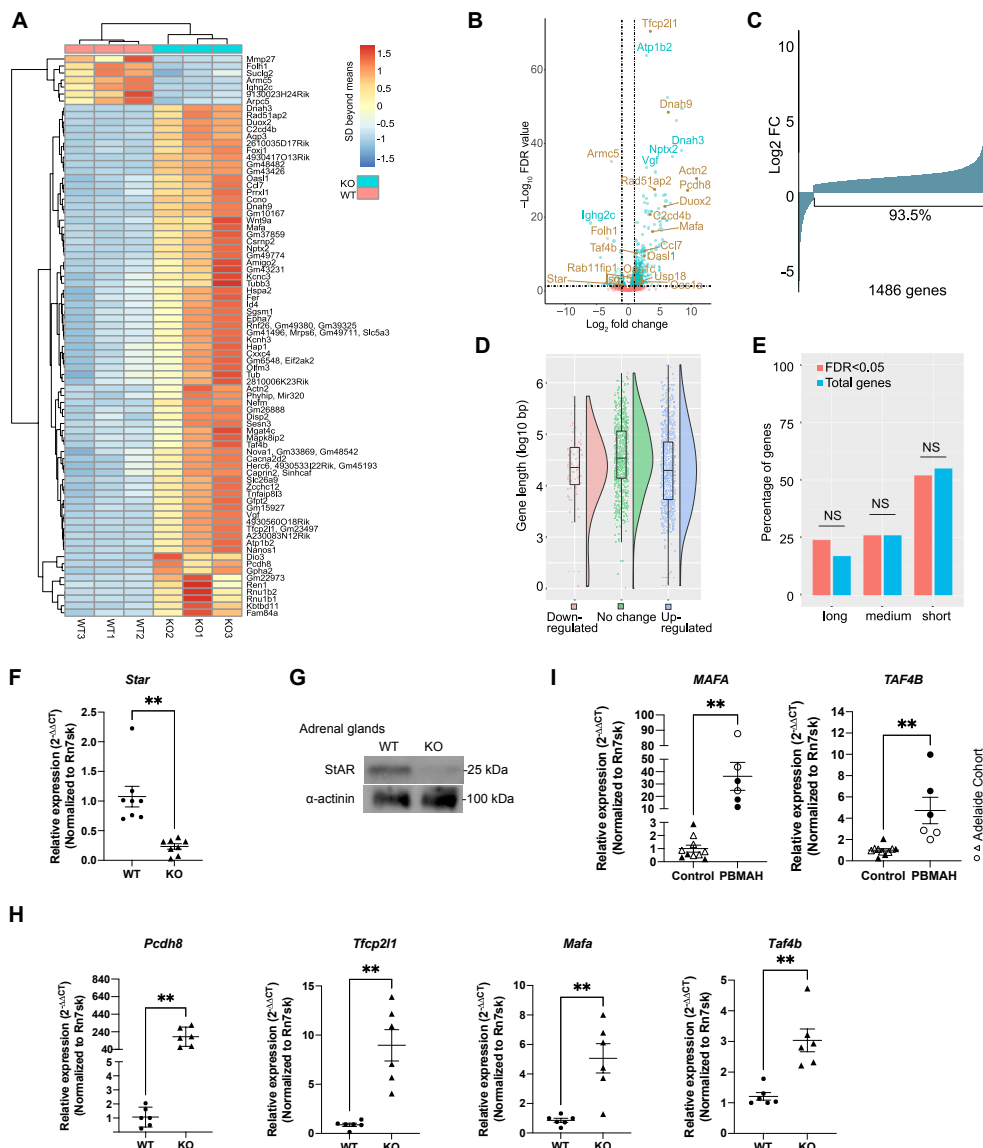
Fig. 3.5. RNA-seq of WT and KO adrenal glands *(see next page)*

Fig. 3.5 *(previous page).* **a**. A heatmap of differentially expressed genes. Eighty genes with the lowest FDR were shown. For a given gene, color represents SDs beyond the mean of raw reads in all the six samples (3 KO and 3 WT) tested. **b**. A volcano plot of FDR and expression fold changes (FC) of all the genes detected in the KO versus WT adrenal glands according to RNA-seq. Some prominently changed ones are annotated, and dysregulation of genes marked in brown was confirmed by RT-qPCR. Dashed horizontal and vertical lines represent FDR = 0.05 and $\log_2$FC = ±1, respectively. **c**. Fold changes of expression for 1486 genes with FDR < 0.05 in the KO versus WT adrenal glands. **d**. The length distribution of the genes with significant upregulation or downregulation or without change. The box graphs show the medium (solid horizontal line in the box), the 75th percentile (upper part of the box), 25th percentile (lower part of the box), 95th percentile (upper whisker), 5th percentile (lower whisker), and outliners (dots beyond the 95th and 5th percentile) of the gene length of each group. The violin plots illustrate the gene size distribution of each group. **e**. The length of the upregulated genes was similar to those in the whole genome. The percentages of short (< 30 kb), medium-sized ($\geq$ 30 kb and < 100 kb), and long ($\geq$ 100 kb) genes of the significantly upregulated genes (FDR < 0.05) and the percentages of genes of these sizes in all the genes in the genome are shown. No significant difference was observed in any length category (p > 0.05; χ^2 test). **f**. Reduced *Star* mRNA levels in the KO adrenal glands according to RT-qPCR. **g**. Reduced *STAR* protein levels in the KO adrenal glands according to immunoblotting. A representative blot from three repetitions is shown. **h**. Elevated mRNA levels of tumor suppressor genes (*Pcdh8* and *Tfcp2l1*) and oncogenes (*Mafa* and *Taf4b*) in the mouse KO adrenal glands, according to RT-qPCR. **i**. Elevated mRNA levels of oncogenes *MAFA* and *TAF4B* in the human PBMAH adrenal gland samples according to RT-qPCR. Genes in **h** and **i** were selected from the ones with significant upregulation in the KO adrenal glands according to RNA-seq. The signal ratios of the test genes versus *Rn7sk* were presented. Means ± SEM are shown. **: *p*< 0.01; ***: *p* < 0.001 (paired two-way Student's *t*-tests for mouse samples, and unpaired two-way Student's *t*-test for human samples).

subunit and has a jaw to bind incoming template DNA. A total of 12,718 genes showed discernable ChIP-seq signals. The distribution of Pol II peaks in different regions of genes was illustrated in **Figure 3.6a**. In both the KO and WT adrenal glands, the introns had the highest peak number, followed by intergenic regions and then the promoter regions. Within the genes, the highest normalized RPB1 readcounts (readcount per million mapped reads (CPM)) were accumulated near the transcription start site (TSS) (**Figure 3.6b**). Representative CPM heatmaps for the region from -2000 bp upstream of TSS to +2000 bp downstream of transcription ending site (TES) of all genes in one pair of WT and KO

samples are illustrated in **Figure 3.6c**. In such metagene analysis (**Figure 3.6a-c**), no visually discernable Pol II peak number or density differences were observed between the KO and WT adrenal glands. However, in a fixed region analysis in which the Pol II density of all the genes for the region spanning from -10 kb to +10 kb surrounding the TSS was measured, the KO tissue had a slightly higher Pol II density across this region according to visual inspection (**Figure 3.6d**). It is to be noted that such metagene illustration (**Figure 3.6b and d**) is not suitable for statistical analysis and is only meant for visual appreciation. The statistical analysis results are presented below.

Statistical analysis revealed that 273 genes of the 12,718 genes that had ChIP-seq signals showed a significant difference (FDR < 0.05) in Pol II density in the KO versus WT adrenal glands (95 genes (all increased) in the TSS region (from TSS -400 bp to TSS +100 bp); 179 genes (172 increased and seven decreased) in the gene body region (from TSS +100 bp to TES -100 bp (transcription ending site)); and 102 genes (94 increased and eight decreased) in the TES region (from TES -100 bp to TES +2,000 bp) (**Figure 3.6e** and **Table 3.S.3, Table 3.S.4, Table 3.S.5**). Thus, for those 273 genes with FDR < 0.05, the majority of them (261 genes) in the KO group presented increased Pol II density. A Venn diagram (left panel, **Figure 3.6f**) illustrates the overlaps of the different gene regions where the higher Pol II density of the 261 genes was located. The higher Pol II density of six genes with concomitant upregulation in the KO adrenal glands was illustrated in **Figure 3.5g**. The obviously increased Pol II density in the TSS region (all the six genes), in the gene body (all the genes), and in the TES region (*Hist1h1a, Oas1a,* and *Ccl7*) could be visually appreciated.

RPB1 ChIP-seq only provided a snapshot of the location of Pol II during the dynamic transcription process. To assess whether these Pol IIs were actively transcribing or stalled, we conducted a combined ChIP-seq and RNA-seq data analysis. In the KO adrenal glands, 69% of the 1,486 dysregulated genes, according to RNA-seq, had detectable ChIP-Seq signals. Among the 970 upregulated genes, 100 presented a significantly higher Pol II density in either TSS, gene-body, and/or TES regions (right panel, **Figure 3.5f**; **Figure 3.6e** and **Table 3.S.6, Table 3.S.7, Table 3.S.8**). Conversely, among the 261 genes with higher

Pol II density, 100 genes were upregulated. On the other hand, none of the genes with an increased Pol II density presented decreased mRNA levels (data not shown). This suggested that the increased Pol II density was associated with increased transcription in a subgroup (38%) of genes. The four confirmed upregulated genes (*Pddh8, Tfcp2l1, Mafa, and Taf4b*) related to tumorigenesis had significantly higher Pol II density in their genes (**Figure 3.5h**).

3.4. Discussion

In the present study, we discovered that ARMC5 complexed with CUL3 and RBX1, forming a novel multiple-unit RING-finger E3 specific for RPB1, the largest subunit of Pol II. This E3 was largely responsible for RPB1 ubiquitination in all tissues tested under a physiological condition. Due to compromised RPB1 ubiquitination and subsequent degradation via the proteasome pathway, *Armc5* deletion causedRPB1 accumulation in all major organs in KO mice. Similar RPB1 accumulation was also found in the hyperplastic adrenal gland tissues from PBMAH patients with *ARMC5* mutations. Such accumulation likely resulted in an increased Pol II pool size. We did not find any evidence of increased Pol II stalling or a generally decreased mRNA transcription in the KO adrenal glands. It is plausible that the enlarged Pol II pool dysregulates some effector genes, which collectively cause the phenotypes observed in KO animals and adrenal gland hyperplasia in PBMAH patients with *ARMC5* mutations.

3.4.1. ARMC5-CUL3-RBX1 was a primary RPB1-specific E3 under a physiological condition

Several RPB1-specific E3s have been reported in yeasts and mammalian cells. Rsp5, an E3 ligase in yeasts, polyubiquitinates RPB1 with an S5 hypo-phosphorylated CTD[368,369]. Rsp5 also mono-ubiquitinates Rpb1, and then a second E3 Elc1/Cul3 adds polyubiquitin K48-linked chains to Rpb1[268]. NEDD4, a mammalian homolog of yeast Rsp5, cooperates with Elongins-CUL5 and catalyzes polyubiquitination via a two-step reaction[268]. However, in the HEK293 cells, the ubiquitination of RPB1 catalyzed by NEDD4 only occurs when

they are UV irradiated[275]. CUL4-CSA has been suggested as an RPB1-specific E3 in irradiated cells[263]. Two RPB1-specific E3 for cell lines without exogenous DNA damage have

Figure 6
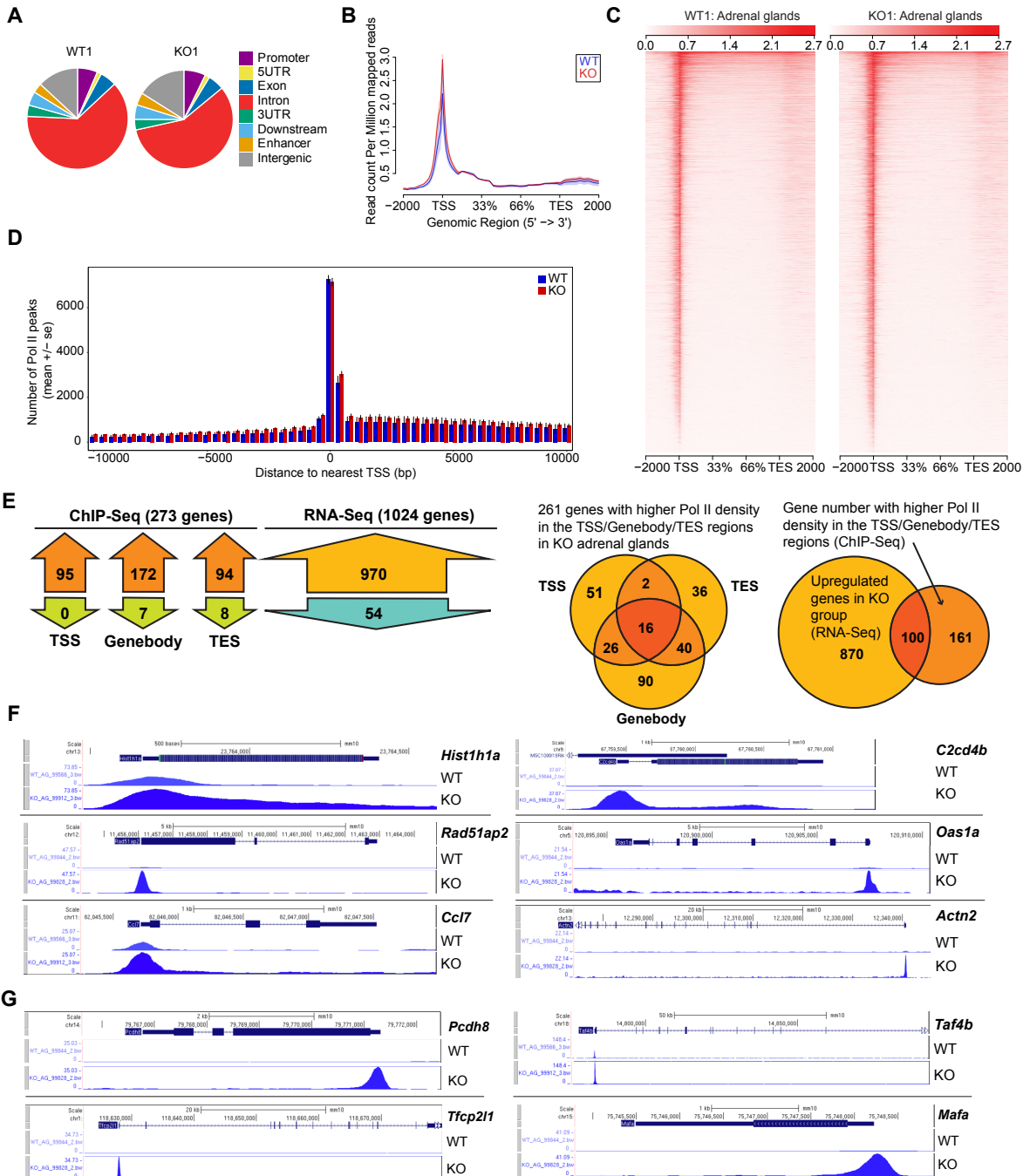
Fig. 3.6. **RPB1 ChIP-seq of WT and KO adrenal glands** *(see next page)*

Fig. 3.6 *(previous page).* **a**. Pol II peak distribution in different gene regions in a representative pair of KO and WT samples. **b**. Means (solid lines) ± SE (shadows) of normalized readcounts (readcounts per million mapped reads) in a metagene analysis for a genomic region from -2 kb of TSS to +2 kb of TES. Data were based on three biological replicates for WT and KO pairs. **c**. Heatmaps of the normalized readcounts based on data from a representative pair of KO and WT adrenal glands. **d**. Pol II peak distribution (mean ± SE) in a fixed region from -10 kb upstream to + 10 kb downstream of TSS. **e**. Combined analysis of RNA-seq and ChIP-seq data. The numbers of genes with significantly dysregulated mRNA levels and Pol II density in different gene regions (TSS, gene body, and TES) are shown in the left panel. The locations (i.e., TSS, gene body, and TES) of the heightened Pol II density of 261 genes with increased Pol II density are illustrated in the middle panel. The overlap of the upregulated genes and genes with increased Pol II density is depicted in the right panel. **f**. ChIP-seq readcount tracks in the gene regions of six genes, with a prominent difference of Pol II density in KO versus WT adrenal glands. **g**. ChIP-seq readcount tracks showing increased Pol II density in the genes with upregulated mRNA. The tracks were normalized so that each value was proportional to the readcount per base pair per 10 million reads.

been reported. pVHL-EloB/EloC-CUL2-RBX1 is specific for RPB1 with hyperphosphorylated CTD, and its activity is found in unperturbed PC12 cells based on pVHL anti-sense knockdown[279]. WWP2, a HECT E3, ubiquitinates the RPB1 CTD domain in the absence of exogenous DNA damage[194]. However, this is only proved *in vitro* in F9 embryonic carcinoma cells, in which *WWP2* siRNA knockdown leads to increased total RPB1 as well as RPB1 containing S2 or S5 hyperphosphorylation of the CTD. The activity of these RPB1-specific E3s highly depends on RPB1 CTD phosphorylation. WWP2[194,285,286] KO mice have been generated, but there is no report on the possible RPB1 accumulation in the organs of these mice. Either the authors have not paid attention to the RPB1 levels, or there is no failed RPB1 degradation under a physiological condition in these KO mice. Thus, to the best of our knowledge, there is no report of an E3 acting on RPB1 in tissues or organs without exogenously induced DNA damage. In theory, such an E3 should exist, but has eluded us until now.

The novel ARMC5-CUL3-RBX1 E3 we discovered in this work represented a major and constitutive RPB1-specific E3 under a physiological condition when its function was compromised by ARMC5 deletion or mutation, there was a considerable accumulation of RPB1

protein in most tissues and organs without exogeneously induced DNA damage or stress. The phosphorylation of S2 and S5 in the RPB1 C-terminal domain repeats is related to the location of RPB1 in the genes[370]. In the absence of ARMC5-CUL3-RBX1 E3, the degradation of RPB1 with hyper- or hypo-phosphorylated S2 or S5 in the CTD, or unphosphorylated RPB1 was all compromised (**Figure 3.3**), resulting in significant accumulation of all these RPB1 species. As RPB1 constantly transforms among these different phosphorylation statuses, we have not determined whether the E3 targets one, a few, or all forms of RPB1. It is interesting to note that even in organs where ARMC5 expression is low, such as the liver and heart[8], the absence of this E3 still caused drastically increased RPB1 levels. Thus, this E3 likely plays a significant role in maintaining Pol II pool size homeostasis. We found that the cytosolic RPB1 level in the KO tissue was also increased (**Figure 3.3i**), suggesting that this E3 was also involved in degrading misassembled Pol II or misfolded RPB1 under a physiological condition. Due to the vital roles of RPB1 and Pol II play in cell biology, other RPB1-specific E3s might serve the much-needed redundancy to allow cells to survive when this major ARMC5-CUL3-RBX1 E3 is dysfunctional or inadequate. However, they cannot fully compensate for the dysfunction of this E3, as evidenced by the accumulation of RPB1 in the KO organs and tissues.

K48-linked ubiquitination ushers proteins to the proteasome for degradation[111]. The KO lymphoid organs and adrenal glands showed reduced K48-linked RPB1 ubiquitination (**Figure 3.4a − d**). These findings were compatible with augmented RPB1 levels in these tissues.

It is to be noted that using tagged ARMC5 overexpression in HEK293 cells or SW-13 is a very useful approach to detect the interaction of ARMC5 with other proteins, identify the interaction regions of ARMC5 with other proteins, or even its role in ubiquitinate RPB1. However, this system is not very useful to assess ARMC5's function on RPB1 levels, which should, in theory, decrease. Such a decrease was not observed in SW-13 cells (**Figure 3.1g**) and was only occasionally found in HEK293 cells at a very moderate degree (**Figure 3.1d and f**). This is mainly due to low plasmid transfection efficiency. In transfected SW-13 cells,

tagged ARMC5 could only be found in about 3% of the cells. Therefore, 97% of the RPB1 detected in the lysates was from cells without ARMC5 overexpression, hence no changes of the RPB1 level (**Figure 3.1g**). The transfection efficiency in HEK293 cells was better, at about 20%. The 80% background noise explains why only a slight decrease of RPB1 levels in HEK293 cells overexpressing ARMC5 could occasionally be noticed.

3.4.2. The structure of ARMC5-CUL3-RBX1 E3 and its accessory molecules

Through a series of deletion mutations, we established that ARMC5 utilized its BTB domain to interact with the cullin repeats of CUL3. The ARM domain and the preceding N-terminal sequences were essential for ARMC5's association with the N-terminal region before the CTD of RPB1. However, the remaining sequences after the ARM domain except the BTB domain in ARMC5 also contributed to RPB1 binding, but to a lesser extent. The critical component, a RING-finger protein RBX1, of this multi-unit RING-finger E3, is well known to bind to CUL3[371,372]. These four molecules formed a basic ensemble. We demonstrated previously by Y2H that ARMC5 interacted with another ARMC5 molecule. We showed that the ARM domain in ARMC5 was critical for its dimerization, according to the deletion study. It is possible that this basic 4-molecule ensemble exists and is dimeric as illustrated in **Figure 3.2k**, or even polymeric. A 3D model of the complex was constructed (**Figure 3.3l**) based on information extracted from Protein Database and computer modeling. Hopefully, this 3D model will be confirmed and detailed in the future by structure biology studies, especially in the context of this E3 in association with Pol II. We will then have a better understanding of how this dimeric E3 works and a better answer to the following questions. 1) Does this dimeric (or polymeric E3) attack only one RPB1 or target different RPB1s in the vicinity simultaneously? 2) Even more intriguingly, will this dimeric or polymeric E3 simultaneously attack other subunits of the Pol II or even other components of the transcription machinery, such as those in the pre-initiation complex?

There should be other obligatory components attached to or near this E3 complex for its E3 enzymatic activity, such as E1, E2, and ubiquitin. We indeed found E1 (UBE1) and ubiquitin (UbC) in the ARMC5 immunoprecipitates according to LC-MS/MS analysis (**Figure 3.1b**). No E2 was found in the ARMC5 immunoprecipitates, but this was not unexpected, as they might not have survived the immunoprecipitation due to lower affinity to the E3 complex.

In this E3 complex, RBX1 is the catalytic subunit, and CUL3 is the scaffold to bridge the catalytic subunit to the substrate recognition subunit. RBX1, in the presence of E1, E2, ubiquitin, and ATP, is likely capable of ubiquitinating anything coming to its vicinity, including CUL3 and RBX1 themselves. Indeed, CUL3-RBX1 auto-ubiquitination has been documented[373]. The function of ARMC5 as a substrate recognition subunit is probably to pull a substrate (e.g., RPB1) close enough to the catalytic subunit and let it stay there long enough to be ubiquitinated. Since ARMC5 is also near the RBX1 catalytic subunit, it should be ubiquitinated by this E3 as well. Such ARMC5 ubiquitination was recently reported by Cavalcante et al.[374]. While technically, ARMC5 could be considered as a substrate of RBX1-CUL3, a more accurate characterization of such ARMC5 ubiquitination is auto-ubiquitination of this E3 complex on its own components. Customarily, we do not consider an enzyme itself as an actual substrate even if it acts on itself. Therefore, RPB1, but not ARMC5 is a true substrate of this novel ARMC5-CUL3-RBX1 E3.

Each E3 often has more than one substrate[375]. Since the ARM domain in an ARM-containing protein can often dock different proteins[291], it follows that ARMC5 may be able to serve as a recognition subunit for several different substrates. Thus, in addition to RPB1, this novel ARMC5-CUL3-RBX1 E3 might have other substrates awaiting to be discovered. A good starting point is those hits found in our LC-MS/MS analysis of ARMC5 precipitates from HEK293 cells. A total of 164 hits with more than 2-fold presence than that in the controls. Among these hits, some could be additional substrates recognized by ARMC5. Dysfunction of these additional putative substrates and RPB1 might collectively contribute to ARMC5 KO/mutation phenotype in mice and humans.

3.4.3. ARMC5 isoforms and cleavage products

ARMC5 has eight isoforms at the mRNA level[349]. Among them, six encode productive proteins. The dominant isoform *ARMC5*-201 encodes a peptide of 935 aa in length. This isoform is expressed in most tissues, albeit at different levels. In our study, the cDNA encoding this 935-aa peptide was employed for transfection. The expressed exogenous ARMC5 always appeared as two bands at 130 kDa and 100 kDa. This size is bigger than the calculated size based on mRNA length, probably due to posttranslational modifications. The relative intensity of these two bands varied in different experiments (**Figure 3.1c, d, e and Figure 3.2d**). This suggests that the lower band is a protease degradation product rather than a peptide translated from a second ATG downstream in the 935-nt mRNA. Since the HA and FLAG tags are at the C-terminus, according to size calculation, this cleavage site is near the end of the ARM domain (aa143 – 444). As expected, when this domain was deleted, the 100-kDa band no longer existed (**Figure 3.2h**, the 3rd lane of the left panel). However, with the deletion of aa143 – 444 or aa445 – 747, there appeared a new band of about 50 kDa in size, probably due to the exposure of a new protease cleavage site (**Figure 3.2h**, 3^{rd,} and 4th lanes). These results suggest that in the cells, a fraction of ARMC5 (30 – 50%) was cleaved into two parts. The cleaved N-terminal part contained the RPB1-binding site and the C-terminal part, the CUL3-binding BTB domain. These fragments can no longer serve as an E3 substrate recognition unit, as they can not bring RPB1 close to CUL3-RBX1. Instead, they might act as negative competitors to interfere with the interaction of the full-size ARMC5 with RPB1 and CUL3.

3.4.4. Decreased RPB1 degradation did not cause Pol II stalling

Transcription by Pol II is a dynamic process. It can frequently pause due to damage of the template DNA, defective Pol II assembly, or stress in growth conditions[376]. In quiescent cells at the G_0 phase, some Pol IIs are paused at promoter-proximal regions of many genes, and they need to be nudged along to increase the transcription rate to allow cells to enter the G_1 phase[377]. The paused Pol II will continue its journey once DNA damage is repaired, the

quiescent cell status is changed, or the stress is relieved[264,263,343]. If the pausing persists, for the transcription to start or to resume, it is believed that the Pol II has to be destroyed by the proteasomes after being ubiquitinated[340,341,342,264,263,343,338,337]. If the major machinery for Pol II degradation was dysfunctional, as was the case in *Armc5* KO or *ARMC5* mutations, we expected to see extensive Pol II stalling, and consequently, a general decrease of mRNA transcription. However, to our surprise, despite the significant accumulation of RPB1 due to its decreased degradation in the KO adrenal glands, we did not observe a generally reduced mRNA transcription according to RNA-seq.

Surprisingly, based on ChIP-seq, we did not find a generalized Pol II density increase in the genes of the KO adrenal glands. Among 12,718 genes with detectable ChIP-seq signals, only 261 genes in the KO tissue had a significantly higher Pol II density. A caveat of this ChIP-seq observation is that with our experimental setup, probably only genes with drastically changed Pol II density are detected. With a better normalization approach, we might be able to detect more genes with increased Pol II density. However, it remains to be true that there was no general transcription-hampering Pol II stalling in the KO tissue, as no gene with higher Pol II density presented lower mRNA levels, and no generally decreased transcription was observed.

There are two possible non-competing explanations for such an unexpected result. Most of our knowledge related to Pol II stalling and the role of the proteasome system to resolve such stalling was derived from experiments where heavy DNA damage was induced[376] or from experiments conducted *in vitro*[378,379]. Maybe the ubiquitination/proteasome system is not required at all for removing the stalled Pol II *in vivo*. This is the case in yeasts[380,381]. Recently, two groups reported that K1268 ubiquitination is responsible for UV-induced RPB1 degradation by the proteasome[263,264]. In HEK293 cells with RPB1 K1628R mutation, in spite of failed RPB1 degradation after UV irradiation, Pol II still comes off from the damaged DNA sites and is recycled rather normally[263]. This is consistent with our finding that there was no apparent Pol II stalling in the absence of this major RPB1-specific ARMC5-CUL3-RBX1 E3. The primary function of this novel E3 is probably to maintain the homeostasis

of the Pol II pool size and to degrade misfolded or misassembled RPB1. Another possible explanation is that under a physiological condition, Pol II stalling is an insignificant event. Hence, the task of removing the stalled Pol II, although being vital, is light. E3 redundancy is in place, as manifested by the existence of several other RPB1-specific E3s[268,194,279,343,194,263]. Although most of the other E3s play a minor role without massive DNA damage, in the absence of ARMC5, they are probably sufficient to carry out the light-duty of removing the stalled Pol II.

Another surprising observation is that the accumulation of RPB1 was correlated to significantly increased mRNA levels of a large number of genes (1,389 genes) in the adrenal glands, while only a small number of genes (97 genes) had reduced mRNA levels (**Figure 3.5c**). Such a heavily skewed expression pattern was rarely seen in RNA-seq datasets with other gene deletions or mutations. Although the steady-state mRNA level is determined by the balance of mRNA transcription and degradation, in most cases, it reflects the rate of mRNA transcription. It is, therefore, reasonable to assume that in the KO adrenal glands, the heavily biased expression pattern towards augmentation was due to a generally increased transcription rate of a large number of genes. The increased transcription is probably due to a bigger Pol II pool in the KO cells. In supporting this hypothesis, Vidakovic et al. reported that K1268R mutation of RPB1 renders it incapable of being degraded by the ubiquitination/proteasome system in UV irradiated HEK293[264]. Consequently, there is an enlarged Pol II pool in cells after UV irradiation, resulting in increased transcription of more than 1,600 genes but decreased transcription of fewer than 400 genes. Thus, both our results and the results of Vidakovic support the notion that a larger Pol II pool favors enhanced transcription, at least for a subgroup of genes in certain tissues[264].

A larger Pol II pool does not affect all the genes, unlike we would expect intuitively. Instead, only a subgroup of genes, 1,389 genes out of 18,500 expressed genes in the adrenal glands, showed increased transcription. Vidakovic et al. reported that a large Pol II pool size preferably increases the transcription of short genes after UV irradiation[264]. Although in our study, 50.2% of the genes in the upregulated group are short ones, this percentage

was not significantly different from that of short genes in the entire genome, in which short genes are the predominant species. Therefore, gene length is not a factor for the Pol II pool size to influence transcription.

In the KO adrenal glands, among 261 genes with increased Pol II density, 100 of them (38%) upregulated at the mRNA level (**Figure 3.6e**). What is the mechanism for a larger Pol II pool to upregulate some genes? For the genes with an increase in both mRNA levels and Pol II density, they must have some special sequences in their gene to enable more active transcription due to a larger Pol II pool. A possible location of such sequences is the TSS region, where most Pol IIs reside according to ChIP-seq (**Figure 3.6b**). About 24% of human genes have a TATA-like element as the transcription start site in the region[209]. Others might have multiple noncanonical Pol II binding sites for transcription initiation[209,382]. It is possible that under a condition of excessive Pol IIs, genes with multiple transcription start sites have an increased transcription rate since they can dock multiple Pol IIs. It will be interesting to identify these Pol II pool size-sensitive noncanonical Pol II binding motifs (transcription initiation sites) in the promoter region, as such motifs could be operational in other genes of different tissues under the condition of an enlarged Pol II pool size. Of course, such augmentation of transcription will likely be subjected to another layer of tissue-specific regulation. Indeed, different types of tissues had different numbers and different sets of upregulated genes under the influence of a larger Pol II. For example, in the *Armc5* KO mouse embryonic fibroblasts, similar to the KO adrenal gland, more than 1,000 genes were upregulated, but most of them were different from those upregulated in the adrenal gland (data not shown). Another example is the upregulated genes in HEK293 cells with RPB1 K1268R mutation[264]. The 1,600 upregulated genes, in that case, were mostly different from the ones we observed in the *Armc5* KO adrenal glands.

It is more difficult to understand why a larger Pol II pool causes the downregulation of some genes. Maybe some of the downregulated genes lack other means to remove the stalled Pol II, which exists functionally but is not being detected as significant in ChIP-seq after multiple-testing correction. It is also possible that some of these downregulated genes are

indirectly influenced by the larger Pol II pool via other upregulated ones. Needless to say, there were false-positive ones in the RNA-seq and ChIP-seq results.

The biological consequence of the Pol II pool size has not attracted much attention until recently[264,263]. The increased Pol II pool size could abnormally regulate many genes, and the consequence is often detrimental, as demonstrated in the RPB1 K1268R knock-in mice[263], which have a Cockayne syndrome-like manifestation. These mice and our *Armc5* KO mice share some common pathologies such as a significant degree of prenatal lethality, defective neural development (our KO mice had a high incidence of neural tube defects; data not shown), and dwarfism. Based on their cellular studies of the effect of the enlarged Pol II pool caused by RPB K1268R mutation, Vidakovic et al. suggested that "the Pol II pool size might contribute significantly to genome instability disorders (e.g., Cockayne syndrome, Fanconi anemia, Blooms syndrome, and Huntington's disease), and perhaps even to the regulation of cell-type-specific transcription programs in normal cells."[264]. Indeed, our current work proved the biological importance of the Pol II pool size in the normal cells and whole animals/humans without irradiation.

3.4.5. ARMC5 mutations and diseases

A general phenotype of the *Armc5* KO cells was reduced proliferation. This was observed in T cells[8], MEFs, neural tube cells, and neural progenitor cells (data not shown). The generally reduced proliferation might also be a contributing factor to dwarfism. How do we reconcile this phenotype with adrenal gland hyperplasia and meningiomas found in PBMAH patients with *ARMC5* mutation? It is possible that some upregulated genes caused by the larger Pol II pool are oncogenes. However, such an oncogenic effect is counteracted by reduced cell proliferation, which is caused by another set of anti-proliferation genes (detailed examples will be given later). As a result of these two opposing effects, *ARMC5* mutations lead to very slow-growing hyperplasia (adrenal gland hyperplasia in mice and PBMAH in humans) or tumors (meningiomas in humans). Indeed, both PBMAH and meningiomas take decades to develop and are often detected after $50 - 70$ years of age[350,383].

Using two cohorts from Adelaide, Australia, and Montreal, Canada, we proved that PBMAH patients with *ARMC5* mutations had abnormally high RPB1 levels in their hyperplastic adrenal nodules. This has demonstrated that our findings of ARMC5-CUL3-RBX1 as a novel RPB1-specific E3 are valid in humans and relevant to PBMAH. The three Australian PBMAH patients were from the same family, and all had an identical missense mutation of R593W caused by a point mutation at C1777T in the *ARMC5* coding sequence. The mutation is located in a region between the ARM domain and BTB domain. According to our deletion studies (**Figure 3.2h and j**), this region contributed to the association between RPB1 and ARMC5. Based on the 3D structure of ARMC5 predicted by AlphaFold, R593 is located at a bending site of the ARMC5 molecule. We speculate that this mutation might alter the bending angle of ARMC5 and consequently cause a weaker binding between ARMC5 and RPB1. Such reduced interaction was proven *in vitro* using ARMC5-R539W mutant according to immunoprecipitation (**Figure 3.3o**). As a consequence, the E3 became less potent, resulting in inadequate RPB1 ubiquitination ensured by RPB1 accumulation. It is not the issue of why one point mutation causes such a dramatic effect. Rather, among the many possible point mutations that occur naturally in *ARMC5*, only the critical ones will render pathological manifestations, hence being identified. Some PBMAH patients of other cohorts have mutations in the ARMC5 BTB domain[17], which was critical for binding to CUL3 based on our deletion experiment. It is likely that any critical mutations affecting the interaction between the subunits of this RPB1-ARMC5-CUL3-RBX1 complex will probably result in compromised RPB1 ubiquitination, which in turn causes RPB1 accumulation and an abnormally larger Pol II pool.

The adrenal gland hyperplasia in KO mice and PBMAH patients is likely caused by increased cortical cell proliferation or reduced apoptosis, or both. We particularly validated the changes of some of the differentially upregulated expressed genes from RNA-seq by RT-qPCR.

Some confirmed upregulated genes seem to function as tumor suppressors and are anti-proliferative. For example, *Pcdh8* encodes a membrane protein belonging to the cadherin

superfamily and is a candidate tumor suppressor in breast cancer[384]. *Tfcp2l1* encodes a transcription factor that is known to maintain the pluripotency of embryonic stem cells[385]. It positively regulates CRYAB protein expression, and its downregulation is associated with thyroid carcinomas[386]. They were both highly upregulated (∼200 fold and ∼6 fold, respectively) in KO adrenal glands.

On the other hand, some confirmed upregulated genes can act as oncogenes. *Mafa* is a case in point. It is a large-*Maf* family member and has a strong transforming capability in fibroblast[387]. It is to be mentioned, though, that its oncogenic activity depends on the cell context, and under some circumstances, it counteracts the oncogenic function of the Ras/Raf/MEK pathway activation[388]. This gene was highly upregulated in KO adrenal glands (5-fold) and hugely so in human PBMAH adrenal gland nodules (40-fold). *TAF4b* also belongs to this oncogene category. Its protein is a subunit of the transcription factor IID (TFIID)[389], which is essential for the initiation of transcription by Pol II[390]. TAF4b is known to activate anti-apoptotic genes and can thus promote cell survival[391]. It was upregulated 3 and 5 folds in the KO mouse adrenal gland and human PBMAH adrenal glands, respectively.

It is possible that the collective effects of these multiple dysregulated genes but not a single gene resulted in the phenotype observed in the KO mice and patients with *ARMC5* mutations. The upregulation of tumor suppressor and anti-proliferative genes such as *Pcdh8* and *Tfcp2l1* curbs the growth of cells and tumors, while elevated levels of oncogenes such as *Mafa* and *Taf4b* trigger oncogenesis. The end results depend on the equilibrium of these two opposing forces in different tissues and often lead to slow-growing hyperplasia or tumors such as PBMAH and meningiomas. The upregulation of the anti-proliferative genes (*Pcdh8* and *Tfcp2l1* or similar ones) can also explain the generally reduced growth rate of many types of KO cells, such as lymphocytes[8], MEFs, and neural progenitor cells (data not shown).

Since *ARMC5* mutation is associated with PBMAH and meningioma risks, can it be characterized as a tumor suppressor gene? To qualify a gene as an oncogene or tumor suppressor gene, it should affect cell transformation, proliferation, or death in most cell types and have clear, immediate mechanism leading to such phenotypes. *ARMC5* does not have such

quality. The abnormally large Pol II pool size after *ARMC5* deletion/mutation might affect some real effector genes (e.g., upregulating certain oncogenes or tumor suppressor genes), causing cell type-dependent indirect effects, be it anti-proliferative or pro-proliferative, and anti-apoptosis or pro-apoptosis. Thus, despite the eventual cell growth or death caused by ARMC5 KO/mutation, it is inappropriate to categorize *ARMC5* neither as a tumor suppressor gene nor an oncogene *per se*, for the same reason that we do not consider ARMC5's downstream target *RPB1* as a tumor suppressor or oncogene even Pol II transcribes every tumor suppressor gene and oncogene.

We have been puzzled that although *ARMC5* mutation is associated with PBMAH risks, only about 25% of the PBMAH patients have *ARMC5* mutations. What causes PBMAH in the remaining 75% of patients? We found that only in PBMAH patients with *ARMC5* mutations but not those without the mutation, the RPB1 protein levels in their adrenal gland nodules were elevated. This finding confirms that *ARMC5* mutation results in RPB1 accumulation. It also clearly proves the heterogeneous nature of PBMAH pathogenesis. For the PBMAH patients with *ARMC5* mutations, the enlarged Pol II pool size is the direct cause, but the real culprits might be the downstream effector genes, some of which could be upregulated oncogenes. For those PBMAH patients without *ARMC5* mutations, they have a normal Pol II pool and function, and their PBMAH might be caused by the mutations of the same effector genes as those found downstream of Pol II, such as *Mafa* and *Taf4b* as alluded above. Thus, pin-pointing the Pol II downstream effector genes in the PBMAH patients with *ARMC5* mutations might help us to find risk genes for those PBMAH patients without *ARMC5* mutations.

Three of our coauthors previously conducted Affimatrix microarray to discover differentially expressed genes in three PBMAH versus two normal adrenal gland tissues[392]. The complete dataset has now been deposited in the Gene Expression Omnibus of NCBI (accession #GSE171558). We compared the mouse RNA-seq and human microarray datasets and found 43 genes that were commonly upregulated in the mouse KO and human PBMAH adrenal gland tissues and four genes that as commonly downregulated (**Table 3.S.9**). The

oncogene *TAF4b* was among the upregulated ones found in both studies. It will be interesting to investigate *TAF4b* and other commonly regulated genes for their roles in PBMAH pathogenesis. There are apparent profile differences between the mouse RNA-seq and human microarray data. This is not unexpected due to the following reasons. A). The sensitivity and specificity of RNA-seq and microarray are different. B). In the KO mouse tissue, all the cells have bi-allelic *Armc5* deletion, while all the PBMAH tissues carry a monoallelic R593W point mutation. C). In the mouse samples, the whole adrenal gland, including the medulla, was used. *Armc5* in all the cell types in the gland was deleted. The human samples were derived from macronodules of the adrenal gland cortex, but they also contained non-malignant cells (e.g., cells in the blood vessel and fibroblasts) without *ARMC5* mutations. In both datasets, there were false-positive and false-negative genes. A case in point is *Mafa*. This gene was significantly upregulated in mouse RNA-seq. This prompted us to assess its expression in the PBMAH samples. The RT-qPCR results showed that it was also highly upregulated (25 – 90 folds) (**Figure 3.5i**) in the same three Adelaide PBMAH samples used in the microarray, which did not identify it as significant. Obviously, there is a species difference.

ARMC5 mutations cause PBMAH accompanied by reduced cortisol biogenesis per cell, although due to the massive adrenal gland hyperplasia, the patients have varying degrees of biochemical hypercortisolism. Among the small number of downregulated genes, *Star* had particular relevance to the reduced per cell cortisol biogenesis. The protein coded by this gene is a transport protein for cholesterol within the mitochondria and is one of the rate-limiting enzymes for the biogenesis of steroid hormones, including glucocorticoids[393]. Thus, *Star* is one of the effector genes downstream of the enlarged Pol II for cortisol biogenesis. The significantly reduced *Star* mRNA and protein levels in the KO adrenal gland likely contribute to the reduced per cell cortisol production. Mechanisms by which the large Pol II pool-caused suppression of *Star* transcription need to be elucidated. Possibly it is an indirect effect mediated by some upregulated genes, which is the majority among the dysregulated ones in the adrenal glands. With that said, we cannot exclude the possibility that for this gene, some

low degree of Pol II stalling did occur, and this gene lacks a degradation-independent Pol II recycling mechanism as alluded to before, resulting in depressed transcription. Similar to those hyperplasia-related effector genes we described above, in humans, STAR dysfunction might also need a second hit either in the *ARMC5* or *STAR* gene. Probably only the germline but not somatic second-hit will be effective, as somatic mutation-caused loss-of-function in a single cell will not create a noticeable phenotype in patients.

The enlarged Pol II pool due to *ARMC5* mutation might cause other diseases, some of which are subtle and can only be revealed if carefully examined. The *Armc5* KO phenotype in mice can serve as a guide to search for the cause of such human diseases.

In summary, we discovered a novel and major RPB1-specific E3 that functioned under a physiological condition. The dysfunction of this E3 led to an enlarged Pol II pool, which rendered dysregulation of many genes and subsequent deleterious phenotypes and diseases such as adrenal gland hyperplasia in both mice and humans.

3.5. Methods

3.5.1. *Armc5* KO mice

Armc5 KO mice and their littermates used in this report were in the129/sv × CD1 background. The details about the generation of *Armc5* KO mice are described previously[8]. All animals were housed under specific pathogen-free conditions and handled in accordance with a protocol approved by the Institutional Animal Protection Committees of the CRCHUM and INRS-IAF.

3.5.2. Cell culture and transfection

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM). MEF cells were derived from E12.5 KO or WT fetuses and cultured in DMEM. SW-13 (ATCC) adrenal gland carcinoma cells were cultured in L-15 medium at 37 °C. All culture media were supplemented with 10% fetal bovine serum (FBS), penicillin (100 U/ml), and streptomycin (100 U/ml). Transfection of HEK293 cells and SW-13 cells was performed using Lipofectamine

2000 and Lipofectin (both from Invitrogen), respectively. Proteasome inhibitor MG132 (10 μM; Cayman Chemical) or nuclear export inhibitor leptomycin B (20 nM; Cell Signaling Technology) was added to cell culture as indicated. These cells were analyzed 48 h after the transfection.

3.5.3. Plasmids

Plasmid ARMC5-HA (EX-H0661-M07) expressing full-length ARMC5 (aa1 – 935) with HA at its C-terminus and plasmid ARMC5-FLAG (EX-H0661-M14) expressing full-length ARMC5 (aa1 – 935) with FLAG at its C-terminus were obtained from GeneCopeia. Plasmid CUL3-Myc (RC208066) expressing full-length human CUL3 with Myc at its C-terminus was obtained from OriGene. Plasmid CUL3-HA expressing full-length human CUL3 with HA tag at its C-terminus was cloned by restriction enzyme-based method using the insert from plasmid CUL3-Myc and the vector from plasmid ARMC5-HA. Plasmid FLAG-RPB1 (Plasmid #35175) expressing full-length human RPB1 with FLAG at its N-terminus, plasmid FLAG-RPB1-ΔCTD (Plasmid #35176) expressing CTD-deleted human RPB1 with FLAG at its N-terminus, and plasmid HA-Ubiquitin (Plasmid #18712) were obtained from Addgene. Following plasmids expressing human ARMC5 and CUL3 deletion mutants were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs): ARMC5(Δaa2 – 142)-HA with aa2 – 142 deleted, ARMC5(Δaa143 – 444)-HA with aa143 – 444 deleted, ARMC5(Δaa445 – 747)-HA with aa445 – 747 deleted, ARMC5(Δaa748 – 816)-HA with aa748 – 816 deleted, and ARMC5(Δaa817 – 934)-HA with aa817 – 934 deleted, CUL3(Δaa31 – 385)-HA with aa31 – 385 deleted, CUL3(Δaa377 – 675)-Myc with aa377 – 765 deleted, CUL3(Δaa695 – 762)-Myc with the neddylation site (aa695 – 762) deleted. Plasmids ARMC5-BTB-HA expressing the ARMC5 BTB domain (aa748 – 935) with HA at its C-terminus, and CUL3(aa1 – 376)-Myc expressing CUL3 N-terminal culling repeats (aa1 – 376) with Myc at its C-terminus were generated by retrieving the needed fragments with PCR using Q5 Hot Start High-Fidelity DNA polymerase (New England Biolabs) from full-length ARMC5 or Cul3 cDNA and re-cloned them to the original vectors.

3.5.4. Abs, Ab-conjugated beads, and recombinant proteins

These reagents and their sources were listed in **Table 3.S.11**.

3.5.5. Immunoprecipitation and Western blotting

For protein-protein interaction experiments, cells were lysed in TNE buffer (50 mM Tris-HCl at pH 7.4, 100 mM NaCl, 0.1 mM EDTA, 1% Triton X100) supplemented with Halt^{TM} Protease Inhibitor Cocktail (ThermoFisher) and Phosphatase Inhibitors (Roche). The extracts were incubated with the corresponding antibodies overnight at 4 °C and then incubated with Protein G magnetic beads (Bio-Rad) for another two hours at 4 °C. The beads were washed five times with wash buffer (50mM Tris-HCl at pH 7.4, 150 mM NaCl, 1mM EDTA, 0.2 mM Sodium Orthovanadate, 1% Triton X100). The bound proteins were eluted by a 2× SDS-loading buffer. For protein ubiquitination experiments, cells or tissues were lysed or homogenized in RIPA buffer (25 mM Tris at pH 7.6, 150 mM NaCl, 1% Nonidet P-40, 1% sodium deoxycholate, 0.1% SDS) supplemented with Halt^{TM} Protease Inhibitor Cocktail, Phosphatase Inhibitors, and 4 mM of N-ethylmaleimide (NEM) (Millipore Sigma). The cleared supernatants were also incubated with Abs overnight at 4 °C followed by protein G-conjugated magnetic beads (Bio-Rad) for another two hours at 4 °C. The beads were washed five times with RIPA buffer. The bound proteins were eluted by a 2× SDS-loading buffer and then resolved by SDS-PAGE. The proteins in the gels were transferred to PVDF membranes. The membranes were blocked with 5% (w/v) milk in TBST (Tris-Buffered Saline, 0.05% Tween 20) and incubated with first Abs for two hours at room temperature or overnight at 4 °C, followed by HRP-conjugated secondary Abs for one hour at room temperature. The signal was revealed by the Western Lightning pro^{TM} ECL (PerkinElmer) and detected with either X-ray film or ChemiDoc Imaging system (Bio-Rad).

3.5.6. Tandem liquid chromatography and mass spectrometry

HEK293 cells were transfected with ARMC5-HA or control plasmid, and after 48 hours, the cells were lysed in TNE buffer supplemented with Halt^{TM} Protease Inhibitor Cocktail

and Phosphatase Inhibitors. Lysates were centrifuged and immunoprecipitated by anti-HA Ab-conjugated agarose beads (Sigma) at 4 °C overnight. The beads were washed five times by wash buffer (50 mM Tris-HCl at pH 7.4, 150 mM NaCl, 1mM EDTA, 0.2 mM sodium orthovanadate, 1% Triton X100), and the precipitated proteins were eluted by HA peptides (GenScript). The eluates were resolved by 4 – 15% Mini-PROTEAN TGX$^{\mathrm{TM}}$ Precast Protein Gels, and the gels were stained with the silver staining kit (ThermoFisher).

Visible bands in the silver-stained gel were excised and destained in 50% MeOH (Sigma-Aldrich). Each band was shrunk in 50% acetonitrile (ACN), reconstituted in 50 mM ammonium bicarbonate with 10 mM Tris (2-carboxyethyl) phosphine hydrochloride (Thermo Fisher Scientific), and vortexed for 1 hour at 37 °C. Chloroacetamide (Sigma-Aldrich) was added for alkylation to a final concentration of 55 mM. Samples were vortexed for another hour at 37 °C. One microgram of trypsin was added, and digestion was performed for 8 hours at 37 °C. Peptide extraction was conducted with 90% ACN. The extracted peptide samples were dried and solubilized in 5% ACN-0.2% formic acid (FA). The samples were loaded on a home-made C18 pre-column (0.3-mm inside diameter × 5 mm) connected directly to the switching valve. Peptides were separated on a home-made reversed-phase column (150-μm inside diameter × 150 mm) with a 56-min gradient from 10 to 30% ACN-0.2% FA and a 600 nl/min flow rate on an Ultimate 3000 HPLC connected to a Q-Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer (MS) (Thermo Fisher Scientific). Each full MS spectrum acquired at a resolution of 70,000 was followed by 12 tandem-MS (MS-MS) spectra on the most abundant multiply charged precursor ions. Tandem-MS experiments were performed using collision-induced dissociation at a collision energy of 27%.

The data were processed using PEAKS 8.5 (Bioinformatics Solutions, Waterloo, ON) and a human database. Mass tolerances on precursor and fragment ions were 10 ppm and 0.01 Da, respectively. Variable selected posttranslational modifications were carbamidomethyl (C), oxidation (M), deamidation (NQ), and phosphorylation (STY). The data were analyzed with Scaffold 4.3.0. A protein was categorized as a hit if it met the threshold of 99%, with at least two peptides identified and a false-discovery rate (FDR) of 1% for peptides.

The hits were further filtered for data presentation, according to the criteria described in the Result section.

3.5.7. Construction of the E3 3D model

The ARMC5 3D structure was obtained from Alphafold Protein Structure Database[361]. The cullin-repeat structure of CUL3 was derived from the crystal structure of the KLHL3-CUL3 complex (Protein Database Band (PDB) 4HXI). The cullin homology domain and C-terminal domain of CUL3 conjugated with RBX1 were extracted from the crystal structure of an RBX1-UBC12~NEDD8-CUL1-DCN1 complex (PDB 4P5O), in which CUL1 was highly homologous to CUL3. The RPB1 structure was obtained from the crystal structure of the human RNA Pol II complex (PDB 6DRD). The structures of UBE2E1 and UBC were derived from the structures of the TRIM21-UBE2E1 complex (PDB 6FGA) and UbcH5A-UBC complex (PDB 4AP4), respectively. The docking of UBE2E1 on RBX1 was modeled according to the structure of TRIM21-UBE2E1 (PDB 6FGA) by replacing the RING domain of TRIM21 with RBX1. The interaction between ubiquitin UBC and UBE2E1 was based on the UbcH5A-UBC structure (PDB 4AP4) by replacing UbcH5A with UBE2E1. The interactions among RPB1, ARMC5, and RPB1 were modeled according to the results of our deletion studies by positioning the interacting domains close to each other. UCSF Chimera[394] was used to extract and position proteins in the 3D model.

3.5.8. Immunofluorescence

MEFs, ARMC5-HA transfected HEK293 cells, and SW-13 cells were grown on coverslips in 6-well plates. In some experiments, nuclear export inhibitor leptomycin (20 nM) was added to the culture for the last 2 – 4 hours of culture, as indicated. After 48 hours, the cells were fixed with 4% (w/v) paraformaldehyde for 30 minutes at room temperature and permeabilized with 0.3% Triton in PBS for 5 minutes. The cells were then blocked with 5% goat serum in PBS for 1 hour at room temperature and reacted with the corresponding first Abs overnight at 4°C. The coverslips were washed by PBS three times and incubated

with Alexa Fluor$^{\text{TM}}$ 488 goat anti-mouse Ab or Alexa Fluor$^{\text{TM}}$ 555 goat anti-rabbit Ab for 2 hours at room temperature. After three times of wash, the coverslips were mounted with ProLong Diamond$^{\text{TM}}$ Antifade Mountant containing DAPI (4,6-diamidino-2-phenylindole) (ThermoFisher Scientific). Images were acquired with a Zeiss microscope. For RPB1 signal quantification in MEFs, the corrected total cell fluorescence (CTCF) intensity was evaluated by Image J using the following formula. CTCF = Integrated density - (Area of selected cells × Mean fluorescence of background readings)

3.5.9. *In vitro* ubiquitination assay

HEK293 cells transfected with plasmids expressing ARMC5-HA or FLAG-RPB1. Forty-eight hours after transfection, the cells were lysed by RIPA buffer. The lysates were immunoprecipitated by anti-HA Ab-conjugated agarose beads or anti-FLAG M2 Ab-conjugated agarose beads. The beads were washed three times with RIPA buffer and then three times more with wash buffer (50 mM Tris-HCl at pH 7.4, 300 mM NaCl, 1mM EDTA, 0.2 mM sodium orthovanadate, 1% Triton-X100). The bound proteins were eluted by 10 μg HA peptides (GenScript) or 10 μg FLAG peptides (GenScript) in 200 μl protein preservation buffer (40 mM Tris, 110 mM NaCl, 2.2 mM KCl, 0.04% Tween 20, 30% Glycerol). ARMC5-HA protein and ARMC5-ΔBTB-HA protein eluates were concentrated by Microcon-30 kDa centrifugal filter devices (Amicon, Millipore). Flag-RPB1 protein eluates were concentrated by Microcon-100 kDa centrifugal filter devices (Amicon, Millipore). The control eluate was extracted from HEK293 cells transfected with empty vectors and underwent the same purification steps. CUL3/RBX1-GST complexes were obtained from BPS Bioscience.

Preliminary screening for an optimal E2 was performed. Ten different E2s (UBE2H, UBE2R1, UBE2D1, UBE2D2, UBE2D3, UBE2E1, UBE2E3, UBE2L3, UBE2C, and UBE2N/UBE2V1) in the UbcH Enzyme Kit (Boston Biochem) were tested in *in vitro* ubiquitination assays. UBE2E1 generated the highest signal difference between samples with and without ARMC5 and was chosen for the final *in vitro* ubiquitination assay.

For a typical *in vitro* ubiquitination reaction, ARMC5-HA, ARMC5-ΔBTB-HA, and FLAG-RPB1 (500 ng each, affinity-purified from transfected HEK293 cells) were added to a mixture containing 100 ng of E1 (UBE1), 500 ng of E2 (UBE2E1), 10 μg of His-ubiquitin (all from BostonBiochem), 260 ng of CUL3/RBX1-GST (BPS Bioscience), and 10 mM ATP in ubiquitination buffer (25 mM Tris-Cl at pH 8.0, 125 mM NaCl, 10 mM MgCl2, 50 μM DTT). The reaction was carried out at 30 °C for 90 minutes and stopped by 20 mM EDTA. The ubiquitinated RPB1 was immunoprecipitated with anti-FLAG Ab plus Protein G conjugated magnetic beads, resolved by 6% SDS-PAGE gel, and immunoblotted with anti-ubiquitin Ab.

3.5.10. RNA-seq

KO and WT adrenal glands were homogenized, and their total RNA was extracted by RNeasy kit (Qiagen). The total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Inc.), and its integrity was assessed on a 2100 Bioanalyzer (Agilent Technologies). rRNA was depleted from 250 ng of total RNA using QIAseq FastSelect (Human 96rxns; Qiagen). cDNA synthesis was achieved with the NEBNext RNA First-Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs). The remaining steps of library preparation were performed using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs), according to the manufacturer's instructions. Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the Quant-iT™ PicoGreen dsDNA Assay Kit (Life Technologies) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). The average fragment size was determined using a LabChip GX (PerkinElmer).

The libraries were normalized and pooled and then denatured in 0.05 N NaOH and neutralized using HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S2 lane using the Xp protocol as per the manufacturer's recommendations. The run was performed for 2 × 100 cycles (paired-end mode). A phiX library was used as a control and

mixed with libraries at a 1% level. Base-calling was performed with RTA v3.x Program bcl2fastq2 v2.20 was then used to demultiplex samples and generate fastq reads.

Reads were trimmed from the 3' end to have a phred score of at least 30. Illumina sequencing adapters were removed from the reads, and all reads were required to have a length of at least 32 bp. Trimming and clipping were performed using Trimmomatic[395].

Upstream processing of FastQ files was performed as described previously[396], using the ENSEMBL GRCm38 (Mus musculus) release 97 mouse reference genome sequences and annotations. Each readset was aligned using STAR[397], which creates a Binary Alignment Map file (.bam). Then, all readset BAM files from the same sample were merged into a single global BAM file using Picard.

The pairwise Pearson's correlation values of samples were calculated. The correlation controlled the general transcripts expression consistency between samples. It could check sample mix-up or errors in name assignment. Thus, samples belonging to the same design group/condition were expected to show a higher level of correlation.

Gene expression levels were quantified using StringTie. Specifically, the count matrix was extracted from StringTie output using the auxiliary script prepDE.py as provided on the StringTie website. A gene here could be a novel gene with no overlap with a known gene. This resulted in count data for StringTie-assembled gene models.

It is generally beneficial to filter out lowly expressed genes before exploratory data analysis, but importantly prior to differential expression analysis (to reduce the multiple testing burden). For this data set, we filtered out genes that do not exceed 1 CPM in at least three samples. Out of the original genes, 17,711 genes are left after filtering.

Each gene was tested for differential expression between WT and *Armc5* KO adrenal glands with EdgeR Likely Ratio Tests. Due to the concern that the augmented Pol II pool caused by *Armc5* deletion might generally affect the transcription of all the gene in the KO adrenal glands, we set normalization factors for each sample as the ratio between the $\log_2$CPM value of *Rn7sk* of the sample to the average of all $\log_2$CPM values of *Rn7sk*

across all samples. *Rn7sk* is transcribed by Pol III and is thus independent of the putative influence of the Pol II pool size. A similar level of *Rn7sk* expression in WT and KO adrenal glands is shown in **Figure 3.S.4**. This normalization was used instead of using edgeR::calcNormFactors, which uses the trimmed mean of M-value normalization by default.

The heatmaps were constructed using R pheatmap. The volcano plots, pie charts, and bar plots were produced using R v3.6.3. ggplot2. The violin plot was generated by Raincloud Plots[398]. The chord pie chart was created by R circlize v0.4.9[399]. Based on a threshold for gene-level significance of 5% FDR, GO analysis of the RNA-seq data was performed using the Cytoscape v3.7.2[400] application ClueGO v2.5.6[401]. The Uniprot Gene Ontology Annotations were used for the classification of the GO terms.

3.5.11. RT-qPCR

Total RNA was extracted by RNeasy kit with on-column DNase A digestion (Qiagen) and reverse-transcribed with SuperScript IV Reverse Transcriptase (Invitrogen). PowerTrack™ SYBR Green Master Mix (ThermoFisher) was used with fast cycling mode (2-minute of enzyme activation at 95 °C, 40 cycles of 5-second denaturation at 95 °C, 30-second annealing and extension at 60 °C) on a 7500 Fast Real-time PCR System (ThermoFisher). *Rn7sk* is used as an internal control. Six pairs of mouse biological samples were analyzed for each gene. The $^{\Delta\Delta}$CT method was applied to the analysis of CT values. The paired two-way Student's *t*-tests was applied for mouse samples, and unpaired two-way Student's *t*-test was used for human samples. The primer sequences are provided in **Table 3.S.10**.

3.5.12. Chromatin immunoprecipitation and sequencing (ChIP-seq)

The adrenal glands from 8- to 12-month-old WT and KO female mice were resected and stored in liquid nitrogen immediately until use. The tissue-specific ChIP method was adapted from Cotney and Noonan's protocol[402]. Each gland was homogenized in 200 μl ice-cold PBS buffer with a handheld Polytron homogenizer. The final volume was brought

to 1 ml with ice-cold PBS. The homogenized samples were crosslinked with 66.7 μl 16% formaldehyde (1% final) at room temperature for 15 minutes. They were quenched with 107 μl 1.25 M glycine (0.125 M final) at room temperature for another 10 minutes in rotating tubes. The samples were centrifuged, and the pellets were washed twice with ice-cold PBS. The crosslinked pellets were re-homogenized with the Polytron homogenizer and suspended in 300 μl cell lysis buffer (50 mM Tris (pH 8.0), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100) and incubated on ice for 20 minutes to release nuclei. The nuclei were harvested by centrifugation, resuspended in 200 μl ChIP sonication buffer (10 mM Tris (pH 8.0), 1 mM EDTA, 0.5 mM EGTA, 0.5% SDS), and incubated on ice for 20 minutes. The nuclei were sonicated with a probe-based sonicator (FB120 with a CL-18 probe; ThermoFisher) at a 20% amplitude setting. The sonication was conducted using 15-second pulses at 15-second intervals for a total of 8 minutes. The sonicated nuclei were harvested by centrifugation and then diluted with 800 μl ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris (pH 8.1), 167 mM NaCl) to reach the final SDS concentration of 0.1%. These samples represented sonicated chromatin ready for immunoprecipitation.

To quantify chromatin and assess the degree of its fragmentation, we used a rapid de-crosslinking protocol by treating 5% of the sonicated nuclei (50 μl/sample) with 10 μg of RNase A for 15 minutes at 37 °C followed by 20 μg of proteinase K for 30 minutes at 65 °C. They were de-crosslinked for 5 minutes at 95 °C. DNA was extracted with the QIAquick PCR Purification Kit (Qiagen). DNA concentration was determined with a Nanodrop 1000 Fluorospectrometer (ThermoFisher). DNA fragment sizes were confirmed to be 200 – 800 bp in length, according to electrophoresis.

For immunoprecipitation, an equal amount of sonicated chromatin, based on their prior DNA measurements of different samples was reacted with anti-RPB1 N-terminal domain Ab (D8L4Y) (1:100) at 4 °C overnight, followed by 40 μl magnetic protein G beads (Bio-Rad) for another 2 hours at 4 °C. The beads were rinsed with wash buffer (100 mM Tris (pH 8.0), 500 mM LiCl, 1% NP-40, 1% deoxycholic acid) for five times and then with TE buffer

once. The chromatin was eluted with elution buffer (50 mM Tris (pH 8.0), 10 mM EDTA, 1% SDS) at 65 °C for 10 minutes. The immunoprecipitated chromatins were de-crosslinked at 65 °C overnight with NaCl adjusted to 200 mM. The chromatins were then treated with 10 μg RNase A/sample at 37 °C for 1 hour, followed by 200 μg proteinase K/sample for 2 hours at 45 °C. DNA of the samples was purified with QIAquick PCR Purification kit and quantified by the Bioanalyzer (Agilent).

Libraries were prepared robotically with 0.2 to 2 ng of fragmented DNA ranging $100 - 300$ bp in length, using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs), as per the manufacturer's recommendations. Adapters and PCR primers were purchased from Integrated DNA Technologies. Size selection was carried out using SparQ beads (Qiagen) prior to PCR amplification (12 cycles). Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). The average size of the fragments was determined using a LabChip GX (PerkinElmer) instrument.

The libraries were normalized and pooled, and then denatured in 0.05 N NaOH and neutralized using HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S4 lane using Xp protocol as per the manufacturer's recommendations. The run was performed for 2×100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1% level. Each library was sequenced at 25 million reads. Base-calling was performed with RTA v3. Program bcl2fastq2 v2.20 was then used to de-multiplex samples and generate fastq reads.

ChIP-seq reads were first trimmed for adapter sequences and low-quality score bases using Trimmomatic[395]. The resulting reads were mapped to the mouse reference genome (GRCm38) using BWA-MEM[403] in paired-end mode at default parameters. Only reads that had a unique alignment (mapping quality > 20) were retained, and PCR duplicates were marked using Picard tools. Peaks were called and annotated using MACS2[404] and HOMER[405] software suites, respectively.

To assess differences in Pol II occupancy patterns between WT and KO samples, we obtained ChIP-seq read counts within the following genomic regions using HOMER: the

promoter region (from TSS (transcription starting site) -400 bp to TSS +100 bp), gene body (from TSS +100 bp to TES (transcription ending site) -100 bp), region downstream of the gene body (from TES -100 bp to TES +2,000 bp), the 5'-untranslated region (5'-UTR), intron, 3'-UTR, the enhancer region (from TSS -5000 bp to TSS -400 bp), the region from -10,000 bp to TSS, the region from TSS to +10,000 bp, and the intergenic region. Since the RPB1 levels in the KO tissues were elevated, we speculated that there would be more Pol II association with the genes, hence a higher RPB1 ChIP signal in the KO promoter regions than in the WT counterparts. Therefore, genes that lacked RPB1 ChIP-seq signal in the promoter region in the KO tissues were filtered out, as these genes were believed to have no signals in WT tissues neither. Raw counts were normalized using edgeR's TMM algorithm[406] and were then transformed to $\log_2$ Counts Per Million ($\log_2$CPM) using the Voom function implemented in the Limma R package[407].

To construct the global metagene Pol II-binding profile, normalized read counts (Fragments per Kilobase of transcript per Million Mapped reads (FPKM) of a full gene length plus 2,000-bp flanks (TSS -2,000 bp to TES +2,000 bp) were obtained from all the genes that passed the filtering. Both flanks were divided into 20 equal-sized bins of 100 bp each. The gene bodies were scaled to 60 bins for the full gene length. FPKM was calculated from BAM input files using ngs.plot[408] with the following parameters: -G mm10 -R genebody -D ensembl -FL 200 -BOX 0 -SE 1 -VLN 0 -LWD 2 -WD 9. These global metagene Pol II binding profiles were only for visualization of differences in Pol II density, and customarily inferential statistics was not conducted for such profiling.

The peak count versus distance (-10 kb to +10 kb from TSS) profile was generated from 51 equal-sized bins of 400 bp for this region of all the genes that passed filtering. This profile was meant to give an overall view of the genomic location of all the peaks, and again, inferential statistics were not conducted.

To test for differential Pol II density in WT and KO tissues, we used the R package DESeq2[409] to analyze the raw counts of the promoter region, gene body, and the region downstream of the gene body. Differential expression analysis of DESeq2 is based on the

Negative Binomial (*a.k.a.* Gamma-Poisson) distribution. Genes with the threshold of 5% FDR were analyzed.

Genome browser tracks were created with the HOMER makeUCSCfile command and bedGraphToBigWig utility from UCSC. Tracks were normalized so that each value represented the read count per base pair per 10 million reads. UCSC Genome Browser (`http://genome.ucsc.edu`) was implemented for track visualization.

3.5.13. Patient cohort information

3.5.13.1. The Adelaide cohort.

PBMAH patients III-1, III-2, and III-3 were three male siblings from the first Australian kindred as we previously reported[364,3]. Briefly, III-1 presented with advanced Cushing's syndrome due to PBMAH. Despite bilateral adrenalectomy, he died from complications of advanced Cushing's syndrome. His two siblings, III-2 and III-3, self-presented for evaluation and were found to have mild hypercortisolism. Both underwent a single adrenalectomy, and at the most recent evaluation, remain eucortisolaemic. Their age, gender, and diagnosis are shown in **Table 3.1**. Additional clinical and laboratory findings of these 3 PBMAH patients were detailed previously[364]. These 3 PBMAH patients were genotyped by whole-exome sequencing, and *ARMC5* mutations were subsequently confirmed by Sanger sequencing[3]. They all carried the same heterozygous Chr16:g.31476121; c.1777C$\rightarrow$T mutation resulting R593W in the ARMC5 protein. The adrenal gland adenomas or hyperplasic adrenal glands from two patients with ACTH-independent Cushing's syndrome, one patient with a large adrenal incidentaloma, and three patients with primary aldosteronism were used as controls, as shown in **Table 3.1**. Also used as controls were two normal adrenal gland tissues, one (N-2) from a patient undergoing nephrectomy for renal cancer and the other (N-1) from normal adrenal gland tissues adjacent to the adrenal gland adenoma resected from patient PA-03 (**Table 3.1**). All the tissue samples were stored at -90 °C until use.

3.5.13.2. The Montreal cohort.

Table 3.1. Clinical information and genotype of the Australian cohort

Patient ID	Age at operation	Gender	Diagnosis	Secretion	Histopathology	Germline *ARMC5* mutation status
III-1	69	M	Cushing's due to PBMAH	Cortisol	PBMAH	Chr16:g.31476121; c.1777C→T; p.(R593W)
III-2	62	M	Cushing's due to PBMAH	Cortisol	PBMAH	Chr16:g.31476121; c.1777C→T; p.(R593W)
III-3	66	M	Cushing's due to PBMAH	Cortisol	PBMAH	Chr16:g.31476121; c.1777C→T; p.(R593W)
CS-01	39	F	Cushing's syndrome	Cortisol	Adrenal adenoma	not tested
CS-02	48	F	Cushing's syndrome	Cortisol	Adrenal adenoma	not tested
PA-01	57	M	Primary aldosteronism	Aldosterone	Adrenal adenoma	not tested
PA-02	76	M	Primary aldosteronism	Aldosterone	Adrenal hyperplasia micro- and macro-nodular	not tested
PA-03	50	M	Primary aldosteronism	Aldosterone	Adrenal hyperplasia micro- and macro-nodular	not tested
N-1	50	M	normal adrenal gland[a]	Nil excess	N/A	not tested
N-2	unknown	M	normal adrenal gland[b]	Nil excess	N/A	not tested

Note: For RPB1 expression analysis, PBMAH macronodules were obtained from resected adrenal glands of PBMAH patients with germline *ARMC5* mutations. Adrenal gland adenomas and adrenocortical carcinoma tissues not known to have *ARMC5* mutations were used as controls. The clinical diagnosis and histopathology of the patients are indicated. [a] The normal adrenal gland of patient N-1 was obtained from normal tissues adjacent to the adrenal gland adenoma of patient PA-03. [b] Normal adrenal gland N-2 was obtained from a patient undergoing nephrectomy for renal cancer. PBMAH: primary bilateral macronodular adrenal gland hyperplasia. N/A: not applicable.

PBMAH patients E35 and E202 were a father and daughter that were described previously[2,365]. They both had Cushing's syndrome secondary to PBMAH, and their cortisol secretion was beta-adrenergic/vasopressinresponsive. These PBMAH patients were genotyped by whole-exome sequencing, and *ARMC5* mutations were subsequently confirmed by Sanger sequencing. They both carried the heterozygous germline pathogenic variant in the *ARMC5* gene c.327_328insC, (p.Ala110Argfs*9) (**Table 3.2**). Patient E191 was a 47-year-old man with PBMAH co-secreting cortisol and aldosterone with clinical and biochemical Cushing's syndrome and primary aldosteronism. Moreover, the patient had a 2.8 cm mass on the pancreatic tail that was resected and compatible with a pancreatic neuroendocrine tumor at pathology. This patient carried a heterozygous germline *ARMC5* deletion of exons 5 – 8. The deletion is predicted to prematurely truncate the protein product and cause loss

of function. The adrenal gland adenomas from 3 patients with primary aldosteronism and one patient with adrenocortical carcinoma co-secreting cortisol and androgens were used as controls (**Table 3.2**). Two PBMAH patients without *ARMC5* mutation were included as additional controls.

Table 3.2. Clinical information and genotype of the Montreal cohort

Patient ID	Age at operation	Gender	Diagnosis	Secretion	Histopathology	Germline ARMC5 mutation status
E35	56	M	Cushing's syndrome	Cortisol	PBMAH	c.327_328insC, (p.Ala110Argfs*9)
E191	37	M	Cushing's syndrome and primary aldosteronism	Cortisol and aldosterone	PBMAH	Deletion of exons 5 to 8
E202	44	F	Cushing's syndrome	Cortisol	PBMAH	c.327_328insC, (p.Ala110Argfs*9)
B148	46	F	Primary aldosteronism	Aldosterone	Adrenal adenoma	not tested
B193	47	M	Primary aldosteronism	Aldosterone	Adrenal adenoma	not tested
B17	55	F	Primary aldosteronism	Aldosterone	Adrenal adenoma	not tested
B183	77	M	Cushing's syndrome and hirsutism	Cortisol and androgens	Adrenocortical carcinoma	not tested
B206	54	F	Cushing's syndrome	Cortisol and androgens	PBMAH	WT
E58	38	F	Cushing's syndrome	Cortisol	PBMAH	WT

Note: For RPB1 expression analysis, PBMAH nodules were obtained from resected adrenal gland macronodules of PBMAH patients with germline *ARMC5* mutations. Adrenal gland adenomas and adrenocortical carcinoma tissues not known to have *ARMC5* mutations were used as controls. Two PBMAH samples with confirmed WT *ARMC5* were included as additional controls. The clinical diagnosis and histopathology of the patients are indicated. PBMAH: primary bilateral macronodular adrenal gland hyperplasia. WT: wild type.

3.6. Data availability

The mouse RNA-seq dataset and ChIP-seq dataset have been deposited to the Gene Expression Omnibus of NCBI (accession #GSE169263 and #GSE169578, respectively). The human PBMAH microarray dataset has also been deposited to the Gene Expression Omnibus of NCBI (accession #GSE171558).

3.7. Ethics statement

All animal studies were approved by the Animal Protection Committee (Comité institutionnel d'intégration de la protection des animaux) of the CRCHUM. All patients provided written, informed consent to the retention of adrenal tissue for research purposes. For the Adelaide cohort, the study was approved by the Royal Adelaide Hospital Human Research Ethics Committee, for the Montreal, by the Ethical Committee at Centre hospitalier de l'Université de Montréal (CHUM) (SL05-054).

3.8. Acknowledgments

3.9. Supplementary Figures

3.9.1. Supplementary Figure 1
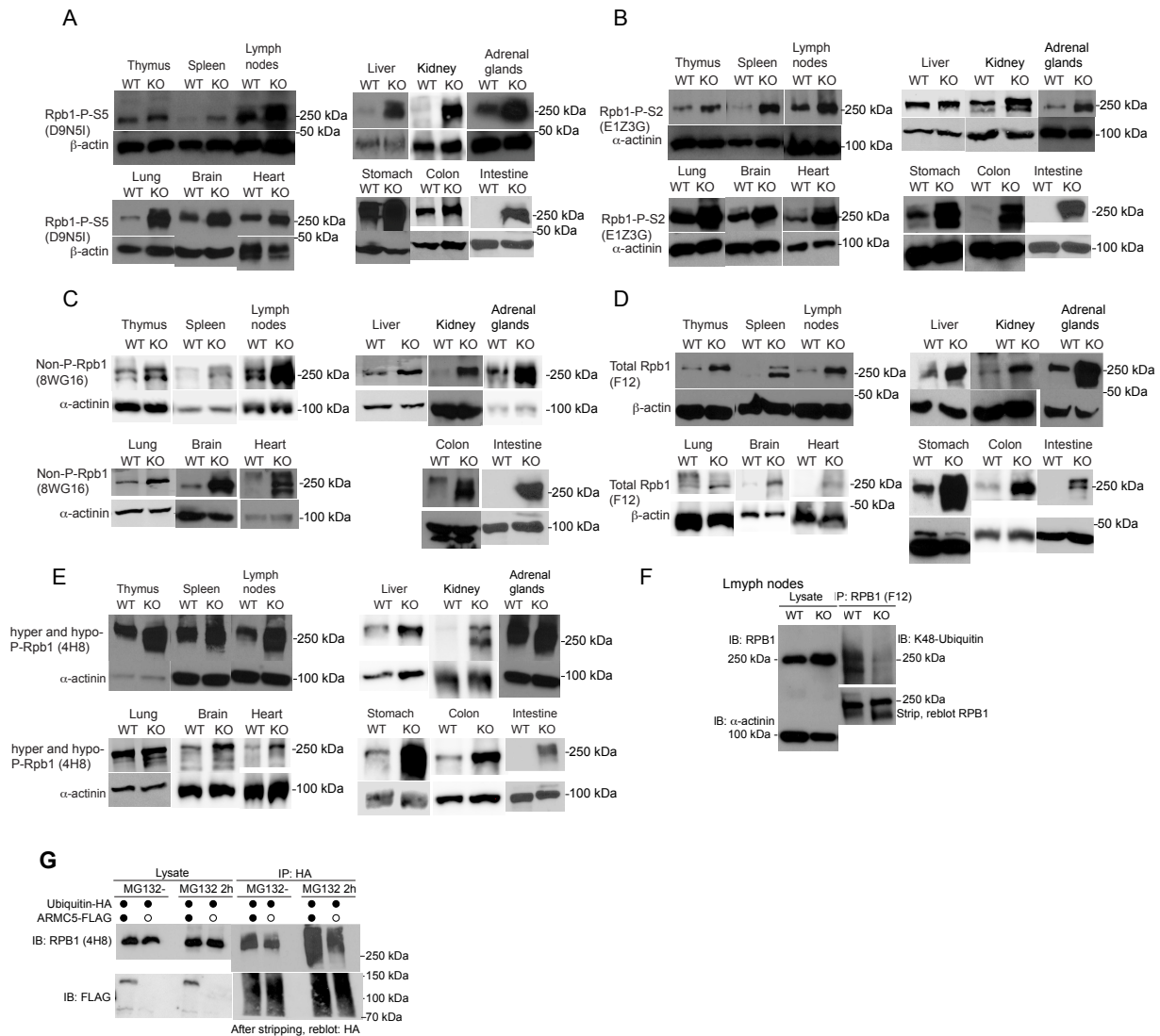
Fig. 3.S.1. The expression and ubiquitination of RPB1 in KO and WT mice/cells The expression of RPB1 of different phosphorylation statuses in the thymus, spleen, lymph nodes, liver, adrenal glands, lung, brain, heart, stomach, colon, and intestine was determined by Western blotting, using different anti-RPB1 Abs. **a**. RPB1 with phosphorylation of S5 in CTD repeats (RPB1-S5; mAb clone D9N5I). **b**. RPB1 with phosphorylation of S2 in CTD repeats (RPB1-S2; mAb clone E1Z3G). **c**. Non-phosphorylated RPB1 (mAb 8WG16). **d**. Total RPB1 (NTD; mAb F12 against N-terminal domain). **e**. RPB1 with both high and low CTD phosphorylation (mAb 4H8). *(see next page)*

3.9.2. Supplementary Figure 2

Fig. 3.S.2. Similar expression of Rn7sk in the WT and KO adrenal glands
The *Rn7sk* levels in the WT and KO adrenal glands were determined by RNA-seq. The results ($\log_2$CPM) of three pairs of WT and KO adrenal glands are shown.

3.9.3. Supplementary Figure 3
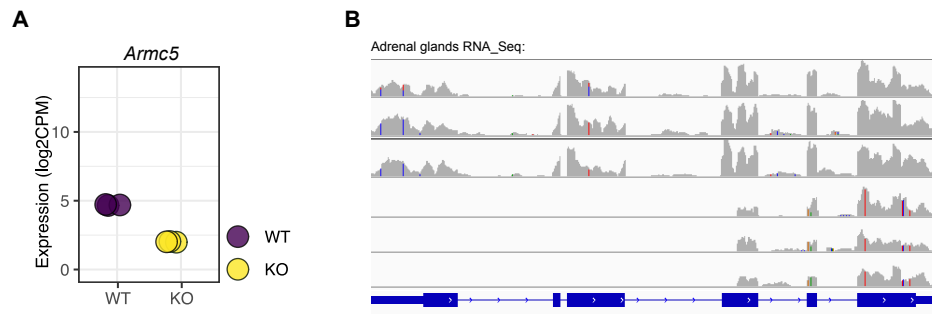
Fig. 3.S.3. *Armc5* **deletion in the adrenal glands according to RNA-seq a**. log$_2$CPM of *Armc5* in the KO and WT adrenal glands. **b**. The RNA-seq reads corresponding to the *Armc5* gene region are shown.

3.9.4. Supplementary Figure 4
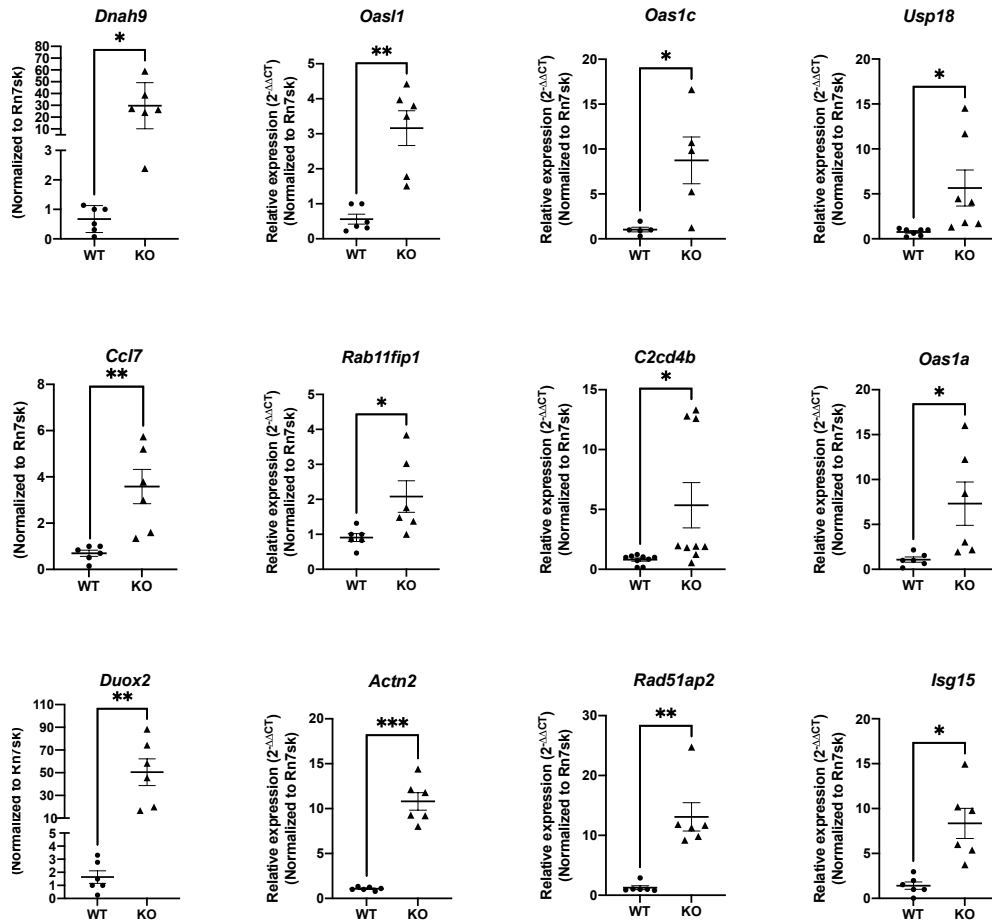
Fig. 3.S.4. RT-qPCR validation of differentially expressed genes in the KO adrenal glands Some genes with prominent expression differences in the KO adrenal gland according to RNA-seq were selected and validated in the KO and WT adrenal glands with RT-qPCR. The signal ratios of the test genes versus *Rn7sk* were presented. Means with 95% CI are shown. *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$ (paired two-way Student's *t*-tests).

3.10. Supplementary Tables

3.10.1. Supplementary Table 1

Table 3.S.1. Proteins found in the ARMC5 precipitates

Protein ID	MW (kDa)	ARMC5-HA	Empty Control	Protein ID	MW (kDa)	ARMC5-HA	Empty Control
ARMC5	98	91	12	RARS	75	9	0
HEL-S-72p	71	70	33	UbC	77	9	1
HNRNPUL1	96	46	13	ACSL1	78	9	0
HSPA9	72	45	17	SLC25A5	33	8	2
HSPA1A	70	34	13	TUBB	50	8	3
EEF2	95	30	4	PABPC4	68	8	0
MTHFD1	102	27	3	DNA helicase	91	8	0
HSP90B1	92	25	3	HSPA4	94	8	0
ENO1	47	24	5	SLFN11	103	8	1
ERP70	73	23	0	DARS	57	7	1
NCL	77	23	1	HSPD1	61	7	0
HSPA5	72	20	7	IGF2BP1	63	7	1
SAMHD1	72	19	0	RBM14	69	7	2
EEF1A1	50	18	8	HEL-S-71p	77	7	3
KARS	68	18	2	ACTN4	105	7	1
XRCC6	70	18	7	DDX20	92	7	0
ATP5A1	60	17	3	RUVBL1	50	6	2
PABPC1	71	15	1	MAGED2	65	6	0
MAGED1	86	15	2	RAF1	70	6	0
UBE1	118	15	0	SDHA	73	6	1
GANAB	96	14	1	AARS	107	6	0
MARS	101	13	0	ACLY	121	6	0
CAD	236	13	0	KIF5B	110	6	0
ATP5B	57	12	3	SND1	102	6	0
TRIM28	89	12	0	TUBB2A	50	5	0
RUVBL2	51	11	1	FUS	53	5	1
PGM3	60	11	0	Septin 9	64	5	0
TRAP1	80	11	0	ATP6V1A	68	5	0
HSP90AB1	83	11	0	MAGED4	81	5	0
FUBP1	70	10	0	PARP1	113	5	0
HSPA2	70	10	1	RPB1	217	5	0
HSPH1	92	10	0	HRNR	282	5	2
MCM4	97	10	0				

Note: Proteins found in the ARMC5 precipitates and met with the following two conditions in any of 3 biological replicates were listed. 1) The protein had equal or more than five peptides corresponding to its sequence in the ARMC5-HA transfected sample; 2) the number of the peptides in the ARMC5-HA transfected sample was more than 2-folds larger than that in the empty vector control.

3.10.2. Supplementary Table 2

Table 3.S.2. Differentially expressed genes in the KO and WT adrenal glands

| Gene name | FC | log$_2$FC | P Value | FDR | Rowcounts | | | | | |
					WT1	WT2	WT3	KO1	KO2	KO3
Tfcp2l1	11.43	3.51	8.22E-53	1.46E-48	731	856	467	7227	4748	8255
Atp1b2	7.31	2.87	1.05E-47	9.31E-44	4750	5985	4234	33895	22353	40293
Dnah9	85.57	6.42	2.51E-43	1.48E-39	11	19	6	823	716	991
Rnu1b2	73.49	6.20	2.76E-38	9.78E-35	18	9	5	1006	482	449
Rnu1b1	68.39	6.10	7.52E-37	2.22E-33	18	9	5	924	450	425
Sesn3	5.81	2.54	1.66E-35	4.20E-32	4331	3902	2986	24322	10294	25649
Actn2	1862.21	10.86	1.44E-31	3.18E-28	0	0	0	314	106	263
Gm10167	20.51	4.36	1.28E-26	1.88E-23	34	35	28	544	511	674
Armc5	-6.38	-2.67	1.84E-26	2.51E-23	2763	3263	2233	395	269	473
Vgf	7.31	2.87	4.73E-26	5.98E-23	293	321	206	1929	1217	2089
Gm48482	49.05	5.62	1.69E-23	1.66E-20	3	3	3	132	120	141
Ccno	13.56	3.76	3.38E-23	2.99E-20	45	66	30	476	484	635
Gm25890	35.34	5.14	6.59E-23	5.55E-20	147	40	22	2512	1561	990
Rad51ap2	18.56	4.21	4.16E-22	3.20E-19	21	25	22	432	296	378
Gm22614	36.90	5.21	4.48E-22	3.23E-19	174	44	23	2817	1758	1213
Pcdh8	737.62	9.53	4.57E-22	3.23E-19	0	0	3	506	800	711
Nptx2	156.74	7.29	6.16E-22	4.19E-19	3	9	0	443	352	597
Gm26232	38.87	5.28	2.60E-21	1.71E-18	154	38	17	2537	1599	987
Gm24830	38.82	5.28	2.73E-21	1.72E-18	154	38	17	2535	1599	984
Aqp3	43.91	5.46	3.86E-21	2.36E-18	5	21	4	358	277	313
C2cd4b	11.16	3.48	6.97E-21	4.12E-18	44	57	29	457	349	413
Rnu1b6	39.68	5.31	8.70E-21	4.97E-18	141	34	15	2330	1467	907
Csrnp2	6.35	2.67	2.63E-20	1.46E-17	498	603	641	3127	2546	4016
Amigo2	5.62	2.49	2.81E-20	1.51E-17	700	923	479	3361	2275	4658
Mgat4c	7.24	2.86	3.61E-20	1.88E-17	4707	6923	3632	35909	17035	47243
Mafa	15.35	3.94	6.81E-20	3.45E-17	14	30	27	270	254	411
Gfpt2	6.91	2.79	2.01E-19	9.88E-17	346	398	175	1836	1272	2227
Gm22042	41.18	5.36	2.46E-19	1.15E-16	56	15	6	1002	633	417
Gm22973	11.27	3.49	2.55E-19	1.16E-16	566	639	349	7313	5424	2904
Nefm	8.70	3.12	1.86E-18	8.26E-16	95	107	84	846	532	812
A230083N12Rik	13.82	3.79	2.10E-18	9.08E-16	17	25	25	290	193	337
Gm49774	36.93	5.21	3.39E-18	1.40E-15	10	49	11	576	464	764
Nanos1	7.07	2.82	9.17E-18	3.61E-15	167	144	123	933	650	1120
2610035D17Rik	22.37	4.48	5.86E-17	2.16E-14	56	32	11	588	443	619
4930560O18Rik	77.42	6.27	3.58E-16	1.30E-13	9	3	0	256	150	282
Folh1	-10.31	-3.37	1.63E-15	5.71E-13	371	321	255	35	11	42
Rnf26	4.96	2.31	2.88E-15	9.81E-13	1058	1374	546	4570	2702	5104
Duox2	62.01	5.95	3.50E-15	1.17E-12	0	5	5	187	149	176
Gpha2	4.81	2.26	3.97E-15	1.30E-12	828	1047	714	2920	3912	3904

Note: According to RNA-seq of KO and WT adrenal glands, top 40 of the most significant genes are listed. Additional information, such as gene ID, fold change (FC), *p*-values, FDRs are raw counts are also provided.

3.10.3. Supplementary Table 3

Table 3.S.3. Genes with differential Pol II density in the TSS region in the KO and WT adrenal glands

Gene region	log$_2$FC (WT vs KO)	FDR
Pcdh8_TSS	-3.25	3.08E-13
Tfcp2l1_TSS	-2.74	1.02E-10
Cyp4f14_TSS	-2.59	1.81E-05
Mafa_TSS	-2.39	7.26E-17
Celf4_TSS	-2.19	5.85E-09
Actn2_TSS	-2.19	7.32E-04
Serpinb1a_TSS	-2.08	5.39E-11
Oas1a_TSS	-2.07	2.00E-06
Oas1g_TSS	-2.06	4.15E-03
Dnah9_TSS	-2.06	1.80E-02
C2cd4b_TSS	-2.00	3.90E-09

Note: The genes with differential Pol II density in the TSS region in the KO versus WT adrenal glands (log$_2$FC $\geq$ 2 or $\leq$ -2) are listed, along with their log$_2$FCs and FDRs.

3.10.4. Supplementary Table 4

Table 3.S.4. Genes with differential Pol II density in the gene body in the KO and WT adrenal glands

Gene region	log$_2$FC (WT vs KO)	FDR
Hist1h1a_GENEBODY	-2.41	3.74E-12
Pcdh8_GENEBODY	-2.32	2.20E-12
Nicn1_GENEBODY	-2.17	2.24E-14
Mafa_GENEBODY	-2.12	2.24E-05
Ccl7_GENEBODY	-2.11	4.36E-06
Hist1h2bm_GENEBODY	-2.00	7.63E-05

Note: The genes with differential Pol II density in the gene body region in the KO versus WT adrenal glands (log$_2$FC $\geq$ 2 or $\leq$ -2) are listed, along with their log$_2$FCs and FDRs.

3.10.5. Supplementary Table 5

Table 3.S.5. Genes with differential Pol II density in the TES region in the KO and WT adrenal glands

gene region	log$_2$FC (WT vs KO)	FDR
Hist1h1a_TES	-2.66	6.46E-11
1700013F07Rik_TES	-2.44	2.57E-09
Serpinb1a_TES	-2.38	2.18E-06
Nabp1_TES	-2.32	4.95E-15
Id4_TES	-2.24	1.87E-06
Atp1b2_TES	-2.16	1.63E-07
Hspa1a_TES	-2.11	1.34E-05
Mc1r_TES	-2.09	7.00E-09
Caprin2_TES	-2.09	6.40E-04
Tuba3a_TES	-2.06	4.88E-05
Cwc22_TES	4.09	5.10E-03

Note: The genes with differential Pol II density in the TES region in the KO versus WT adrenal glands (log$_2$FC ≥ 2 or $\leq$ -2) are listed, along with their log$_2$FCs and FDRs.

3.10.6. Supplementary Table 6

Table 3.S.6. ChIP-seq FCs and FDRs in the TSS region of the significantly upregulated genes according to RNA-seq

RNA-seq				ChIP-seq		
Gene	FC	log$_2$FC (KO vs WT)	FDR	Gene region	log$_2$FC (WT vs KO)	FDR
Pcdh8	737.62	9.53	3.23E-19	Pcdh8_TSS	-3.25	3.08E-13
Tfcp2l1	11.43	3.51	1.46E-48	Tfcp2l1_TSS	-2.74	1.02E-10
Mafa	15.35	3.94	3.45E-17	Mafa_TSS	-2.39	7.26E-17
Celf4	2.12	1.09	1.96E-03	Celf4_TSS	-2.19	5.85E-09
Actn2	1862.21	10.86	3.18E-28	Actn2_TSS	-2.19	7.32E-04
Oas1a, Oas1g	6.13	2.62	6.46E-05	Oas1a_TSS	-2.07	2.00E-06
Oas1a, Oas1g	6.13	2.62	6.46E-05	Oas1g_TSS	-2.06	4.15E-03
Dnah9	85.57	6.42	1.48E-39	Dnah9_TSS	-2.06	1.80E-02
C2cd4b	11.16	3.48	4.12E-18	C2cd4b_TSS	-2.00	3.90E-09

Note: The ChIP-seq parameters are based on the analysis of the TSS regions. Only genes with log$_2$FC $\geq$ 2 or $\leq$ -2 from ChIP-seq analysis are listed here.

3.10.7. Supplementary Table 7

Table 3.S.7. ChIP-seq FCs and FDRs in the gene body of the significantly upregulated genes according to RNA-seq

	RNA-seq				ChIP-seq	
Gene	FC	$\log_2$FC (WT vs KO)	FDR	Gene region	$\log_2$FC (WT vs KO)	FDR
Hist1h1a	2.57	1.36	9.98E-05	Hist1h1a_GENEBODY	-2.41	3.74E-12
Pcdh8	737.62	9.53	3.23E-19	Pcdh8_GENEBODY	-2.32	2.20E-12
Mafa	15.35	3.94	3.45E-17	Mafa_GENEBODY	-2.12	2.24E-05
Ccl7	6.21	2.63	6.01E-10	Ccl7_GENEBODY	-2.11	4.36E-06

Note: The ChIP-seq parameters are based on the analysis of the genebody regions. Only genes with $\log_2$FC $\geq$ 2 or $\leq$ -2 from ChIP-seq analysis are listed here.

3.10.8. Supplementary Table 8

Table 3.S.8. ChIP-seq FCs and FDRs in the TES region of the significantly upregulated genes according to RNA-seq

	RNA-seq			ChIP-seq		
Gene	**FC**	**log$_2$FC (KO vs WT)**	**FDR**	**gene region**	**log$_2$FC (WT vs KO)**	**CFDR**
Hist1h1a	2.57	1.36	9.98E-05	Hist1h1a_TES	-2.66	6.46E-11
Nabp1	2.78	1.48	1.21E-03	Nabp1_TES	-2.32	4.95E-15
Id4	5.84	2.55	8.31E-09	Id4_TES	-2.24	1.87E-06
Atp1b2	7.31	2.87	9.31E-44	Atp1b2_TES	-2.16	1.63E-07
Caprin2	2.69	1.43	6.06E-10	Caprin2_TES	-2.09	6.40E-04

Note: The ChIP-seq parameters are based on the analysis of the TES regions. Only genes with log$_2$FC $\geq$ 2 or $\leq$ -2 from ChIP-seq analysis are listed here.

3.10.9. Supplementary Table 9

Table 3.S.9. Comparison of differentially expressed genes in human PBMAH adrenal gland macronodules and mouse *Armc5* KO adrenal glands

Human gene name	FC [1]	p-value [1]	FDR [1]	Mouse gene name	FC [2]	p-value [2]	FDR [2]
Commonly upregulated genes in human PBMAH and mouse KO adrenal glands							
F2RL1	4.89	6.40E-04	2.83E-02	F2rl1	16.08	1.06E-04	2.52E-03
NEFM	4.55	1.07E-03	3.57E-02	Nefm	8.70	1.86E-18	8.26E-16
ATP1B2	11.06	6.40E-07	1.18E-03	Atp1b2	7.31	1.05E-47	9.31E-44
CRTAC1	11.59	4.20E-07	1.09E-03	Crtac1	7.26	4.06E-08	3.63E-06
GFPT2	8.35	8.21E-06	3.53E-03	Gfpt2	6.91	2.01E-19	9.88E-17
BAIAP2	5.15	4.33E-04	2.40E-02	Baiap2	4.44	4.13E-12	8.81E-10
SLC4A8	5.87	1.58E-04	1.38E-02	Slc4a8	4.43	6.21E-04	9.92E-03
ZCCHC12	10.86	7.60E-07	1.18E-03	Zcchc12	4.01	2.78E-10	4.14E-08
MT3	4.09	2.18E-03	4.94E-02	Mt3	3.81	2.78E-08	2.63E-06
MFHAS1	5.31	3.45E-04	2.13E-02	Mfhas1	3.74	4.42E-03	4.40E-02
PCP4	7.82	1.47E-05	4.56E-03	Pcp4	3.56	9.02E-07	5.20E-05
NRIP3	6.14	1.11E-04	1.19E-02	Nrip3	3.54	1.60E-07	1.17E-05
CIT	5.66	2.10E-04	1.61E-02	Cit	3.40	8.57E-04	1.27E-02
TUBB3	4.85	6.80E-04	2.88E-02	Tubb3	3.20	1.89E-08	1.87E-06
NOVA1	7.76	1.56E-05	4.69E-03	Nova1	3.06	3.20E-11	5.85E-09
LRRTM3	7.73	1.61E-05	4.73E-03	Lrrtm3	2.89	2.62E-05	8.22E-04
NETO2	5.13	4.45E-04	2.41E-02	Neto2	2.76	2.34E-07	1.63E-05
EPHA7	6.62	6.01E-05	8.80E-03	Epha7	2.69	1.58E-08	1.62E-06
HSPA2	6.21	1.02E-04	1.14E-02	Hspa2	2.68	2.65E-11	4.99E-09
SCN8A	5.59	2.32E-04	1.72E-02	Scn8a	2.59	7.96E-05	2.01E-03
BMPR1B	6.70	5.45E-05	8.32E-03	Bmpr1b	2.55	1.65E-04	3.57E-03
CNTN1	8.43	7.55E-06	3.47E-03	Cntn1	2.51	2.80E-04	5.41E-03
HAP1	11.65	3.90E-07	1.09E-03	Hap1	2.49	5.73E-11	9.48E-09
TAF4B	6.88	4.38E-05	7.56E-03	Taf4b	2.47	3.01E-09	3.61E-07
GLDC	5.11	4.60E-04	2.43E-02	Gldc	2.46	5.03E-04	8.55E-03
CHGB	7.90	1.34E-05	4.36E-03	Chgb	2.43	3.98E-05	1.16E-03
GATA3	4.13	2.04E-03	4.76E-02	Gata3	2.34	2.42E-07	1.68E-05
RORB	4.58	1.02E-03	3.48E-02	Rorb	2.28	3.42E-03	3.62E-02
NAP1L3	4.92	6.12E-04	2.79E-02	Nap1l3	2.20	1.75E-04	3.72E-03
ALCAM	7.85	1.41E-05	4.47E-03	Alcam	2.00	2.19E-06	1.08E-04
VAMP1	9.68	2.19E-06	1.66E-03	Vamp1	2.00	7.19E-06	2.88E-04
MMP16	12.97	1.40E-07	8.30E-04	Mmp16	1.98	5.01E-03	4.82E-02
KIF5C	6.08	1.19E-04	1.22E-02	Kif5c	1.90	4.41E-05	1.26E-03
FAM117A	4.93	5.96E-04	2.76E-02	Fam117a	1.90	1.23E-04	2.82E-03
PGM2L1	6.52	6.79E-05	9.42E-03	Pgm2l1	1.87	5.01E-03	4.82E-02
RGS7	10.30	1.24E-06	1.36E-03	Rgs7	1.84	2.08E-03	2.47E-02
CSRP2	6.61	6.10E-05	8.80E-03	Csrp2	1.79	7.46E-04	1.14E-02
SLC16A1	5.57	2.40E-04	1.72E-02	Slc16a1	1.62	3.54E-03	3.72E-02
ZFHX4	4.17	1.93E-03	4.59E-02	Zfhx4	1.62	3.10E-04	5.87E-03
EPS8	5.95	1.43E-04	1.34E-02	Eps8	1.62	5.19E-03	4.95E-02

Human gene name	FC [1]	*p-value* [1]	FDR [1]	Mouse gene name	FC [2]	*p-value* [2]	FDR [2]
SLC12A2	4.19	1.87E-03	4.52E-02	Slc12a2	1.54	2.20E-03	2.60E-02
RNPC3	4.08	2.22E-03	4.99E-02	Rnpc3	1.51	3.76E-03	3.89E-02
INTS6	4.49	1.17E-03	3.69E-02	Ints6	1.49	9.56E-04	1.38E-02

Commonly downregulated genes in human PBMAH and mouse KO adrenal glands

TAGLN	-5.99	1.36E-04	1.32E-02	Tagln	-1.75	1.41E-03	1.83E-02
MPHOSPH9	-7.64	1.79E-05	4.91E-03	Mphosph9	-2.06	8.55E-05	2.13E-03
RASSF4	-4.17	1.94E-03	4.59E-02	Rassf4	-2.40	1.22E-05	4.45E-04
IL1RL1	-4.28	1.62E-03	4.25E-02	Il1rl1	-4.56	4.08E-05	1.19E-03

Note: Microarray data were generated using resected adrenal gland macronodules from three Adelaide PBMAH patients and two normal adrenal glands, as detailed in Table 3.1. Mouse RNA-seq was conducted using adrenal glands from three pairs of KO and WT mice. The differentially expressed genes in each dataset (FDR < 0.05) were compared, and those with upregulation or downregulation in both datasets are listed along with their FC, *p*-value, and FDR.

[1]Human PBMAH vs normal adrenal
[2]Mouse KO vs WT adrenal glands

3.10.10. Supplementary Table 10 (DNA Sequences)

Table 3.S.10. DNA sequences

Name	Sequence
Primers for cloning	
ARMC5 Δaa(2 − 142), Forward	GGGGCGTGCCGGACCGAA
ARMC5 Δaa(2 − 142), Reverse	CATGGTACCGAATTCCTTCAAGCCTGCTTTTTTG
ARMC5 Δaa(143 − 444), Forward	CCTGAGCGGGCACAGGGT
ARMC5 Δaa(143 − 444), Reverse	TTCCGTACAGCAATCGGCTAGG
ARMC5 Δaa(445 − 747), Forward	CCCGACCTGCACTTCCTG
ARMC5 Δaa(445 − 747), Reverse	GGTCCTCTCCTCAGGAAAG
ARMC5 Δaa(748 − 816), Forward	CTGGGGCCCGTGCCCCCA
ARMC5 Δaa(748 − 816), Reverse	AGCTGGGACAGGGGCTGGG
ARMC5 Δaa(817 − 934), Forward	TACCTCGAGTGCGGCCGC
ARMC5 Δaa(817 − 934), Reverse	CGGGGGTGTGGGGCTGCC
ARMC5(BTB) (aa748 − 935), Forward	AGCAGGTACCATGCCCGACCTGCACTTCCTG (KpnI+∼)
ARMC5(BTB) (aa748 − 935), Reverse	GCGGCCGCGGCAGCCCCACACCCCCG (∼+ NotI)
CUL3 Δaa(31 − 385), Forward	GAATACCTCTCATTATTTATTGATG
CUL3 Δaa(31 − 385), Reverse	TACATATTTTTCATCCATGGTC
CUL3 Δaa(377 − 675), Forward	ACAGTTGCTGCCAAACAAG
CUL3 Δaa(377 − 675), Reverse	ATACTCAAAGTCACCCGC
CUL3 Δaa(695 − 762), Forward	GTATACACATATGTAGCAAC
CUL3 Δaa(695 − 762), Reverse	TTTCTGCCTTGTTTCTTTC
CUL3 aa(1 − 376), Forward	GTA TAC ACA TAT GTA GCA AC
CUL3 aa(1 − 376), Reverse	ATA CTC AAA GTC ACC CGC
Primers for RT-qPCR (mouse)	
Rpb1, Forward	CAC TGT CAT CAC CCC TGA CC
Rpb1, Reverse	ATA CTG GCT GTT TCC CCT GC
7skRNA, Forward	TCA CCC CAT TGA TCG CCA GGG T
7skRNA, Reverse	CAC ATG GAG CGG TGA GGG AGG A
Mafa, Forward	ATT CTG GAG AGC GAG AAG TGC CAG
Mafa, Reverse	CGC CAA CTT CTC GTA TTT CTC CTT
StAR, Forward	ATT TTG GGG AGA TGC CGG AG
StAR, Reverse	GCC ACC CCT TCA GGT CAA TAC
Tfcp2l1, Forward	CAG CCT CTA TCC AGG ATG CAC A
Tfcp2l1, Reverse	CTC TGG ACA TCT TCA GGA GGT C
Pcdh8, Forward	CAT GCA GAG TGG ACT GTG GGC G
Pcdh8, Reverse	TAC GTT GGG TCC GGC ACA GGA T

Table 3.S.10 continued from previous page

Name	Sequence
Actn2, Forward	CGG GAT TAC CGT CGT AAG CA
Actn2, Reverse	TGA TCC GAA GCT TGG TCT GC
Oasl1, Forward	TGA AGA GCC TCC TTC GGT TGG T
Oasl1, Reverse	TCC AGC CTG AAG TTG GCA TCC T
Oas1a, Forward	GAG GTG GAG TTT GAT GTG CTG C
Oas1a, Reverse	GTG AAG CAG GTA GAG AAC TCG C
Oas1c, Forward	AGC TCG ACT TCC ATC TGT CC
Oas1c, Reverse	GTC CAC CCC TTT CTG GCA ATT A
Rad51ap2, Forward	CTC CAT TCT ACT CCT GAG GAA GA
Rad51ap2, Reverse	GAG GCT GTA CTC TGG AAA TCC C
Isg15, Forward	GAC CAG TTC TGG CTG AGC TT
Isg15, Reverse	GGG GCT TTA GGC CAT ACT CC
Usp18, Forward	CAG GAG TCC CTG ATT TGC GT
Usp18, Reverse	GGG CTG GAC GAA ACA TCT CA
C2cd4b, Forward	CTG CTT CGT TCC TGC GAC TA
C2cd4b, Reverse	GAG CGA AGT CGA CCG AGA AG
Dnah9, Forward	AGC CAG TCT CAG ATG CCA TAG A
Dnah9, Reverse	CCA GGA GTT CAG CCA TTC CC
Ccl7, Forward	AAG TGG GTC GAG GAG GCT AT
Ccl7, Reverse	CCA TTC CTT AGG CGT GAC CA
DuoX2, Forward	CGT TCA TCA ACC GGA CTC CT
DuoX2 Reverse	GGC CCC ATT ACC TTT TTG CC
Rab11fip1, Forward	CCT TTG AGG ACG TGC AGA TCT C
Rab11fip1, Reverse	TTT GGG CTG AGC CTC TGG AGA A

Primers for RT-qPCR (human)

POLR2A, Forward	ACG CTG CTC TTC AAC ATC CA
POLR2A, Reverse	GGC AGA CAC ACC AGC ATA GT
7skRNA, Forward	TCA CCC CAT TGA TCG CCA GGG T
7skRNA, Reverse	CAC ATG GAG CGG TGA GGG AGG A
MAFA, Forward	GCT TCA GCA GGA GGA GGG TCA T
MAFA, Reverse	TCT GGA GTT GGC ACT TCT CGC T
TAF4B, Forward	CAA GAG CCG AGA CCA CAA GT
TAF4B, Reverse	AGG AAG GCT TCG GAA CAG TG

3.10.11. Supplementary Table 11 (Key reagents and resource)

Table 3.S.11. Key reagents and resource

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
HA-Tag (6E2) (HRP Conjugate)	Cell Signaling Technology	2999S; RRID:AB_1264166
c-Myc (9E10)	Santa Cruz	sc-40; RRID:AB_627268
CUL-3 (G-8)	Santa Cruz	sc-166110; RRID:AB_2245478
RBX1	Cell Signaling Technology	11922; RRID:AB_2797769
RPB1 (4H8)	BioLegend	904001; RRID:AB_2565036
RPB1 (8WG16)	BioLegend	664906; RRID:AB_2565554
Rpb1 NTD (D8L4Y)	Cell Signaling Technology	14958; RRID:AB_2687876
Rpb1 NTD (F12)	Santa Cruz	sc-55492; RRID:AB_630203
Phospho-Rpb1 CTD (Ser2) (E1Z3G)	Cell Signaling Technology	13499S; RRID:AB_2798238
Phospho-Rpb1 CTD (Ser5) (D9N5I)	Cell Signaling Technology	13523S; RRID:AB_2798246
Ubiquitin (F-11)	Santa Cruz	sc-271289; RRID:AB_10611436
α-actinin	Cell Signaling Technology	6487; RRID:AB_11179206
β-actin	Cell Signaling Technology	4967L; RRID:AB_330288
K48-ubiquitin (Apu2)	Millipore Sigma	05-1307; RRID:AB_1587578
Anti-mouse IgG-HRP	Cell Signaling Technology	7076S; RRID:AB_330924
Anti-rabbit IgG-HRP	Cell Signaling Technology	7074S; RRID:AB_2099233
Anti-HA-Tag	Millipore Sigma	H3663; RRID:AB_262051
HA-Tag agarose beads	Millipore Sigma	A2095; RRID:AB_257974
FLAG (M2)	Millipore Sigma	F1804; RRID:AB_262044
FLAG M2 affinity beads	Millipore Sigma	A2220; RRID:AB_10063035
Anti-HA-Tag	Cell Signaling Technology	3724S; RRID:AB_1549585
Alexa Fluor 488 goat anti-mouse IgG (H+L)	Invitrogen	A11001; RRID:AB_2534069
Alexa Fluor 488 Phalloidin	Thermo Fisher Scientific	A12379
Alexa Fluor 555 goat anti-rabbit IgG (H+L)	Thermo Fisher Scientific	A21428; RRID:AB_141784
Bacterial		
One Shot TOP10 Chemically Competent E. coli	Thermo Fisher Scientific	C404003
NEB® 10-beta Competent E. coli	New England BioLabs	C3019H
Chemicals, Peptides, and Recombinant Proteins		
HA peptides	GenScript	RP11735
FLAG peptides	GenScript	RP10586
CUL3/Rbx1, GST-tag	BPS Bioscience	80409
UBE1	BostonBiochem	E-305
His6-UbcH6/UBE2E1	BostonBiochem	E2-630
UbcH(E2) Enzyme Kit	BostonBiochem	K-980B
His6-Ubiquitin	BostonBiochem	U-530

Table 3.S.11 continued from previous page

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BLUelf Prestained Protein Ladder	FroggaBio	PM008-R500
Precision Plus Protein Dual Color Standards	Bio-Rad	1610374
SureBeads Protein G Magnetic beads	Bio-Rad	161-4023
(R)-MG132	Cayman Chemical	13697
N-Ethylmaleimide (NEM)	Millipore Sigma	E3876
16% Formaldehyde Solution(w/v) Methanol-free	Thermo Fisher Scientific	28908
SuperScript IV Reverse Transcriptase	Thermo Fisher Scientific	18090010
DNase I, Amplification Grade	Thermo Fisher Scientific	18068015
RNase A, DNase and protease-free)	Thermo Fisher Scientific	EN0531
PROTEINASE K	WISENT	800-030-EM
Methanol	Sigma	34860
2-Chloroacetamide	Millipore Sigma	C0267
Tris-(2-Carboxyethyl)phosphine, Hydrochloride (TCEP)	Thermo Fisher Scientific	T2556
TRIzol Reagent	Thermo Fisher Scientific	15596026
Q5 Hot Start High-Fidelity DNA polymerase	New England BioLabs	M0493S
KOD Hot Start DNA polymerase	Millipore Sigma	71086
Critical commercial assays		
Lipofectamine 2000	Thermo Fisher Scientific	11668019
Lipofectin	Thermo Fisher Scientific	18292011
Halt™ Protease Inhibitor Cocktail	Thermo Fisher Scientific	78438
PhosSTOP	Roche	04 906 837 001
PowerTrack SYBR Green Master Mix	Thermo Fisher Scientific	A46113
RNeasy kit	QIAGEN	74104
Immobilon-P PVDF Membrane	Millipore Sigma	IPVH00010
Western Lightning ECL Pro	PerkinElmer	ORT2505
ProtoGlow ECL (enhanced)	National diagnostics	CL-300
HyBlot, Autoradiography film	Denville Scientific	E3218
Pierce Acetonitrile (ACN), LC-MS Grade	Thermo Fisher Scientific	51101
Q5 Site-Directed Mutagenesis Kit	New England BioLabs	E0554S
Pierce Silver Stain for Mass Spectrometry	Thermo Fisher Scientific	24600
Monarch DNA Gel Extraction kit	New England BioLabs	T1020S
PureLink HiPure Plasmid Maxiprep kit	Thermo Fisher Scientific	K210007
4 − 15% Mini-PROTEAN TGX Precast Protein Gels	Bio-Rad	4561084
Amico Ultra Centrifuge Filters -30K	Millipore Sigma	UFC503096
Amico Ultra Centrifuge Filters -100K	Millipore Sigma	UFC510024

Table 3.S.11 continued from previous page

REAGENT or RESOURCE	SOURCE	IDENTIFIER
QIAseq FastSelect	QIAGEN	334376
NEBNext RNA First-Strand Synthesis	New England BioLabs	E7525
NEBNext Ultra Directional RNA Second Strand Synthesis Modules	New England BioLabs	E7550
NEBNext Ultra II DNA Library Prep Kit	New England BioLabs	E7645
Quant-iT PicoGreen dsDNA Assay Kit	Thermo Fisher Scientific	P7589
SparQ beads	QuantaBio	95196-060
QIAquick PCR Purification Kit	QIAGEN	28106
Plasmids		
Human ARMC5-HA	GeneCopeia	EX-H0661-M07
Human ARCM5-FLAG	GeneCopeia	EX-H0661-M14
Human Cullin3-Myc-DDK	OriGene	RC208066
Human FLAG-Pol-II WT	Addgene	Plasmid #35175
Human FLAG-Pol-IIΔ	Addgene	Plasmid #35176
HA-Ubiquitin	Addgene	Plasmid #18712
Human ARMC5 (Δaa2-142) -HA	This study	N/A
Human ARMC5 (Δaa143-444) -HA	This study	N/A
Human ARMC5 (Δaa445-747) -HA	This study	N/A
Human ARMC5 (Δaa748-816) -HA	This study	N/A
Human ARMC5 (Δaa817-934) -HA	This study	N/A
Human Cullin3-HA	This study	N/A
Human Cullin3(Δaa31-385)-HA	This study	N/A
Human Cullin3 (Δaa377-675)-Myc-DDK	This study	N/A
Human Cullin3 (Δaa695-762)-Myc-DDK	This study	N/A
Human ARMC5-BTB-HA	This study	N/A
Human Cullin3 (aa1-376)-Myc-DDK	This study	N/A
Softwares		
Scaffold 4.3.0	Proteome software	`http://www.proteomesoftware.com/products/scaffold`
PEAKS 8.5	Bioinformatics Solutions	`https://www.bioinfor.com`
AlphaFold	(Jumper, J et al., 2021)	`https://alphafold.ebi.ac.uk`
UCSF Chimera	(Pettersen, et al., 2004)	`https://www.cgl.ucsf.edu/chimera`
AxioVision	Carl Zeiss	N/A
Trimmomatic	(Bolger et al., 2014)	`http://www.usadellab.org/cms/?page=trimmomatic`
STAR	(Dobin et al., 2013)	`https://github.com/alexdobin/STAR`
Picard	("Picard Toolkit." 2019.)	`https://broadinstitute.github.io/picard`

Table 3.S.11 continued from previous page

REAGENT or RESOURCE	SOURCE	IDENTIFIER
StringTie	(Kovaka et al., 2019)	`https://ccb.jhu.edu/software/stringtie`
EdgeR	(Robinson et al., 2010)	`https://bioconductor.org/packages/release/bioc/html/edgeR.html`
ggplot2	(Wickham et al., 2016)	`https://ggplot2.tidyverse.org`
Raincloud Plots	(Allen et al., 2021)	`https://github.com/RainCloudPlots/RainCloudPlots`
R circlize v0.4.9	(Gu et al., 2014)	`https://jokergoo.github.io/circlize_book/book/`
BWA-MEM	(Li and Durbin, 2009)	`https://github.com/bwa-mem2/bwa-mem2`
MACS2	(Zhang et al., 2008)	`https://hbctraining.github.io/Intro-to-ChIPseq/lessons/05_peak_calling_macs.html`
HOMER	(Heinz et al., 2010)	`http://homer.ucsd.edu/homer`
DESeq2	(Love et al., 2014)	`https://bioconductor.org/packages/release/bioc/html/DESeq2.html`
Deposited data		
RNA-seq	Accession number: GSE169263	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169263`
ChIP-Seq	Accession number: GSE169578	`https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE169578`
Experimental models: cell line		
HEK-293 (ATCC CRL-1573)	ATCC	CRL-1573
SW-13 (ATCC CCL-105)	ATCC	CCL-105

Chapter 4

ARMC5 is part of a novel POLR2A-specific ubiquitin ligase E3 and its mutation is a modifier of neural tube defect risks

[*][#][1]Hongyu Luo, [*][1]Linjiang Lao, [2]Kit Sing Au, [2]Hope Northrup, [1]Xiao He, [3]Diane Forget, [3]Marie-Soleil Gauthier, [3,4]Benoit Coulombe, [1,5,6]Isabelle Bourdeau, [1]Wei Shi, [7-10] Lucia Gagliardi, [11]Maria Candida Barisson Villares Fragoso [1]Junzheng Peng, and [#][1,6,12] Jiangping Wu

From the [1]Centre de recherche and [5]Division of Endocrinology and [12]Division of Nephrology, Centre hospitalier de l'Université de Montréal (CHUM), Montreal, Quebec, Canada;[4]Department of Biochemistry and Molecular Medicine and [6]Department of Medicine Université de Montréal, Montreal, Quebec, Canada; [2]Department of Pediatrics, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, USA; [3]Department of Translational Proteomics, Institut de Recherches Cliniques de Montréal, Montreal, Quebec, Canada; [4]Department of Biochemistry and Molecular Medicine, Université de Montréal, Montreal, Quebec, Canada; [7]Adelaide Medical School, University of Adelaide, Adelaide, Australia; [8]Endocrine and Metabolic Unit, Royal Adelaide Hospital, Adelaide, Australia; [9]Department of Genetics and Molecular Pathology, SA Pathology, Adelaide, Australia; [10]Endocrine and Diabetes Unit, The Queen Elizabeth Hospital, Adelaide, Australia; [11]Unidade de Suprarrenal Disciplina de Endocrinologia e Metabologia, Hospital das Clínicas, Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil

[#]Address correspondence to Dr. Hongyu Luo, CRCHUM, 900 Saint-Denis Street, Room R12.426, Montreal, Quebec, Canada H2X 0A9. Telephone: (514) 890-8000 Extension 25394, Fax: (514) 412-7944, e-mail: hongyu.luo@umontreal.ca; or Dr. Jiangping Wu, CRCHUM, 900 Saint-Denis Street, Room R12.428, Montreal, Quebec, Canada H2X 0A9. Telephone: (514) 890-8000 Extension 25164, Fax: (514) 412-7944, e-mail: jianping.wu@umontreal.ca

This work is under reviewing in *EMBO Reports*.

4.1. Abstract

ARMC5 is a protein containing an armadillo domain (ARM) and a BTB (Broad-complex, tramtrack, and bric-à-brac) domain. Its gene knockout in mice caused many phenotypes, including dwarfism, compromised T-cell immunity, and adrenal gland hyperplasia. *ARMC5* mutation in humans is associated with bilateral macronodular adrenal gland hyperplasia. We found that *Armc5* KO mice had an increased incidence of neural tube defects (NTDs). Whole-exome sequencing of 511 myelomeningocele (MM) patients revealed nine highly deleterious mutations in the ARMC5 coding sequence, showing the relevance of our finding in mice to human NTD. We revealed that ARMC5 complexed with CUL3 and POLR2A and was part of a novel dominant POLR2A-specific ubiquitin ligase (E3) under a physiological condition. A deleterious mutation p.Arg429Cys found in MM patients drastically weakened the interaction between ARMC5 and POLR2A, which likely diminishes the function of this E3 and provides its pathogenic mechanism at the molecular level. *Armc5* gene knockout (KO) caused diminished POLR2A ubiquitination and compromised POLR2A degradation via proteasomes. Surprisingly, the absence of this E3 did not lead to generalized Pol II stalling and the subsequent generalized decrease of mRNA transcription. The lack of the E3 dysregulated 108 genes in neural precursor cells (NPCs), some of which are involved in processes critical to neural development. *Armc5* KO in the intestine downregulated FOLH1 (folate hydrolase 1) expression, which is essential in folate absorption. Our results indicate that this novel ARMC5-CUL3-RBX1 E3 plays a critical role in Pol II pool homeostasis, and *ARMC5* mutation is a modifier of NTD risks in mice and humans.

Keywords: ARMC5, ubiquitin ligase, POLR2A, POLR2A-specific E3, neural tube defects, myelomeningocele, FOLH1, Pol II pool size, Cullin3

4.2. Introduction

During embryonic development, the neural plate folds into a neural tube to form the future brain and spinal column. The process occurs during days 17 – 28 of human gestation or e8.5 – 10.5 (embryonic days) in mice[410,411]. Defective closure of neural tube

causes a collection of manifestations, such as spina bifida, anencephaly, encephalocele, and iniencephaly. The three latter forms of neural tube defects (NTDs) are severe and frequently result in miscarriage or stillbirth. There are three principal types of spina bifida: meningocele, myelomeningocele (MM), and spina bifida occulta[412]. The latter has no or very mild symptoms. NTD is one of the most common birth defects, with a prevalence of 18.6/10,000 live births worldwide[413]. During the process of neural tube closing, the neural plate cells need to undergo the necessary proliferation, differentiation, shape change, and migration[414,415,416,417], and these steps can be affected by genetic and environmental factors, which all contribute to NTD risks[418]. It has been well established that sufficient maternal dietary folate intake is essential for proper neural tube closing[419]. The genetic contribution to NTD is polygenic. Based on candidate gene study approaches, some genes in the folate metabolic pathway have been found to be associated with NTD risks in humans[420,421,422]. Mutations/deletions in more than 200 genes are known to cause NTD in mice, although most of them are yet to be confirmed in humans.

ARMC5 is a protein containing an armadillo domain, which consists of seven armadillo repeats. Each repeat is about 40 amino acids (aa) long and consists of three α-helices[290]. Human and mouse ARMC5 proteins share ∼90% aa sequence homology and have similar structures. Both have the Armadillo domain towards their N-terminus and a BTB (Broad-complex, tramtrack, and bric-à-brac) domain towards their C-terminus[318,317,319]. The dominant human ARMC5 protein isoform is 935 aa in length (NP_001098717.1)[349], and mouse ARMC5, 926 aa (NP_666317.2). In humans, at least four other transcript isoforms derived from the same gene exist, but most of them vary at the 5'- and 3'-regions. However, one uses two extra exons at the 5' end of the gene and translates into a longer protein isoform of 1030 aa in length (NP_001275696.1)[349]. ARMC5 does not have any enzymatic activity, and its mechanism of action is unknown.

We recently reported the phenotype of *Armc5* gene knockout (KO) mice. The KO mice were smaller in size from the fetal stage until old age and were born below the expected Mendelian ratio from heterozygous parents[8]. The function of T lymphocytes of the KO

mice was compromised in that they had reduced proliferation and differentiation *in vitro*, and reduced autoimmune responses, and defective viral clearance[8]. At old age, the KO mice presented adrenal gland hypertrophy[8], similar to primary bilateral macronodular adrenal gland hyperplasia (PBMAH). Approximately $21 - 26\%$ of PBMAH patients carry *ARMC5* mutations[18,1,2,14].

DNA-directed RNA polymerase II (Pol II) is responsible for the transcription of all the mRNA, and most small nuclear RNA, microRNA and long non-coding RNA[334,423,424]. Pol II is highly conserved. Human and mouse Pol II both have 12 subunits[425]. POLR2A is the largest and catalytic subunit. During mRNA transcription, Pol II might pause due to various reasons, such as template DNA damage, cell stress, or gene activation status. It will continue its journey along the template DNA once these adverse conditions are resolved. Permanent Pol II stalling will block the transcription, and the stalled Pol II needs to be removed to resume transcription. It is believed that ubiquitination followed by proteasome degradation is a process to remove the stalled Pol II[341,340,342]. The proteasome-mediated Pol II degradation is also critical for Pol II pool size homeostasis. Whether an abnormal Pol II pool size affects all genes or just a subset of genes, and the pathogenic roles of the abnormal Pol II pool size, are an understudied area. Recently, Vidakovic et al. and Nakazawa et al. reported that Lys1268 ubiquitination is necessary and sufficient for POLR2A degradation in cells after UV irradiation. Lys1268Arg mutation prevents POLR2A ubiquitination, resulting in POLR2A accumulation and an enlarged Pol II pool size in cells with irradiation-induced massive DNA damage[263,264]. The enlarged Pol II pool in the mutant cells in this model selectively leads to faster transcription recovery of short genes and upregulates a subset of genes (about 1600 genes)[264]. These studies clearly demonstrate that the effect of Pol II pool size is not universal to all genes in the irradiated cells, although we do not know whether the same is true under a physiological condition without massive DNA damage.

Ubiquitination is involved in protein degradation and function. Protein ubiquitination is catalyzed by a cascade of enzymes, i.e., E1 (Ub-activating enzyme), E2 (Ub-conjugating enzyme), and E3 (Ub ligase)[344]. The specificity of the cascade is determined by E3, which

has three families: Ring-finger (single or multiple subunits), HECT, and RBR[346]. The Ring-finger E3s are the largest family. A multiple subunit Ring-finger E3 contains a RING-finger protein (e.g., RBX1 or RBX2), a cullin (CUL) protein (CUL1, 2, 3, 4A, 4B, 5, and 7), and a substrate recognition unit[347].

Due to Pol II's and its largest subunit POLR2A's central role in cell biology, POLR2A-specific E3 is of vital interest to cell biologists. Several such E3s have been reported before, but most of them only have convincing activities in cultured cells after irradiation- or drug-induced massive DNA damage[275,268,426,194,427,285]. Two of these E3s do have activity in unmanipulated cell lines, but such observation has not extended to tissues and organs[194,279]. A POLR2A-specific E3 that functions under physiological conditions in organs and tissues, although such an E3 should exist in theory for Pol II pool size homeostasis and for removing stalled Pol II in the absence of artificially induced DNA damage.

In the present work, we revealed that the KO mice had an increased incidence of NTD. A human genetic study discovered nine highly deleterious variants in the ARMC5 coding sequence of myelomeningocele (MM) patients. Our results showed that ARMC5 interacted with CUL3, and the former was the substrate recognition subunit of POLR2A-specific E3, which was essential for POLR2A degradation via the proteasomes and Pol II pool size homeostasis under a physiological condition. ARMC5 deletion compromised POLR2A ubiquitination and caused its accumulation, hence an enlarged Pol II pool. However, failed POLR2A degradation did not result in generalized Pol II stalling nor generalized transcription depression. The abnormally large Pol II pool in the KO tissues dysregulated the transcription of some genes, such as *Folh1*, *Cdkn1*, *Gadd45b*, *Mafa*, and *Pcdh8* that are critical in folate absorption, cell cycling, and growth. These processes are vital in neural tube development.

4.3. Results

4.3.1. Increased incidence of NTD in *Armc5* KO mice

In situ hybridization revealed that *Armc5* was highly expressed in the mouse e10 neural tube (**Figure 4.1a**). In addition to smaller body sizes as reported before[8], live-born

Armc5 KO mice in the CD1 × C57BL/6 F1 background presented significantly high incidences of kinky tails upon visual inspection (**Figure 4.1b**) or micro-CT imaging of skeletons (**Figure 4.1c**). Kinky tails were observed in 27.7% of the live-born KO mice, compared to 3.7% of the wild-type (WT) counterparts (**Figure 4.1d**). The incidence rates did not show any apparent sex bias, with 27.6% and 27.9% in male and female KO mice, respectively (**Figure 4.1d**).

The KO mice from heterozygous male and female parents were born below the Mendelian ratio. Only about 10% of KO pups were born alive, instead of the expected 25%. This suggests a certain degree of embryonic lethality of KO fetuses. Indeed, about 14.9% of KO fetuses but none of WT fetuses manifested anencephaly, a more severe form of NTD (**Figure 4.1e and f**). The total incidence rate of NTD in this genetic background (live-born mice with kinky tails plus fetuses with anencephaly) amounted to 43% in the KO versus 4% in the WT mice.

4.3.2. Decreased proliferation and increased apoptosis of cells in KO fetal neural tubes

Cell proliferation in e9.5 KO and WT fetal neural tubes at the rostral hindbrain level was assessed by Ki-67 immunofluorescence staining (**Figure 4.2a**). The KO neural tubes showed a significantly reduced percentage of Ki-67-positive proliferating cells (pseudo-green) compared to their WT counterparts.

The apoptosis of cells in the e9.5 neural tubes was evaluated using the fluorescent terminal deoxynucleotidyl transferase dUTP nick end-labeling (TUNEL) assay (**Figure 4.2b**). The TUNEL-positive cells among total neural tube cells in the areas viewed were quantified (lower panel). The number of TUNEL-positive apoptotic cells (pseudo-green) at the rostral hindbrain level was significantly increased in the KO neural tube.

Neural stem cells and neural precursor cells are involved in neural tube development[428,429]. They are hereafter called neural precursor cells (NPCs) in this work. We isolated these cells from the KO and WT central nervous system and expanded them *in vitro* for 10 – 12
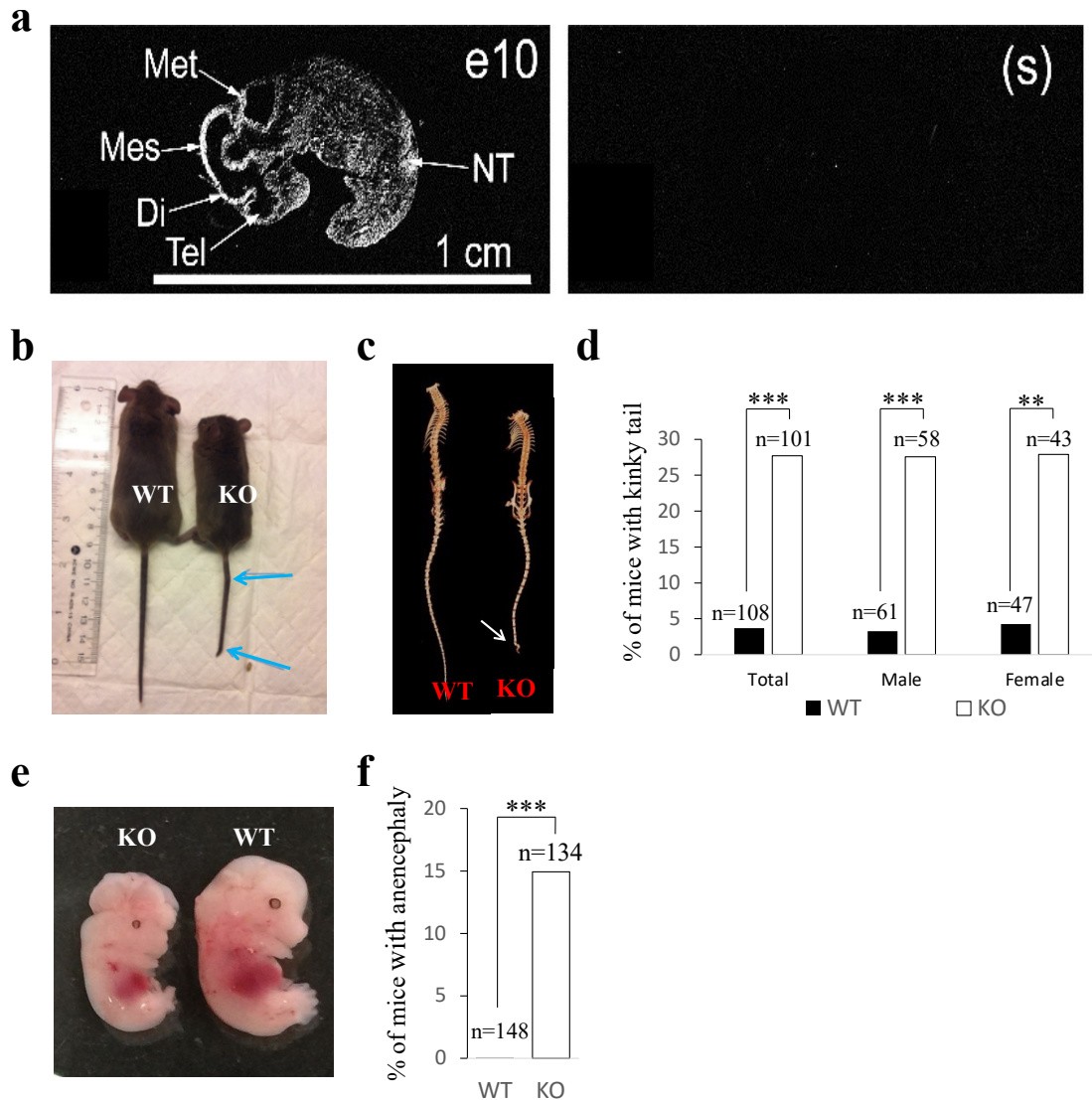
Fig. 4.1. High incidence of NTD in *Armc5* KO mice (a). High *ARMC5* expression in the neural tube of an e10 WT C57BL/6 fetus according to X-ray film autoradiography of *in situ* hybridization. NT: rostral neural tube; Met: metencephalon; Mes: mesencephalon; Di: diencephalon; Tel: telencephalon. S: sense. **(b).** A KO mouse with a kinky tail (arrows: kinks in the tail). **(c).** Skeletons of WT and KO mice according to micro-CT scan. Arrow: the curled tail in a KO mouse. **(d).** Increased incidence of KO mice with kinky tails. The numbers of male, female and total mice examined are indicated. **(e).** An e12.5 KO fetus with anencephaly. **(f).** High incidence of anencephaly in KO fetuses. The numbers of fetuses examined on e12.5 are indicated. ***: p $<$0.001; **: p $<$0.01(χ^2 test).

days. The purity of these cells was about 85%, according to SOX2 and NESTIN staining (**Figure 4.2c**). The KO NPCs proliferated significantly slower than WT ones at different

input cell numbers (**Figure 4.2d**) or at a constant input cell number but different epidermal growth factor (EGF) concentrations (**Figure 4.2e**).

We also measured the apoptosis of KO and WT NPCs cultured in different EGF concentrations, using annexin V staining followed by flow cytometry. Both KO and WT NPCs manifested an increased apoptosis rate inversely correlated to EGF concentrations, but KO NPCs presented a significantly higher degree of apoptosis (**Figure 4.2f**).

4.3.3. *ARMC5* mutations were risk factors for human NTD

We next assessed whether *ARMC5* mutation was relevant to human NTD using a cohort of MM patients, MM being a severe form of NTD. Single nucleotide variants (SNV) in *ARMC5* transcripts of 511 MM subjects were assessed by whole-exome sequencing. Among the 511 MM subjects, 257 were Americans of European descent, and 254 were Mexican Americans. The control populations of Non-Finnish Europeans and Ad Mixed Americans in the genome aggregation database (gnomAD) were used as reference controls. These controls were not selected for or against MM.

The MM subjects'age, alternate SNV position and protein mutation, alternate allele counts, allele number, and allele frequency are shown in **Table 4.1**. The allele numbers varied for different single nucleotide variants (SNVs) in both MM subjects and controls due to differences in exon capturing techniques and batch effect. A larger allele number variation occurred in gnomAD controls because the data were compiled from multiple projects.

The deleteriousness of the SNVs in *ARMC5* of the MM subjects was calculated according to Combined Annotation Dependent Depletion Phred score (C-score)[430]. The top 5% most deleterious SNVs (i.e., with C-score > 13.01) found in *ARMC5* transcripts of MM subjects are listed in **Table 4.1**. Their positions in two ARMC5 isoforms are illustrated in **Figure 4.3a**. Both isoforms are coded by the same *ARMC5* gene. The 935-aa isoform (NP_001098717.1) is the most abundant one presented in almost all tissues[349]. The 1030-aa isoform (NP_001275696.1) is the longest and has an N-terminal 95-aa region coded by two extra exons[349]. These nine SNVs were all missense variants. It is to be noted that the
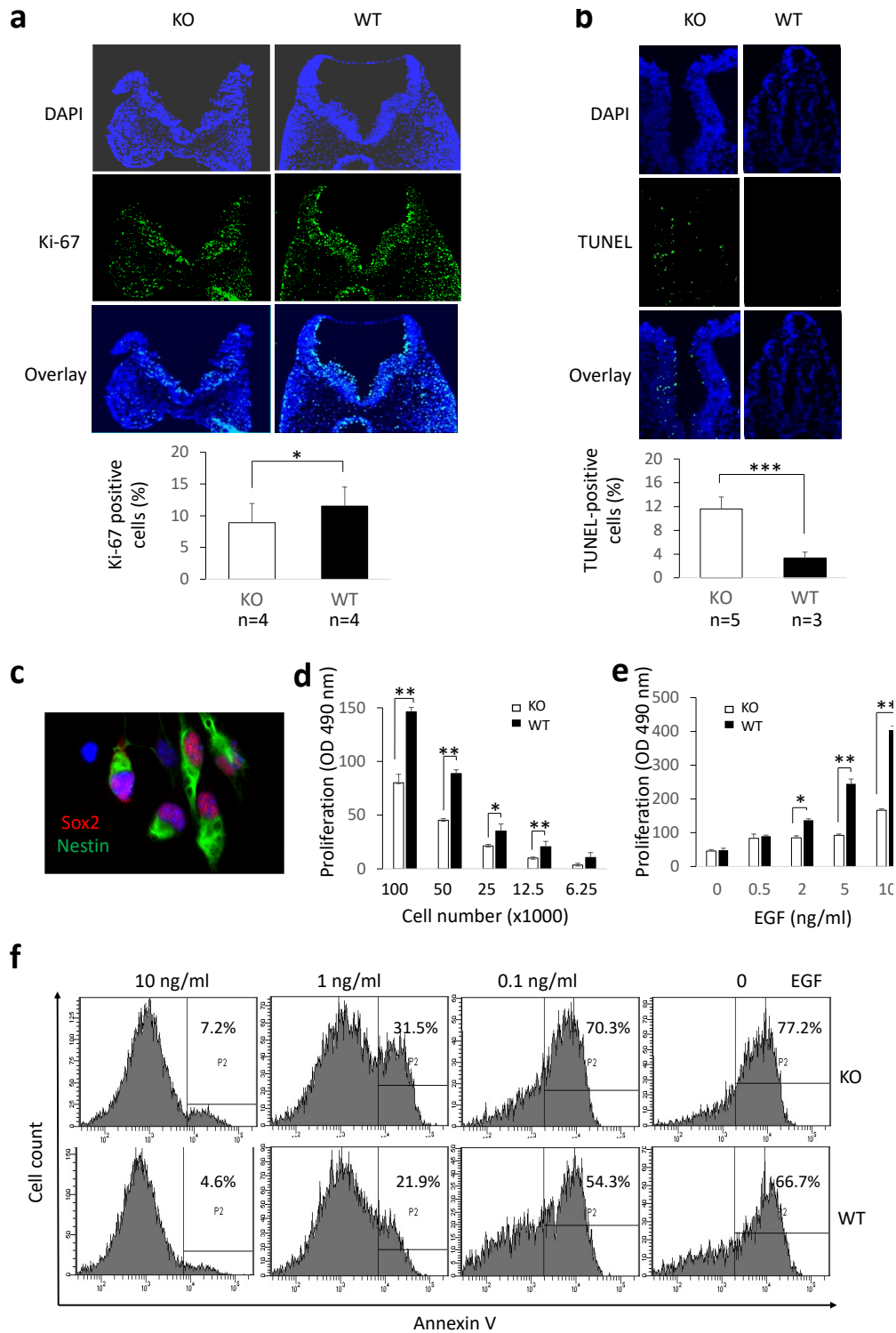
Fig. 4.2. The reduced proliferation and increased apoptosis of cells in KO neural tubes and NPCs *(see next page)*

Fig. 4.2 *(previous page).* **(a).** Reduced cell proliferation in KO neural tubes. KO and WT neural tubes at the level of the rostral hindbrain (sectioned transversally) on e9.5 were stained with a proliferation marker Ki-67 (upper panel). The percentage (means ± SD) of Ki-67-positive cells among total cells in randomly selected view areas in the neural folds is presented in a bar graph (lower panel). Results were based on the counting of 3 different sections per fetus and four fetuses per group (KO or WT). P-value was indicated (paired two-way Student's t-test). **(b).** Increased cell apoptosis in KO neural tubes. Cell apoptosis in KO and WT neural tubes was determined by TUNEL (upper panel). The percentage (means ± SD) of TUNEL-positive cells among total cells in randomly selected view areas in the neural folds is presented in a bar graph (lower panel). Results were based on the counting of three different sections per fetus (KO: 5 fetuses; WT: 3 fetuses). P-values were indicated (paired two-way Student's t-test). **(c).** Characterization of NPCs. NPCs were stained with NPC markers Sox2 and Nestin. **(d)** and **(e).** Reduced KO NPC proliferation. Different numbers of NPCs, as indicated, were cultured in the presence of a fixed concentration of EGF (20 ng/ml) **(d)**, or a fixed number of NPCs were cultured in the presence of different concentrations of EGF as indicated **(e)**. After 72 hours, proliferation was measured by an MTS-based CellTiter96 AQueous Assay. Samples were in triplicate, and means ± SD of OD490 nm (after subtracting background absorbance based on OD490 nm of wells without cells) of a representative experiment out of four independent ones are shown. *: $p < 0.05$ (Two-way Student's t-test). **(f).** Augmented apoptosis of KO NPCs cultured in the presence of different concentrations of EGF. NPCs were cultured for 20 hours, and their apoptosis was measure by annexin V staining followed by flow cytometry. Histograms of a representative experiment out of four repetitions are shown.

mutations in any isoform can be presented in the **Table 4.1**. Using the longest isoform for the numbering purpose does not mean that these SNVs only exist in this longest isoform. In all likelihood, most of these SNVs, except three in the first 95-aa region in the N-terminus **Figure 4.3a**, are in the most abundant 935-aa isoform.

These nine SNVs were rare ones (defined as having allele frequency < 0.01). There were four and five rare missense variants found in European American and Mexican American MM subjects, respectively. Two of the four rare variants found in European American MM subjects and four of the five rare variants found in Mexican American MM subjects were assigned as the top 1% deleterious variants (C-scores > 20). Two variants (p.Thr12Ala rs979451735 and p.Arg429Cys rs539440145) found in European American MM subjects were not present in gnomAD non-Finnish European controls, and they were considered as novel SNVs. Their alternate allele counts were significantly higher than that of the Non-Finnish

European controls (p < 0.05). The p-value of the alternate allele count of one rare SNV found in Mexican American MM subjects was approaching significant (p = 0.083). Since all variants identified in the approximately 9-kb *ARMC5* loci were in linkage disequilibrium, these p-values were not subjected to a multiple-testing penalty.

Table 4.1. *ARMC5* alternate SNVs in myolomeningocele subjects

Subject code	Gender	Birth year	Mutation ID	Protein mutation	rsID (db-SNP151)	Adjusted AC (EUR)	Adjusted AN (EUR)	AF (EUR)	Controls AC (NFE)	Controls AN (NFE)	Controls AF (NFE)	OR	*p-value*	CADD phred
B39-407	M	1998	16:31469751:A:G	p.Thr12Ala	rs979451735	1	510	0.20%	0	18.978	0.00%	∞	0.026	0.1033
BC42-1019	F	1960	16:31470942:C:T	p.Pro128Ser	rs200309618	1	450	0.22%	32	41.112	0.08%	2.86	0.302	0.23056
B34-267	M	1993	16:31473868:C:T	p.Arg429Cys	rs539440145	1	494	0.20%	0	42.506	0.00%	∞	0.011	0.58774
E75-466	F	1985	16:31477780:G:A	p.Arg888Gln	rs199498431	1	470	0.21%	19	42.368	0.04%	4.75	0.198	0.43285
A23-456	F	1991	16:31470793:G:A	p.Arg78His	rs920446902	1	500	0.20%	2	13.094	0.02%	13.12	0.106	0.80969
D68-941	M	2004	16:31470811:C:T	p.Ala84Val	rs202112554	1	500	0.20%	13	14.204	0.09%	2.19	0.384	0.10959
D46-697	M	1989	16:31474085:G:A	p.Arg501Gln	rs749775865	1	500	0.20%	3	17.048	0.02%	11.39	0.109	0.28886
C84-353	F	1998	16:31474132:G:A	p.Gly517Ser	rs372472557	1	498	0.20%	2	17.048	0.01%	17.15	0.083	0.45764
F78-368	M	1986	16:31476020:C:T	p.Pro654Leu	rs200115942	1	488	0.20%	7	16.78	0.04%	4.92	0.205	0.26627

Note: A total of 511 subjects from North America were selected for whole-exon sequencing. Rare variants found in ARMC5 are presented. Protein mutation is presented with amino acid position numbered according to ARMC5 isoform c (ENSP00000386125 or NP_001275696) with 1030 aa. Mutations of amino acids at positions 128, 429, 501, 517, and 654 are present in all ARMC5 isoforms. Amino acid at position 888 is present in three ARMC5 isoforms with a longer coding sequence, including the isoform with 935 aa (ENST00000268314.4 or NP_001098717). Amino acids positions 12, 78, and 84 are not present in the isoform with 935 aa (ENST00000268314.4 or NP_001098717) or the shortest isoform with 725 aa (ENSP00000399561 or NP_079018). EUR: European descent; Mex: Mexican American; NFE: non-Finnish European; AMR: Ad Mixed American; AC, allele count; AN: allele number; AF: allele frequency; OR: odds ratio; CADD Phred: Combined Annotation Dependent Depletion Phred score (C-score).

4.3.4. ARMC5 physically interacts with CUL3 and POLR2A

We previously conducted a yeast-2-hybrid (Y2H) assay to identify ARMC5-binding proteins. Seventeen significant hits were obtained, and CUL3 and POLR2A were among the top six in the list[8]. To validate the findings of Y2H, we transfected HEK293 cells with FLAG-tagged ARMC5-expressing plasmids. POLR2A and CUL3 were significantly associated with ARMC5 (false discovery rate (FDR) < 0.05 and fold change (FC) > 2) according to anti-FLAG Ab immunoprecipitation followed by liquid chromatography with tandem mass spectrometry (LC-MS/MS) (**Figure 4.3b**).

Additional validation of the interaction among ARMC5, POLR2A, and CUL3 was carried out employing immunoprecipitation followed by immunoblotting in neuronal cells, which were more relevant to NTD than HEK293 cells. SK-N-SH human neuronal cells were transfected with plasmids expressing human ARMC5-HA. The cell lysates were precipitated with

anti-HA Ab and then immunoblotted with either anti-CUL3 or anti-POLR2A Abs. Endogenous POLR2A and CUL3 were detected in the precipitates, confirming that ARMC5 indeed physically interacted with these two molecules (**Figure 4.3c**). In the ARMC5 immunoblotting (**Figure 4.3c and d**), there were always two prominent bands, one at 130 kD and the other at 100 kD. This smaller band's intensity varied in different experiments, suggesting that this was a cleavage product of the full-length larger 130 kD ARMC5, but not an isoform initiated from a downstream start codon during translation.

ARMC5 p.Arg429Cys mutation (position based on the 1030-aa isoform; corresponding to p.Arg334Cys in the 935-aa isoform) was significantly associated with MM. This mutation was in the 5th repeat in the ARM domain (**Figure 4.3a**). When HA-ARMC5(WT) or HA-ARMC5(p.Arg429Cys) mutant was transfected into HEK293 cells, the latter co-precipitated significantly less POLR2A, according to immunoblotting (**Figure 4.3d**), indicating that this mutation hampered the interaction between ARMC5 and POLR2A. In PBMAH patients, ARMC5 p.Arg315Trp mutation (position based on the 935-aa isoform) was significantly associated with the disease[16]. Arg315 is also located in the 5th repeat in the ARM domain (**Figure 4.3a**). We transfected FLAG-ARMC5(WT) or FLAG-ARMC5(p.Arg315Trp) mutants into HEK293 cells. LC-MS/MS analysis of the FLAG-precipitates showed that the association of POLR2A to the ARMC5 p.Arg315Trp mutant was about four-fold lower than that of WT ARMC5 (FDR < 0.05) (**Figure 4.3e**). These results indicated that the ARM domain, particularly the 5th repeat in the domain, was extremely important for the interaction between ARMC5 and POLR2A. The two clinically relevant ARMC5 mutations, i.e., p.Arg429Cys (p.Arg334Cys according to the 935-aa isoform) in MM and Arg315Trp (according to the 935-aa isoform), in PBMAH patients, likely exerted their deleterious function by reducing their POLR2A-binding capability. It is interesting to note that in addition to POLR2A, the binding of the ARMC5 p.Arg315Trp mutant to other components of Pol II such as POLR2B, PLOR2C, and POLR2K was also reduced (**Figure 4.3e**). This indicates that WT ARMC5 binds to these Pol II subunits (whether directly or via POLR2A is yet to be determined), but the ARMC5 Arg315Trp mutant is less capable of doing so.

4.3.5. ARMC5 KO resulted in failed degradation of POLR2A

CUL3 is often part of a multiple-subunit RING-finger E3 complex, in which CUL3 interacts with a RING-finger protein RBX1[431]. In such complexes, CUL3 also interacts with a BTB domain-containing protein, which serves as an E3 substrate recognition subunit[348]. Since ARMC5 contains a BTB domain towards its C-terminus and interacts with both
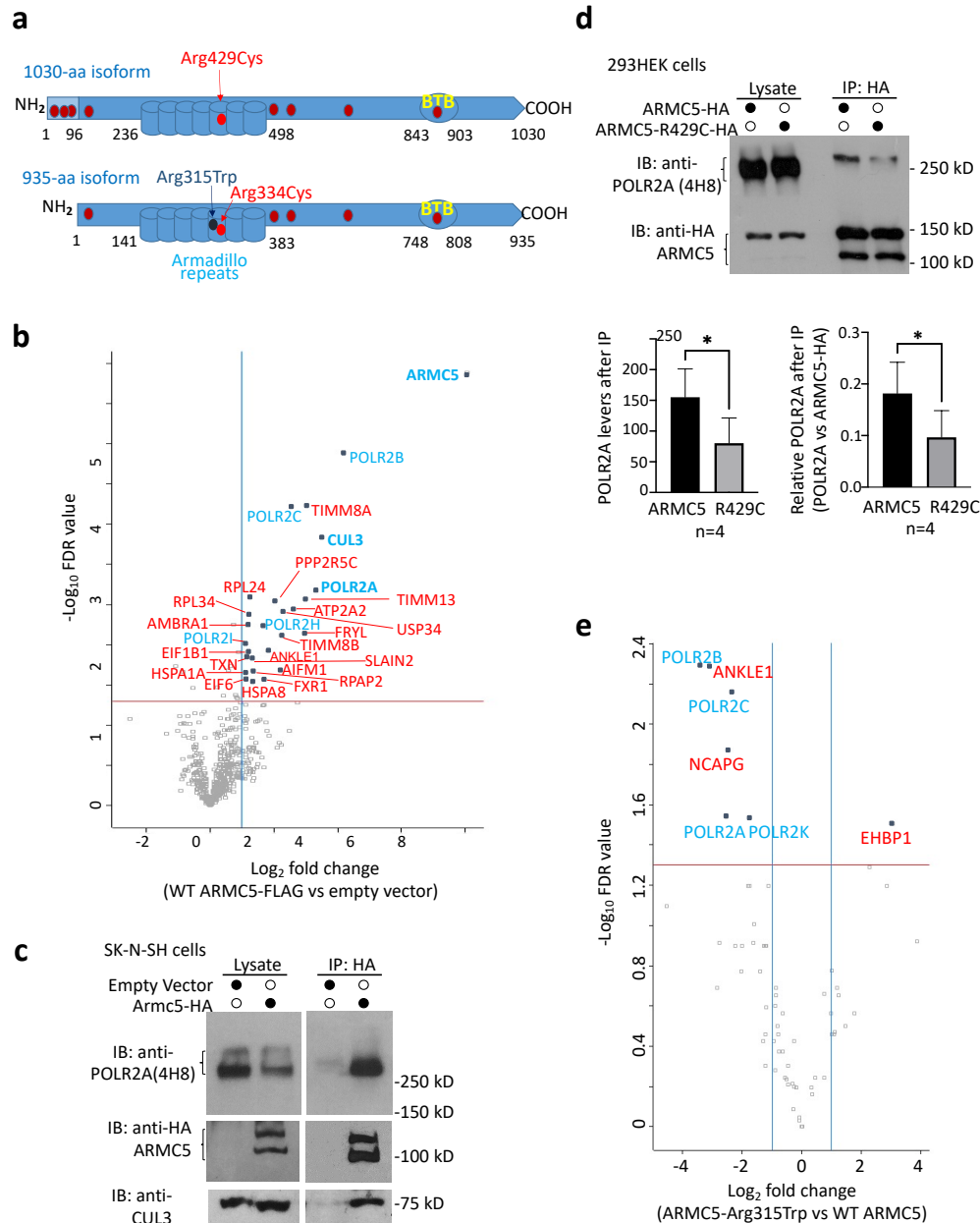
Fig. 4.3. ARMC5 interacted with CUL3 and POLR2A. *(see next page)*

Fig. 4.3 *(previous page).* **(a).** A schematic illustrates the ARMC5 protein structure. The upper schematic shows the longest 1030-aa ARMC5 isoform, which has 95 extra aa in the N-terminus. The red dot: the Arg429Cys mutation (numbered based on the 1030-aa isoform) found in the MM cohort. The maroon dots: the other eight missense mutations found in the MM cohort. The lower schematic shows the most abundant 935-aa ARMC5 isoform. The black dot: Arg315Trp mutation (numbered based on the 930-aa isoform) found in PBMAH patients. The red dot and maroon dots: the Arg334Cys and other mutations (numbered based on the 930-aa isoform) found in the MM cohort. **(b).** A volcano plot showing ARMC5 interacted with endogenous CUL3 and POLR2A in HEK293 cells according to immunoprecipitation followed by LC-MS/MS. HEK293 cells were transfected with ARMC5-FLAG-expressing plasmids. The anti-FLAG precipitates were analyzed by LC-MS/MS. The vertical line indicates 2-fold changes, and the horizontal line, FDR = 0.05, based on three biological replicates. **(c).** ARMC5 interacts with endogenous CUL3 and POLR2A in SK-N-SH neuronal cells according to immunoprecipitation followed by immunoblotting. SK-N-SH cells were transfected with a human ARMC5-HA-expressing plasmid or an empty vector. ARMC5 in the lysates was immunoprecipitated by anti-HA Ab. The presence of endogenous POLR2A (upper right panel) and CUL3 (lower right panel) in the immunoprecipitates was revealed by immunoblotting with anti-POLR2A and anti-CUL3 Abs. The presence of ARMC5-HA in the cell lysates and immunoprecipitation products was confirmed by immunoblotting with anti-HA Ab (middle panel). The endogenous POLR2A and CUL3 in the empty vector- or ARMC5-expressing plasmid-transfected cell lysates were revealed by immunoblotting (upper left and lower left panels, respectively). The experiments were conducted more than three times, and representative results are shown. **(d).** Reduced binding between the ARMC5 Arg429Cys mutant and POLR2A in HEK293 cells according to immunoprecipitation followed by immunoblotting. The lower panels show densitometry results based on four independent experiments: lower left panel, POLR2A signals without normalization; lower right panel, POLR2A signals normalized with ARMC5-HA signals. *: $p < 0.05$ (paired two-way Student's t-tests). **(e).** A volcano plot showing reduced binding between the ARMC5 Arg315Trp mutant and POLR2A in HEK293 cells according to immunoprecipitation followed by LC-MS/MS. The horizontal line: FRD = 0.05 based on three biological replicates. The vertical lines: +2- and -2-fold changes.

CUL3 and POLR2A, we hypothesize that it was the substrate recognition subunit of a novel POLR2A-specific E3.

One of the consequences of protein ubiquitination, particularly K48-linked ubiquitination, is to channel substrate proteins to the proteasome for degradation[344]. E3 controls the substrate specificity of the cascade of enzymes involved in protein ubiquitination. If ARMC5-CUL3-RBX1 was a POLR2A-specific E3, we might observe an accumulation of

POLR2A protein in the ARMC5 KO tissues. The POLR2A C-terminal domain has different phosphorylation levels at its S2 or S5 residues for Pol IIs at different gene regions[432,192]. We assessed POLR2A protein levels in KO tissues and cells relevant to NTD, using Abs recognizing POLR2A of different phosphorylation statuses. The levels of hyper- and hypo-phosphorylated POLR2A proteins (identified by mAb 4H8) in KO e9.5 neural tubes and NPCs were drastically elevated compared to those in their WT counterparts (**Figure 4.4a**). Using mAb specific to POLR2A N-terminus (representing total POLR2A), phosphorylated C-terminal domain S2, or phosphorylated C-terminal domain S5, we demonstrated that POLR2A of these different phosphorylation statutes in the KO neural tubes and NPCs was all increased (**Figure 4.4c-d**). This suggests that POLR2A at different gene regions during the transcription process are all accumulated, reflecting the generally decreased POLR2A degradation. As Pol IIs constantly recycle during transcription, an enlarged Pol II pool due to compromised POLR2A degradation will obviously lead to more abundant Pol IIs in all the transcription stages in different gene regions.

Polr2a mRNA levels in the KO and WT neural tubes (**Figure 4.4e**) were similar. For NPCs, its *Polr2a* mRNA levels were even moderately decreased (**Figure 4.4f**), probably due to the direct or indirect effect of a larger Pol II pool. At any rate, these results show that POLR2A protein accumulation in KO cells is a post-transcriptional event.

4.3.6. ARMC5 KO resulted in compromised POLR2A ubiquitination

We next investigated whether there was reduced POLR2A ubiquitination in the KO cells. The total ubiquitination of POLR2A with high and low phosphorylation (recognized by mAb 4H8) was significantly reduced in KO NPCs (**Figure 4.5a**). Similarly, the K48-linked ubiquitination of total POLR2A (recognized by mAb F12) in KO NPCs was reduced (**Figure 4.5b**). So was the K48-linked POLR2A with high and low phosphorylation (recognized by mAb 4H8; (**Figure 4.5c**). The K63-linked ubiquitination of total POLR2A (recognized by mAb F12) was also reduced in KO NPCs (**Figure 4.5d**). In the presence
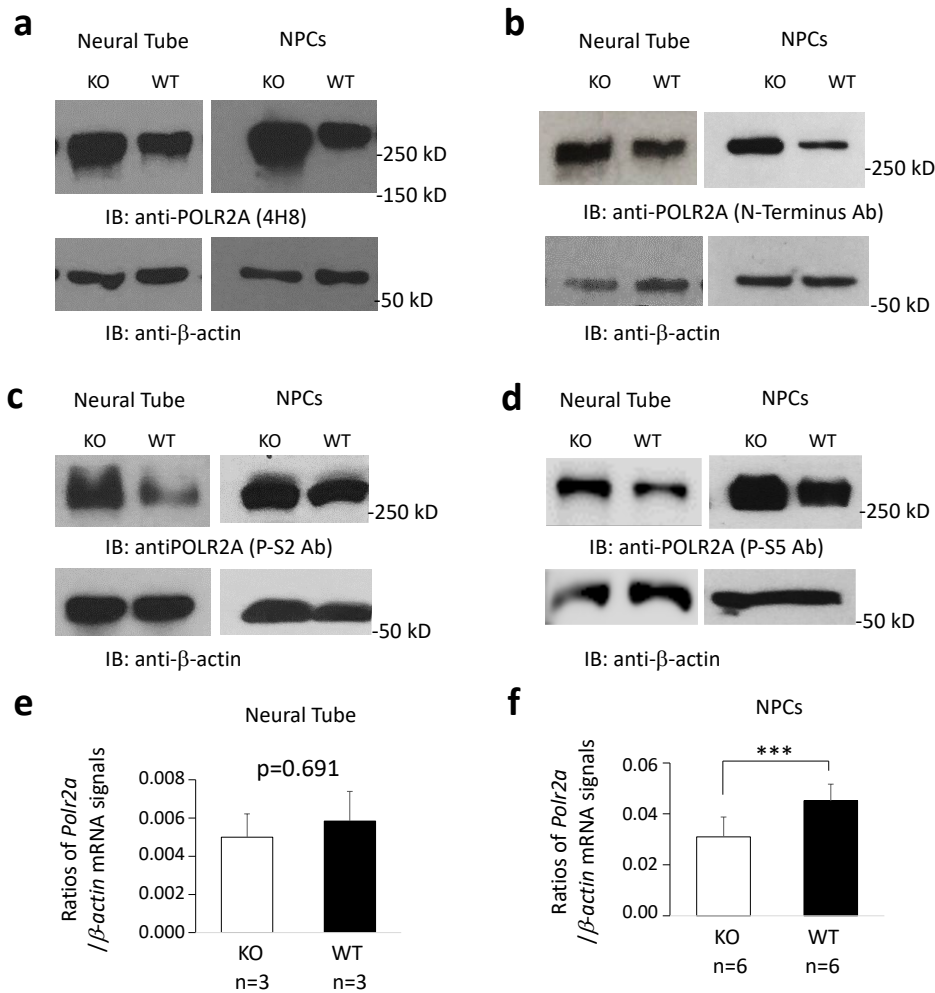
Fig. 4.4. POLR2A protein accumulation in KO neural tubes and NPCs. (a) and **(b).** Increased levels of hyper- and hypo-phosphorylated POLR2A protein (recognized by mAb 4H8) and total POLR2A protein (recognized by anti-N-terminal mAb F12) in e9.5 KO neural tubes and KO NPCs according to immunoblotting. **(c) - (d).** Increased levels of S2-phosphorylated POLR2A (c; recognized by mAb E1Z3G), and S5-phosphorylated POLR2A (d; recognized by mAb D9N5I) in e9.5 KO neural tubes and KO NPCs. **(e).** *Polr2a* mRNA levels of KO and WT e9.5 neural tubes were similar according to RT-qPCR (paired two-way Student's t-test). **(f).** *Polr2a* mRNA levels in the KO NPCs were lower than the WT counterparts. ***: p < 0.001 (paired two-way Student's t-test). The experiments were conducted three times, and representative results are shown.

of a proteasome inhibitor MG132, the total ubiquitination or K48-linked ubiquitination of POLR2A in both WT and KO NPCs was increased (**Figure 4.5a-c**), suggesting that the

ubiquitinated POLR2A is usually sent to the proteasome for degradation. However, the levels of the ubiquitinated POLR2A were consistently lower in the KO NPCs. The presence of some ubiquitinated POLR2A in KO cells also implied that in the absence of ARMC5-CUL3-RBX1 E3, another minor E3(s) is capable of ubiquitinating POLR2A, albeit less efficaciously. On the other hand, the amount of K63-linked total POLR2A only changed marginally in both KO and WT NPCs in the presence of the proteasome inhibitor (**Figure 4.5d**). This is compatible with our current knowledge that K63-linked ubiquitination mainly modifies protein function or channels the protein to lysosomes[433,434].

Figure 4.5e depicts the basic structure of this novel ARMC5-CUL3-RBX1 E3 and its interaction with its substrate POLR2A, based on our results and existing literature.

In SK-N-SH neuronal cells, ARMC5 was found in both cytosol and nuclei (**Figure 4.5f**). As POLR2A is a nuclear protein, the nuclear localization of ARMC5 is consistent with its function as the POLR2A-recognition subunit of a POLR2A-specific E3.

4.3.7. The impact of ARMC5 KO on the NPC transcriptome

We conducted RNA sequencing (RNA-seq) of KO and WT NPCs to evaluate how an enlarged Pol II pool affected NPCs' transcriptome. In NPCs, 47,059 transcripts from 16,475 genes showed detectable expression after filtering out those with less than one count per million reads. Due to concerns that the abnormal Pol II pool size might systematically skew all transcribed genes, we employed *Rn7sk* RNA as an internal control to normalize all the reads of each sample. *Rn7sk* is transcribed by Pol III and is not subjected to the possible influence of Pol II[435,436]. Indeed, *Rn7sk* expression in both KO and WT NPCs was similar (**Supplementary Figure 4.S.1**). A threshold for transcript-level significance of FDR < 0.05 was applied to the paired comparison of RNA-seq results from 3 KO and 3 WT NPC biological replicates. After filtering out transcripts that were not true positives (true positives were defined as having a complete exact match of intron chains with a GffCompare class code of "="[437]), we obtained 111 transcripts from 106 unique genes that showed significantly different expressions between KO and WT NPCs. These transcripts and genes were

listed in **Table 4.S.1**, along with their FDRs, fold changes, and raw reads. It is to be noted that three genes (i.e., *Fam172a*, *Slx1b*, and *Slc25a53*) each had one transcript upregulated and one downregulated **Table 4.S.1**. This resulted in 46 unique genes with upregulated
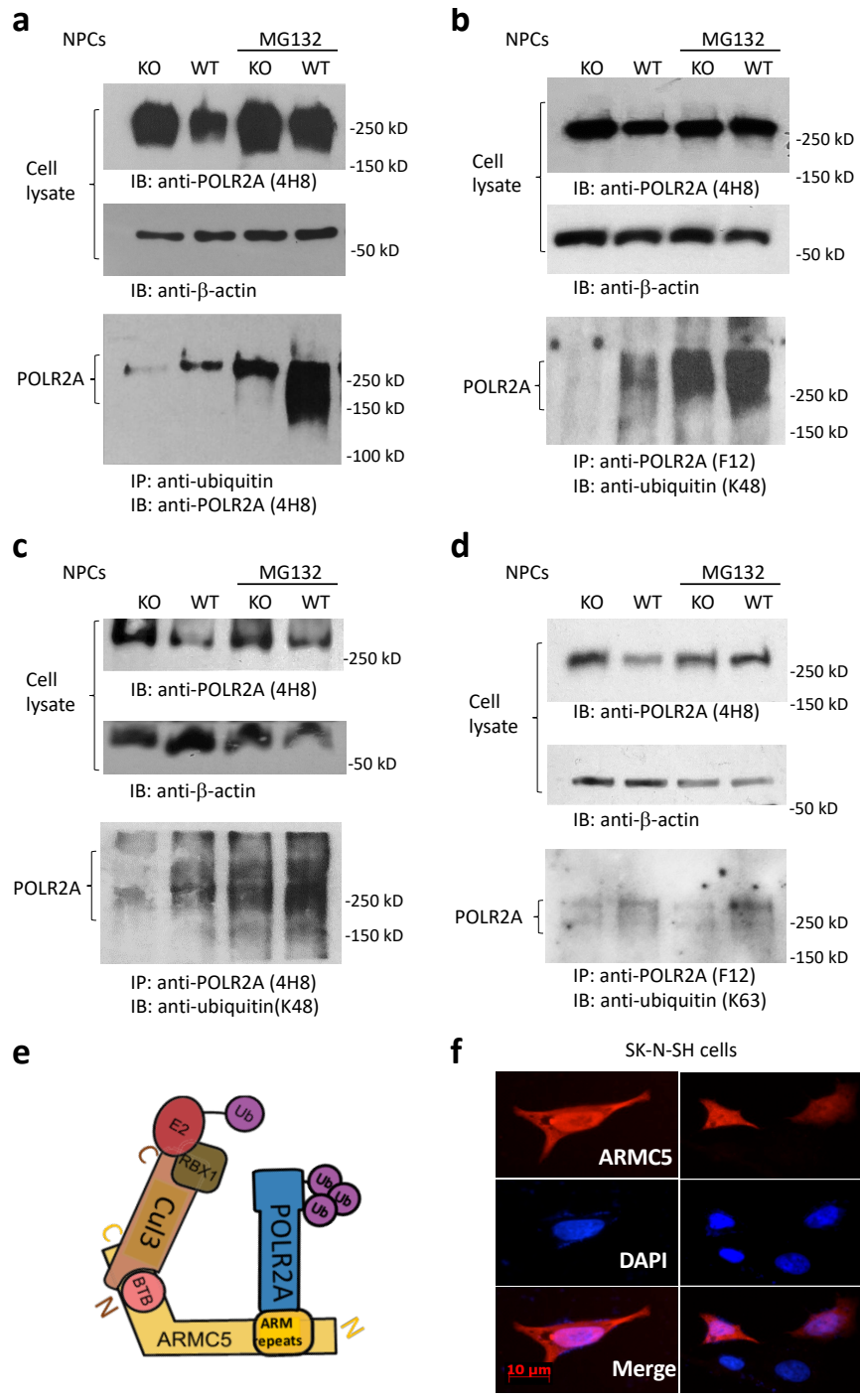
Fig. 4.5. **POLR2A-CUL3-RBX1 as a novel POLR2A-specific E3.** *(see next page)*

240

Fig. 4.5 *(previous page).* **(a - c).** Reduced K48-linked POLR2A ubiquitination in KO NPCs. KO and WT NPCs were cultured in the absence or presence of MC132 (10 μM). Ubiquitinated proteins in the lysates were precipitated with anti-ubiquitin Ab, and ubiquitinated POLR2A of both high and low CTD phosphorylation was revealed by immunoblotting with mAb 4H8 **(a).** In **a**, all the ubiquitinated proteins were precipitated by anti-ubiquitin Ab and then blotted with anti-POLR2A mAb 4H8. In **b** and **c**, the total POLR2A **(b)** or POLR2A with high and low CTD phosphorylation **(c)** was precipitated with mAb F12 or mAb 4H8, respectively. Their K48-linked ubiquitination was determined by immunoblotting using anti-ubiquitin Ab. **(d).** *Armc5* KO affected K63-linked POLR2A ubiquitination. Total POLR2A in KO and WT NPCs were precipitated by mAb F12, and K63-linked POLR2A ubiquitination was determined by immunoblotting using anti-K63 ubiquitin Ab. **(e).** A schematic illustrates the structure of POLR2A-specific ARMC5-CUL3-RBX1 E3. **(f).** ARMC5 was detected in both cytosol and nuclei. ARMC5-HA-expressing plasmid-transfected SK-N-SH cells were fixed 48 hours after transfection. The presence of ARMC5 in the nuclei and cytosol was determined by immunofluorescence. ARMC5-HA: pseudo-red; nucleus staining by DAPI: pseudo-blue. All experiments were repeated three times, and representative ones are shown. IP: immunoprecipitation; IB: immunoblotting.

transcripts and 63 unique genes with downregulated transcripts. Sixty transcripts in this list with the lowest FDRs are shown in a heatmap (**Figure 4.6a**), in which the color represents the standard deviation beyond the mean expression of the gene in all the samples tested (i.e., 3 WT and 3 KO NPC samples). A volcano plot illustrates the fold-change and FDR of the significantly changed genes, with several prominently changed ones annotated (**Figure 4.6b**). *Armc5* was among the downregulated ones, as expected.

One of the possible purposes of POLR2A Ubiquitination is to remove persistently stalled Pol II to allow transcription to resume in the case of DNA damage or cellular stress. Failure to remove the stalled Pol II is believed to cause a general decrease in transcription. However, to our surprise, this was no generalized depression of transcription in NPCs according to RNA-seq. Only 111 transcripts were significantly dysregulated, 48 (43.2%) being upregulated and 63 (56.8%) being downregulated (**Figure 4.6c**). As five genes had one transcript upregulated and one downregulated, there were 106 unique genes being dysregulated. For the vast majority of the genes (16,475 - 106 = 16,369 genes) that had detectable expression

in NPCs, their expression was not influenced by the failed degradation of POLR2A. This issue will be discussed later.

We performed gene ontogeny (GO) analysis of the significantly changed genes for their relationship to biological processes. Twenty-eight significant terms were identified. In addition, eight terms with high relevance to neural tube development were also chosen, even they were not statistically significant. The genes associated with each of these terms are registered in **Table 4.S.2**. Fifteen terms with known relevance to NTD were selected, and the number of the significant genes associated with a particular term is depicted in a bar graph (**Figure 4.6d**).

We conducted nuclear run-on assays on several genes (i.e., *Dnah9*, *Ifi44*, *Irf8e7*, and *Tgfb1*) that were upregulated according to RNA-seq at the transcript level (*Dnah9* and *Ifi44*) or at the gene level (*Irf8* and *Tgfb1*) and confirmed that their de novo transcription was upregulated, consistent with their steady-state mRNA levels according to RNA-seq (**Figure 4.6e**). The nuclear run-on assay was also used to assess another group of four genes (*Gapdh*, *Rpl10*, *Rplp0*, and *Ubc*) that were not modulated in the KO NPCs according to RNA-seq, and as expected, their de novo transcription was similar to their WT counterparts. These results corroborate those of RNA-seq, suggesting that the RNA-seq results largely reflect the rate of mRNA transcription, and there is no generalized transcription suppression in KO NPCs.

4.3.8. The effect of compromised POLR2A degradation on gene-associated Pol II density

The accumulation of POLR2A in the KO cells raised the question whether it was part of the stalled Pol IIs due to failed POLR2A degradation. We, therefore, conducted POLR2A ChIP-seq in NPCs, and the results were analyzed along with RNA-seq data. POLR2A signals were customarily used as a surrogate marker of Pol II[264,263], as it is the largest and catalytic Pol II subunit, and its jaw binds the incoming template DNA. Indeed, a very large fraction of POLR2A is localized in the same place in the nucleus as other components of

Pol II[438], and most POLR2As are also co-localized with DNA-bound histone protein HTA2, suggesting that they are part of DNA-bound Pol IIs[438]. This is consistent with the notion that most Pol IIs are engaged in the genes[439]. A total of 12,107 genes had discernable ChIP-seq signals. The distribution of Pol II peaks in different regions of the genome was
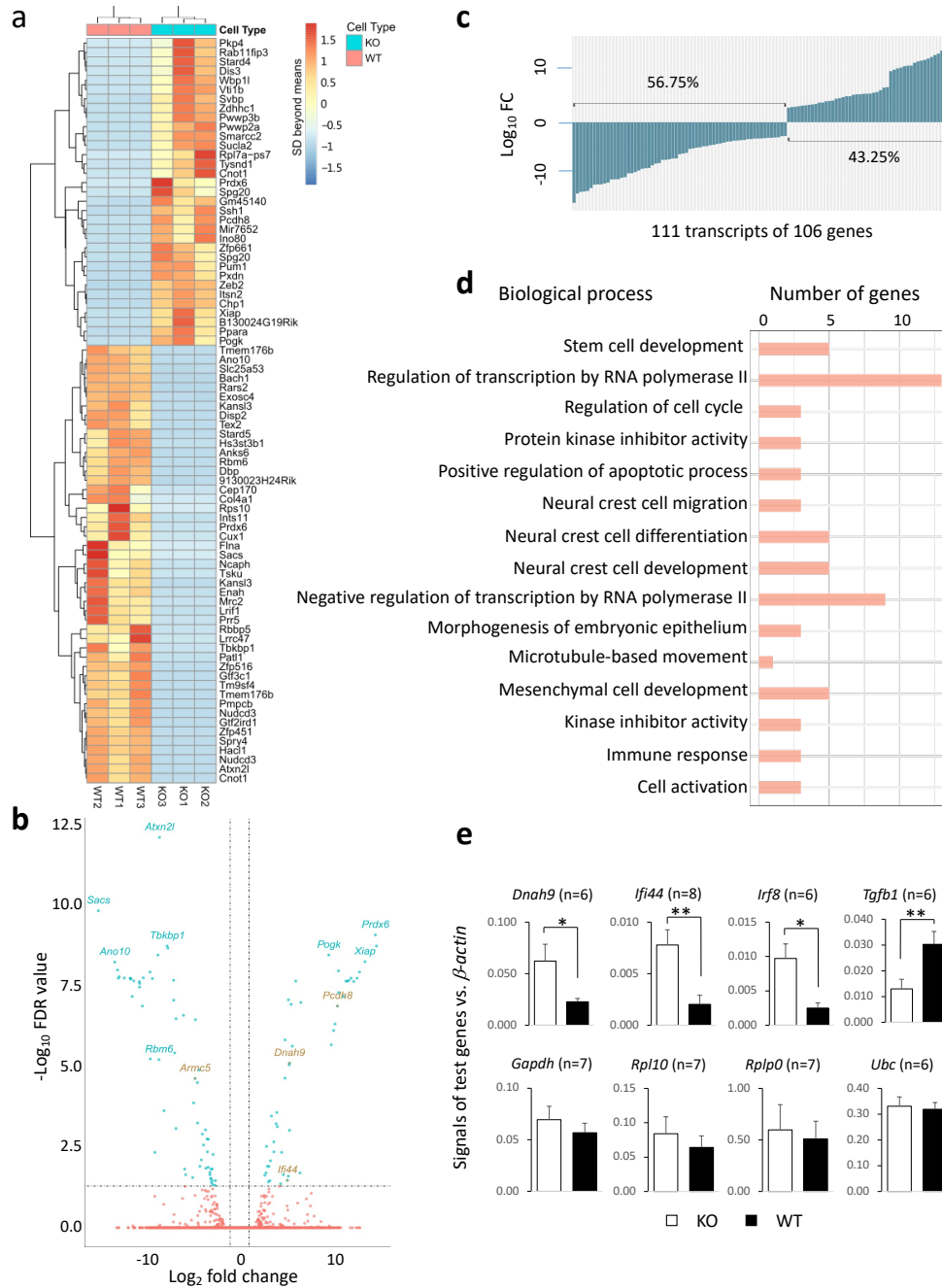
Fig. 4.6. Transcriptome analysis of KO and WT NPC by RNA-seq *(see next page)*

243

Fig. 4.6 *(previous page).* RNA-seq was conducted using three biological replicates for WT and KO NPCs in pairs. **(a).** Heatmap of 60 genes with the lowest FDRs among 106 genes with FDR < 0.05 found in NPC RNA-seq. Results of 3 biological replicate pairs (WT and KO) of NPCs are presented. Colors represent SDs beyond the normalized means of each gene. **(b).** A volcano plot depicting fold changes and FDRs of mRNA levels in KO versus WT NPCs according to RNA-seq. The dashed vertical lines indicate $\log_2$ fold changes (2-fold increase or decrease), and the dashed horizontal line marks FDR = 0.05. The names of some prominently changed genes with the lowest FDR or biggest fold changes are indicated. **(c).** $\log_{10}$ fold changes of all the 111 transcripts (106 genes) with FDR < 0.05. The percentages of transcripts with upregulation and downregulation are shown. **(d).** GO analysis in terms of biological processes for the genes with FDR < 0.05. Fifteen terms with high relevance to NTD were selected out of 28 significant terms. The percentages of the FDR significant genes found in RNA-seq among all the genes belonging to a particular term in the GO databank are indicated. **(e).** Nuclear run-on validation of RNA-seq results for four genes with significant increase and four genes without change in the KO NPCs. The number of repetitions is indicated. *: p < 0.05; **: p < 0.01 (paired two-way Student's t-tests).

illustrated in **Figure 4.7a**. In both KO and WT NPCs, the introns had the highest peak number, followed by intergenic regions and then the promoter regions. Within the genes, the highest normalized *Polr2a* read counts (read count per million mapped reads (CPM)) were accumulated near the transcription start site (TSS) (**Figure 4.7b**). Representative counts per million heatmaps for the region from -2,000 bp upstream of TSS to +2,000 bp downstream of the transcription end site (TES) of all genes in one pair of WT and KO samples are illustrated in **Figure 4.7c**. The Pol II density of all genes for a fixed region spanning from -10 kb to +10 kb surrounding the TSS was shown in **Figure 4.7d**. These metagene analyses (**Figure 4.7a-d**) are for a visual appreciation of all the genes at the same time, but are not suitable for statistical analysis. No visually discernable Pol II density differences were observed between KO and WT NPCs in these metagene analyses.

However, statistical analysis of individual genes revealed that a total of 59 of them showed significant difference (FDR < 0.1) in Pol II density in terms of normalized read counts in KO versus WT NSCs (23 genes in the TSS region (from TSS -400 bp to TSS +100 bp); 33 genes in the gene body region (from TSS +100 bp to TES -100 bp); and 3 genes in the TES region (from TES -100 bp to TES + 2,000 bp) (**Table 4.S.3**). Interestingly, with the exception of

three genes (i.e., *Tex14*, *Ttyh1*, and *Adcyap1r1*) in the gene bodies, the *Polr2a* counts per million in these significant regions and genes were all increased in the KO NPCs. The Pol II density tracks of 4 genes (i.e., *Cdkn1a*, *Gadd45b*, *Mafa*, and *Pcdh8*), which had either the highest increase of Pol II density in the KO NPCs or had known relevance to NTD, were illustrated in **Figure 4.7e**. The higher Pol II density in these genes was associated with increased mRNA levels, as confirmed by RT-qPCR (**Figure 4.7f**). This finding suggests that for a subset of genes, a larger Pol II pool promotes their transcription. As will be discussed later, some of these affected genes are repressive to cell proliferation, and their upregulation might contribute to the observed lower proliferation rates in the neural tubes and NPCs (**Figure 4.2a, d, and e**). It is to be noted that among these four RT-qPCR-validated upregulated genes, only *Pcdh8* but not the other three were identified in RNA-seq as differentially expressed genes. Such false negativeness is probably due to stringent multiple testing corrections. Further RT-qPCR validation of genes with increased Pol II density might find more of such genes with a concomitant increase of Pol II density and transcription.

We determined the pausing index (PI), which is defined as the ratio of POLR2A CPM in the TSS region versus that in the gene body region of all genes with ChIP-seq signals. None of them showed significant differences in PI between KO and WT NPCs (**Figure 4.7g**), suggesting that there is no genome-wide decrease of transcription, consistent with the RNA-seq and RT-qPCR results.

4.3.9. ARMC5 mutation related to *FOLH1* expression and MM

NTD has multifactorial pathogenic mechanisms, and dysfunctional NPCs probably only contribute to some extent of the pathogenic process. Other critical contributing factors include folate intake and metabolism[419]. Obviously, genes involved in folate metabolism have their major expressing in tissues other than NPCs or neural tubes. *Folh1* encodes folate hydrolase, which is necessary for breaking down polyglutamylated folates in the food into monoglutamyl folates for the folate uptake in the intestine[440]. Since this enzyme has glutamate carboxypeptidase activity, it is also expressed in a large number of other tissues such as

the brain, liver, adrenal glands, etc.[441]. In a separate project where we conducted an RNA-seq of the adrenal glands, we noticed that *Folh1* expression in the KO tissue was significantly reduced (**Figure 4.8a**). This finding was confirmed by RT-qPCR (**Figure 4.8b**). In NPCs, the Folh1 mRNA level was too low to be detected by RNA-seq, but RT-qPCR confirmed significantly lower Folh1 expression in the KO NPCs (**Figure 4.8b**). More importantly,
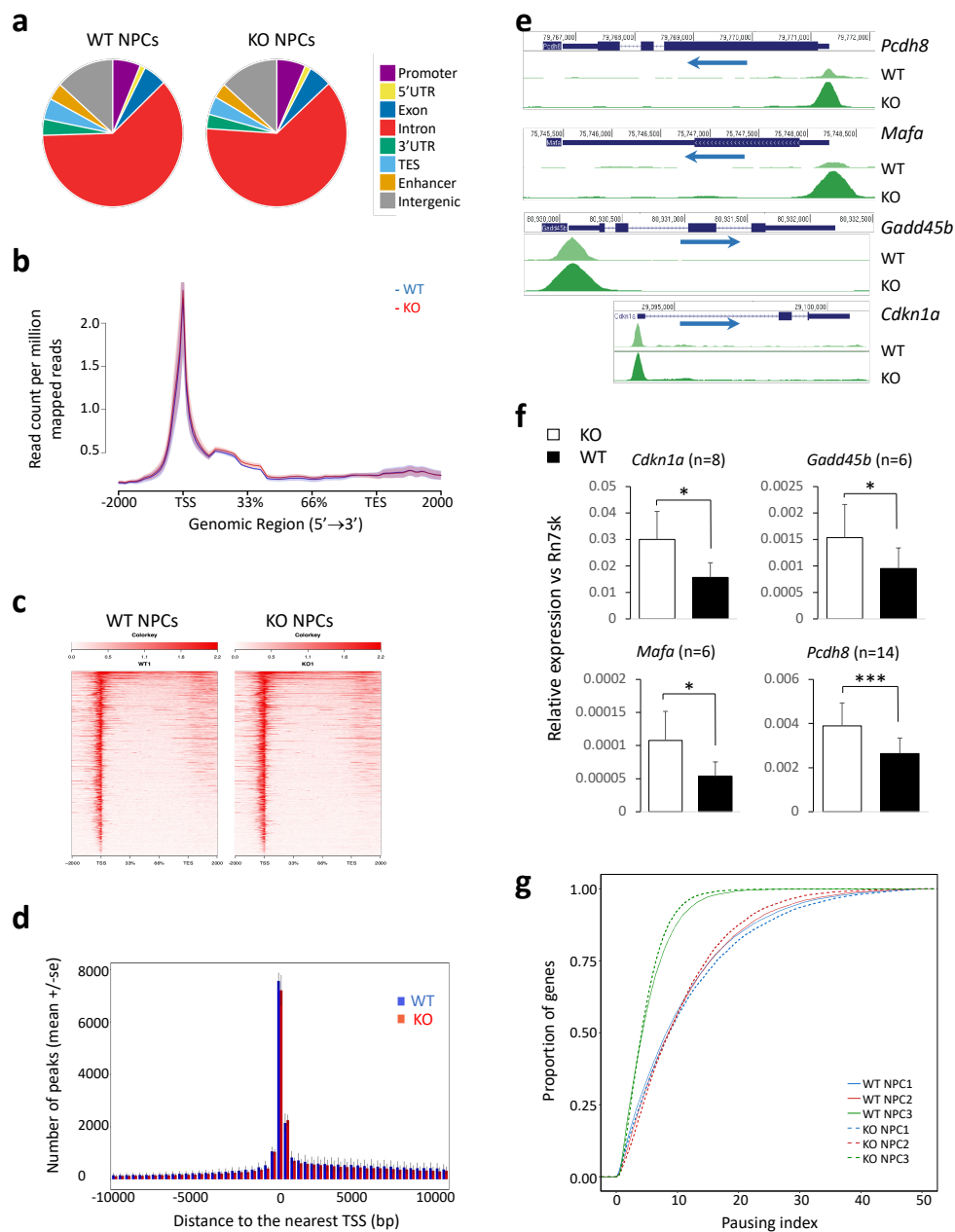
Fig. 4.7. **POLR2A ChIP-seq analysis of KO and WT NPCs** *(see next page)*

Fig. 4.7 *(previous page).* **(a).** ChIP-seq was conducted using three biological replicates for WT and KO NPCs in pairs. Pol II peak distribution in different gene regions of a representative pair of KO and WT NPC samples. **(b).** Means (solid lines) ± SE (shadows) of normalized read counts (read counts per million mapped reads) in a metagene analysis for a genomic region from -2 kb of TSS to +2 kb of TES. Data were based on three biological replicates. **(c).** Heatmaps of counts per million reads based on data from a representative pair of KO and WT NPC samples. (d). Pol II peak distribution (mean ± SE) in a fixed region from -10 kb upstream to + 10 kb downstream of TSS. **(e).** Read count tracks in the genes of *Cdkn1a*, *Gadd45b*, *Mafa*, and *Pcdh8*. The tracks were normalized so that each value was proportional to the read count per base pair per 10 million reads. **(f).** Upregulated mRNA levels of four genes (i.e., *Cdkn1a*, *Gadd45b*, *Mafa*, and *Pcdh8*) that had elevated Pol II density in their genes. The number of repetitions is indicated. *: $p < 0.05$; ***: $p < 0.001$ (paired two-way Student's t-tests). **(g).** Pausing indices (PIs) of genes with detectable signaling in ChIP-seq. The percentages of genes with different PI are plotted. Solid lines: WT NPC samples; dashed lines: KO NPC samples. No statistically significant difference of PIs between the KO versus WT NPC samples for all the genes with ChIP-seq signals was found (paired two-way student t-tests followed by multiple-testing correction).

in the KO intestine, the FOLH1 protein level was significantly lower than that in the WT counterpart (**Figure 4.8c**).

4.4. Discussion

We report here that *Armc5* deletion in mice significantly augmented NTD risks. Several SNVs in the human ARMC5 coding region were highly associated with MM in patients. This finding in human genetics established the relevance of our results in mice to human NTD. ARMC5 was found to be part of a novel dominant POLR2A-specific ubiquitin ligase under a physiological condition. In the absence of ARMC5, POLR2A protein ubiquitination and its subsequent degradation were compromised, leading to an abnormally large Pol II pool. The enlarged Pol II pool only specifically influenced the transcription of 106 genes in NPCs, some of which are involved in processes critical for neural tube development. However, the lack of this dominant POLR2A-specific E3 did not cause general Pol II stalling or transcription depression. *Armc5* KO resulted in reduced intestinal FOLH1 protein levels, which is critical for folic acid absorption.
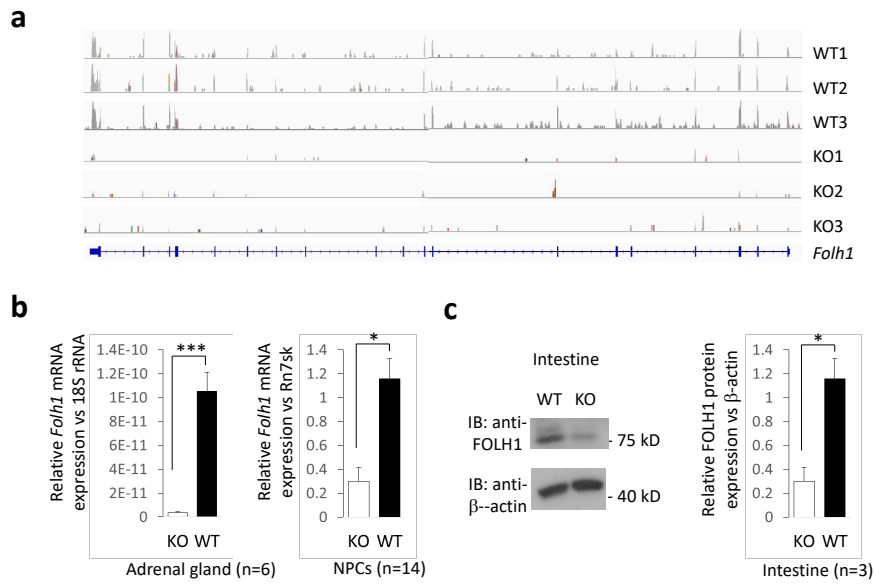
Fig. 4.8. *Armc5* **KO results in reduced** *Folh1* **expression in the intestine. (a).** Reduced *Folh1* RNA-seq read density in KO adrenal glands. *Folh1* RNA-seq read density tracks of three pairs of WT and KO adrenal glands are illustrated. **(b).** Reduced *Folh1* mRNA in the KO adrenal glands (left panel) and NPCs (right panel) according to RT-qPCR. The number of repetitions is shown. *: $p < 0.05$; ***: $p < 0.001$ (paired two-way Student's t-tests). **(c).** Decreased FOLH1 protein levels in the intestine of KO mice according to immunoblotting. A representative blot is shown. The relative ratios (mean $\pm$ SD) of FOLH1 versus β-actin signals of KO and WT intestines based on three independent experiments are shown in the bar graph. *: $p < 0.05$ (paired two-way Student's t-test).

4.4.1. ARMC5-CUL3-RBX1 is a novel POLR2A-specific E3.

We demonstrated that ARMC5 physically interacted with CUL3, which is often a component of multiple subunit RING-finger E3s. This class of E3s routinely contains a RING-finger protein, RBX1, which interacts with CUL3[442]. The interaction of CUL3 and RBX1 is well documented in the literature[442,356,103]. CUL3 also recruits a BTB domain-containing protein as a substrate-recognizing subunit. In our case, BTB domain-containing and CUL3-interacting ARMC5 served as the substrate-recognizing unit for substrate POLR2A, as ARMC5 also bound POLR2A. Therefore, ARMC5-CUL3-RBX1 plus other necessary molecules such as E2 formed a large complex for POLR2A ubiquitination. The function of this E3 was evidenced in that in the absence of ARMC5, there was reduced POLR2A

ubiquitination (**Figure 4.5a - d**). The function of the substrate recognition subunit is to pull a substrate close to the complex, so the cascade of E1, E2, and E3 can sequentially act on it, as depicted in (**Figure 4.8e**). Anything near enough to this complex, including the components within the E3 complex, will probably be ubiquitinated. Indeed, the self-ubiquitination of E3 components has been well-documented[443]. Recently, Cavalcante et al. reported that ARMC5 was ubiquitinated by CUL3[374]. This is compatible with our finding that ARMC5-CUL3-RBX1 is an E3. Technically, ARMC5 can also be considered as a substrate of E3 due to its self-ubiquitination. However, an enzyme's true substrate needs to propagate a cascade of events further downstream and realize a biological consequence. In this sense, ARMC5-self ubiquitination restricts the reaction within this E3 complex but does not propagate the event to downstream effector molecules. Therefore, POLR2A but not ARMC5, is the real substrate of this E3 complex.

Human Pol II has 12 subunits, and POLR2A is the largest subunit[335]. Along the same vein that an E3 ubiquitinates anything near enough to it and its specificity depends on what molecules its substrate recognition subunit pulls close enough to its vicinity, we noticed that multiple other Pol II subunits (i.e., POLR2B, POLR2C, POLR2H, POLR2I, and POLR2K; (**Figure 4.3a and e**)), in addition to POLR2A, were found in the ARMC5 co-precipitates. This raises two possibilities. First, these other subunits might directly associate with ARMC5 and are additional substrates of this novel E3. A more likely possibility is that these subunits are pulled down as parts of the whole Pol II and only indirectly interact with ARMC5 via POLR2A. If this is the case, the subunits proximal to POLR2A in the huge Pol II complex might also be ubiquitinated by this novel E3, while the more distal ones from POLR2A might not. The end result of both possibilities will be the same, i.e., this E3 ubiquitinates multiple Pol II subunits and even other accessory molecules in the huge transcription machinery, and causes their subsequent degradation. If it is the case, this E3 will certainly play a vital role in regulating the Pol II pool size, as it ubiquitinates multiple Pol II subunits. An E3 that directly or indirectly ubiquitinates multiple subunits of a protein complex has

been documented before: UBE2O E2/E3 can ubiquitinate multiple (more than 7) ribosome subunits of the ribosome complex[444].

We initially found that ARMC5 was mainly located in the cytosol in unmanipulated HEK293 cells[8]. This contradicted the clear evidence that ARMC5 interacted with POLR2A and other Pol II subunits that are largely located in the nucleus. The nuclear localization of ARMC5 in SK-N-SH cells resolved part of the paradox in this type of NTD-relevant cells. In HEK293 cells, we recently found that when they were treated with a nuclear export inhibitor leptomycin B, ARMC5 became detectable in both the nucleus and cytosol[445], suggesting that ARMC5 actively shuttles between the cytosol and nucleus. The rates of import versus export are likely different in different types of cells. In SK-N-SH neuronal cells, maybe the export is slower so that we can detect the nuclear presence of ARMC5. On the other hand, for HEK293 cells, the export might be faster than the import. As a consequence, the ARMC5 nuclei presence can only be revealed if the export inhibitor is applied. In any case, this finding resolves the earlier paradox that ARMC5 needs to act on nuclear proteins, but it cannot be found in the nuclei in some types of cells.

Ubiquitination is carried out by a cascade of enzymes[344]. E3 controls the substrate specificity of ubiquitination. Each protein has its specific E3, sometimes more than one. Several POLR2A-specific E3 in mammalian cells have been reported, such as Nedd4[275], Wwp2[194], pVHL-ElonginBC-Cul2-RBX1[279,446], and ElonginA-ElonginB/C-Cul5-RBX2[282]. These POLR2A-specific E3s only have demonstrable activities after massive DNA damage and cellular stress caused by irradiation or chemicals when there is an excessive need to remove stalled Pol II. The activities of pVHL-EloB/EloC-CUL2-RBX1 and WWP2 can be detected in the absence of exogenous DNA damages in cell lines[194,279], but have not been extended to tissues or organs. Theoretically, there should exist a POLR2A-specificity E3 in tissues and organs to deal with stalled Pol II caused by daily low-degree DNA damage or cellular stress. Such an E3 is also needed to remove misassembled Pol II, misfolded POLR2A, and permanently paused Pol II at the promoter region. However, this hypothetical E3 has so far eluded us.

The novel ARMC5-CUL3-RBX1 E3 we reported here was such an E3. It was the dominant POLR2A-specific E3 in normal neural tubes and primary NPCs, as ARMC5 deletion caused a drastic increase of POLR2A protein level in these tissues and cells in the absence of any artificially induced DNA damage or cellular stress. A similar huge accumulation of POLR2A was also observed in all major organs (i.e., thymus, lymph nodes, kidney, spleen, liver, lung, brain, heart, stomach, small intestine, colon, and adrenal gland) of the KO mice[445]. Thus, this novel ARMC5-CUL3-RBX1 E3 is the dominant POLR2A-specific E3 under a physiological condition in most organs and tissues. Other POLR2A-specific E3s may be needed only when there are excessive demands, such as significant DNA damage or severe cellular stress. In KO NPCs, in the presence of a proteasome inhibitor, there was a low-degree increase of both total POLR2A and ubiquitinated POLR2A. This suggests that some other DNA damage-independent E3s play a lesser role in POLR2A ubiquitylation and subsequently channel it to the proteasome for degradation. WWP2 and, to some extent, VHL-ElonginB/C-CUL2-RBX1 might be such E3s, as their activities could be observed in the absence of massive DNA damage[279,194,446], although the latter's activity is more obvious after UV irradiation.

The phosphorylation of S2 and S5 in the POLR2A C-terminal domain repeats is related to the location of POLR2A in the gene[370]. The previously documented POLR2A-specific E3s are often specific to POLR2A with a particular type of CTD phosphorylation. For example, VHL-ElonginB/C-CUL2-RBX1 and ElonginA/B/C-CUL3-RBX1 act on POLR2A with hyper-phosphorylated CTD S5[446,279,282,272]. WWP2 ubiquitinates total POLR2A as well as POLR2A with hyperphosphorylated CTD[194]. Nevertheless, our newly found ARMC5-CUL3-RBX1 E3 is a universal one and acts on Pol II anywhere in the gene, ubiquitinating POLR2A with hyperphosphorylated or hypophosphorylated CTD S2 or S5, or unphosphorylated POLR2A (**Figure 4.4**).

The critical function of the dominant DNA damage-independent POLR2A-specific E3 is evidenced by the severe phenotypes observed in *Armc5* KO mice, such as small body size, a live-birth rate below the expected Mendelian ratio, compromised T-cell immunity,

increased incidences of neural tube defects, and adrenal gland hypertrophy[8]. The existence of other minor POLR2A-specific E3 can probably partially compensate for the missing E3 in the case of *Armc5* KO, so some KO mice can still survive, and most of their organs are still functional. It will be prudent to state that in addition to POLR2A and possibly other Pol II subunits, this ARMC5-CUL3-RBX1 E3 might have other substrates, the dysfunction of which after *Armc5* KO or mutation contributing to some of the observed phenotypes in KO mice and patients.

4.4.2. The effect of compromised POLR2A degradation on the transcriptome

During mRNA transcription, if the transcription machinery encounters template DNA damage or cellular stress, Pol II will stall until the damage is repaired or the stress relieved[336]. It is believed that persistent Pol II stalling prevents transcription from resuming unless the stalled Pol II is degraded by proteasomes[341,342,447]. It follows that if POLR2A ubiquitination is compromised, there will be a general decrease in mRNA transcription. We assessed the transcriptome of KO NPCs by RNA-seq, but to our surprise, only 63 genes out of the 16,475 expressed genes in the KO NPCs showed reduced mRNA levels. The steady mRNA is the sum of newly transcribed mRNA and the degradation of existing ones. We evaluated the transcription rates of eight selected genes (four upregulated and four unchanged according to RNA-seq) by the nuclear run-on assay, which measures the transcription rate. The results were compatible with their steady-state mRNA levels according to RNA-seq, suggesting that the steady-state mRNA levels determined by RNA-seq largely reflected the transcription rates. Pol IIs in the TSS region are considered static or are in the process of being nudged into active transcription. On the other hand, Pol II in the gene body is more likely to be in the active transcription process. The pausing index, which is the ratio of Pol II density in the promoter region versus that in the gene body, is often used to gauge the transcription activity of a gene[194,427,442]. According to POLR2A ChIP-seq, none of the genes detected by RNA-seq in the KO NPCs had a significantly higher pausing index (**Figure 4.7g**). These

results collectively indicate that there's no general decrease in the transcription rate in KO cells, despite the failed POLR2A degradation and the subsequent POLR2A accumulation.

These findings suggest two non-competing possibilities. It is possible that under a physiological condition, Pol II stalling is an insignificant event. Even without this novel dominant E3 to remove the stalled ones, some other POLR2A-specific E3 will be sufficient to remove the small amount of stalled Pol II. Consequently, transcription is not systemically compromised. Equally possible is that our current knowledge about removing stalled Pol II by proteasomes is based on experiments using cells with massive DNA damage[336] or based on *in vitro* experiments[378,379]. Maybe the ubiquitination and proteasome-mediated Pol II degradation is not needed to remove the stalled Pol II at all, and there are other mechanisms to recycle stalled Pol II. Indeed, this is the case in yeasts[380,381]. Recent studies show that POLR2A K1268 ubiquitination is definitely required for UV irradiation-induced Pol II degradation[263,264]. When K1268 is mutated, POLR2A cannot be degraded. However, Pol II can still come off the damaged DNA template and recycle rather normally, suggesting that the process of removing the stalled Pol II can be independent of the ubiquitination/proteasome system[264]. The end result of both scenarios is the same: the loss of the major POLR2A-specific E3 does not cause generalized Pol II stalling under a physiological condition. Indeed, in the KO NPCs, with the exception of 59 genes (out of 12,107 genes with Pol II density signals; using FDR = 0.1 as a threshold), the majority of the genes (12,107 - 59 = 12,048 genes) had similar Pol II density in the KO and WT NPCs.

Since this E3 is not for removing stalled Pol II, its major function is likely to control the Pol II pool size. How the Pol II pool size affects transcription and cellular function is a question infrequently visited, especially under a physiological condition without artificially induced massive DNA damage. Intuitively, we would believe that since the same Pol II works for all genes, its pool size should universally affect all of them and probably increase their transcription. Clearly, this is not the case. The accumulation of POLR2A due to *Armc5* KO in NPCs only selectively influenced a limited number of genes. Forty-three genes were clearly upregulated and fifty-eight downregulated. In addition to these 101 genes, there were

five genes, each having one transcript upregulated and one downregulated, probably caused by different expression levels of different isoforms.

How does the Pol II pool size affect the expression of a subset of genes? First, it does not work alone, but needs additional tissue-specific transcription factors to modulate the transcription rate jointly. Recently, Vidakovic et al. reported that an enlarged Pol II pool due to p.Lys1268Arg mutation-caused POLR2A accumulation resulted in the upregulation of more than 1,600 genes in HEK293 cells, but only a few hundred downregulated ones[264]. We similarly observed predominantly upregulated genes in the KO adrenal glands and MEFs[445]. The abnormally expressed genes in these different cell types (NPCs (our currently RNA-seq data), HEK293 cells[264], and adrenal glands[445] were not identical but showed vast differences. Likely, in these different types of cells, the interaction between the enlarged Pol II pool and tissue-specific transcription factors decides whether a gene is abnormally expressed. In other words, some genes are more susceptible to the influence of the Pol II pool size than others in a given type of tissue.

What are the possible mechanisms by which the enlarged Pol II pool increases the transcription of certain genes? There is not much literature related to this topic, and we can only speculate at this time. The TSS region normally docks most of the Pol IIs (**Figure 4.7b-c**), and for a subset of genes, the higher Pol II density in the TSS region was correlated to higher transcription (**Figure 4.7e, f**). In addition to tissue-specific regulation, the number of TSSs in a given gene might be another factor deciding the susceptibility of a gene to the Pol II pool size. This is especially relevant to the upregulated genes. Only about 20% of the genes have a classical TATA box as the TSS, and others may harbor one or several non-canonical or cryptic TSSs[370]. Logically, the presence of more TSSs in a promoter favors the docking of more Pol IIs, making the gene more sensitive to the Pol II pool size and increasing the transcription rate.

It is more difficult to understand why a larger Pol II pool causes the downregulation of some genes. Maybe some of the downregulated genes lack other means to remove the stalled Pol II, which exists functionally but is not being detected due to statistical false-negativeness

in ChIP-seq. Because of the putative Pol II stalling in such genes, the transcription of these genes is compromised. It is also possible that some of these downregulated genes are indirectly influenced by the larger Pol II pool via other upregulated ones.

4.4.3. *ARMC5* mutation as an NTD risk modifier

In *Armc5* KO mice, a total of 43% of KO mice suffered from NTD. However, since NTD penetrance in the KO mice was not 100%, it means that *Armc5* mutation alone is not sufficient to cause NTD. Instead, such mutation is a modifier of NTD risks. Our mice were in a C57BL/6 × CD1 background. When the mice were bred to a pure C57BL/6 background, the NTD incidence in KO mice was less than 5%, suggesting that the genetic background of CD1 mice contributes to NTD. This might also be the case in humans, whose *ARMC5* mutations increase the risk of NTD caused by additional genetic factors.

Our human genetic study detected in MM patients nine missense mutations caused by SNVs, and they were among the top 5% in terms of functional deleteriousness (CADD > 13). There was one (p.Pro128Ser) in the N-terminus between aa96 and aa140 before the ARM domain, one (p.Arg429Cys) in the ARM domain, and three (p.Arg501Gln, p.Gly517Ser, and p.Pro654Leu) in the region between the ARM domain and BTB domain (**Figure 4.3c**). According to our deletion study, all these three regions contributed to the binding of ARMC5 to POLR2A (data not shown). We functionally assessed the highly deleterious mutant p.Arg429Cys, which was the top 1% deleterious mutation (CADD > 20) with a p-value of 0.011, and confirmed by immunoprecipitation followed by immunoblotting that this point mutation in the ARM domain drastically reduced association between ARMC5 and POLR2A. Another nearby missense mutation reported in PBMAH patients and in the same ARM repeat caused a similar decrease of binding between ARMC5 and POLR2A according to immunoprecipitation followed by LC-MS/MS. These results suggest that the missense mutations in these three regions likely reduce the interaction between ARMC5 and POLR2A, make the substrate recognition unit less effective, and compromise the enzymatic activity of this E3.

One SNV p.Arg888Gln was located in the ARMC5's BTB domain at the C-terminus, and this domain is known to interact with CUL3[348]. This mutation might reduce the interaction between ARMC5 and CUL3 and prevent it from forming an active E3. Three SNVs (i.e., p.Thr12Ala, p.Arg78His, and p.Ala84Val) only existed in the longest 1,030 aa *ARMC5* isoform, which has 95 extra amino acids in the N-terminus than the most abundant 935-aa isoform and is expressed only in the brain plus a limited number of other tissues[374]. At this time, we do not know the function of this 95-aa region nor how the mutations in this region affect the E3 activity.

These nine highly deleterious SNVs had augmented alternate allele frequencies and odds ratio (OR) in MM patients than in the control group, although only two novel SNVs attained statistical significance (p.Thr12Ala and p.Arg429Cys). For those deleterious SNVs found in MM and did not reach significance, they existed in a smaller number of individuals in the control group. MM is a severe form of NTD, with a prevalence of 0.06% of births in the US[448]. On the other hand, some milder forms of NTD have feeble or no clinical manifestations. The reported prevalence of spina bifida can be as high as 12% in the general population[449]. Many spina bifida patients remain undiagnosed. The gnomAD control data come from the "control" populations of multiple projects, controlling for just the phenotypes of individual projects. Spina bifida and MM were not the exclusion criteria for these studies. Therefore, there is a possibility that a few patients with spina bifida or even MM were unknowingly included in the controls, and they had those deleterious *ARMC5* SNVs. The presence of these patients in the control groups might have reduced the statistical significance of these SNVs. We need to emphasize that *Armc5* KO (in mice) or *ARMC5* mutation (in humans) is only a modifier of NTD and is not sufficient to cause NTD, as alluded to above. The contribution of *Armc5* KO to NTD in one allele is even smaller, as no NTD phenotype was observed in the heterozygous KO mice. The deleterious *ARMC5* SNVs in humans are only in one allele, and their effects are small. They will increase the risks of NTD but will need additional genetic or environmental factors for NTD manifestation. Thus, it is also possible that such additional factors were present in the MM patients harboring the SNVs but not

in the controls with such SNVs. This can alternatively explain the presence of deleterious SNVs in a few control individuals, assuming they did not have insidious NTD.

The KO mice and MM patients with *ARMC5* mutations had NTD in common. The KO mice had adrenal gland hypertrophy, which was also found in PBMAH patients with *ARMC5* mutations[18,2]. However, the KO mice manifested other phenotypes not found in humans with *ARMC5* mutations, such as a high embryonic death rate, dwarfism, and immune deficiency[8]. There are several explanations for this discrepancy. There is no functional ARMC5 protein in the KO mice due to the bi-allelic deletion of the *Armc5* gene. We did observe a dosage effect of *Armc5* mutation, as heterozygous KO mice were had no observable phenotypes. In humans, the mutations are often mono-allelic, and many are missense point mutations. Therefore, in these patients, ARMC5 function is not totally lost, and the mutation phenotypes tend to be milder, hence the absence of more severe phenotypes such as dwarfism. Moreover, humans with highly detrimental bi-allele *ARMC5* mutations might have perished during the embryonic or perinatal stage. For some less obvious phenotypes, such as compromised T-cell immune responses found in the KO mice, unless we intentionally search for them, they might not be noticed or might be misdiagnosed due to mild and nonspecific manifestations, as the immune system is highly redundant.

4.4.4. How does the larger Pol II pool size modify NTD risks?

The abnormally large Pol II pool size has serious negative consequences in cell biology and in many biological systems, as evidenced by many compromised phenotypes in our KO mice[8]. Recently, Nakazawa et al. generated *Polr2a*-(K1268R) knock-in mice[263], similar to our KO mice. The transgenic mice manifest a Cockayne syndrome-like phenotype with dwarfism. The Cockayne syndrome often involves in the impaired nervous system development[450]. Our and Nakazawa's data collectively suggest that the enlarged Pol II pool is detrimental to the development of the nervous system and proper neural tube folding. The large Pol II pool is the root cause upstream, and then it will dysregulate some downstream effector genes that influence neural tube development.

Neural tube folding is a complex event involving many cellular processes such as migration, proliferation, apoptosis, etc. We found that 108 genes were dysregulated in KO NPCs. As shown in (**Figure 4.7e and f**), *Cdkn1a*, *GAdd45b*, *Mafa*, and *Pcdh8* presented enhanced Pol II density accompanied by mRNA upregulation. CDKN1A (a.k.a. P21cip1) is a cyclin-dependent kinase inhibitor and vitally important in regulating cell cycle progression at the G1 and S phase[451]. It is an inhibitor for S-phase DNA synthesis[452]. High levels of CDKN1A render cells to enter the G0 quiescent state[453]. The upregulated *Cdkn1a* in the KO cells could well contribute to the reduced proliferation in KO NPCs and neural tubes. If this gene is confirmed to be upregulated in other types of KO cells, it might explain in part the generally compromised proliferation in those cells.

GADD45B is a protein interacting with CDKN1A and has negative control over cell growth[454]. It is a cyclinB1 kinase inhibitor and regulates cell cycling at the S and G2/M check point[455]. Its downregulation is associated with the risk of hepatocellular cancers[456]. Thus, *Gadd45b* upregulation might also play a role in the reduced proliferation of KO NPCs and neural tubes, and in reduced growth of other types of KO cells.

Mafa encodes a transcription factor and has dual functions as an oncogene and tumor suppressor gene, depending on the cellular context[388]. *Pcdh8* is a member of the cadherin superfamily and is reported as a tumor suppressor gene[384]. The upregulated *Mafa* and *Pcdh8* could well impart a negative effect on the NPCs and cells in the neural tubes. More likely, it is possible that multiple genes dysregulated by a larger Pol II each contribute to some extent to the occurrence of NTD, leading to an overall increase of NTD risks.

NPCs are but only one of the many types of cells involved in neural tube development. Other neuronal and interstitial cells at different locations, times, and developmental stages in the neural tube are also critical in the pathogenesis of NTD. The transcriptome of all these cells in the neural tube is changing by the hour. Conceivably, the abnormal Pol II pool size can cause dysregulation of some critical genes in many of these cell types at a given moment and increases the NTD risks.

Armc5 KO or mutation can affect other organs and tissues, whose function might be indirectly needed for proper neural tube development. FOLH1 expression in the intestine is a case in point. FOLH1 is a transmembrane protein and is a glutamate carboxypeptidase[440]. It is well established that sufficient folate is required for proper neural tube development[419]. Folate needs to be absorbed as an essential nutrient from the food[419]. Dietary folate deficiency and dysfunction of folate absorption and metabolism increase NTD risks[420,421,422]. Dietary folate exists in a polyglutamate form and needs to be digested by FOLH1 into monomers to be uptaken by the small intestine[457]. Homozygous *Folh1* KO in mice was embryonically lethal, indicating the vital function of FOLH1 in development[458,459]. Several human studies showed that *FOLH1* mutations are associated with low serum folate levels and increased NTD risks[440,460,461]. In our NPC RNA-seq, *Folh1* was not among the genes with a significantly different expression in KO cells, probably due to the detection sensitivity of RNA-seq and low expression levels of *Folh1* in these cells. However, there was a drastic decrease of *Folh1* expression in the adrenal glands according to the RNA-seq. This prompted us to assess the FOLH1 expression in the KO small intestine, as folates are mainly absorbed in the small intestine. Indeed, a drastically reduced FOLH1 protein level in the KO intestine was revealed, which might compromise folate absorption and consequently increase NTD risks. Due to the demonstrated importance of folate in neural tube development, it will be interesting to assess other enzymes in the one-carbon metabolism pathway in the relevant KO tissues to see whether the enlarged Pol II pool has a negative impact on the expression of these enzymes.

ARMC5-CUL3-RBX1 E3 might have additional substrates, the failed ubiquitination of which might also contribute to NTD susceptibility and other pathological conditions observed in KO mice and patients with deleterious *ARMC5* mutations.

In summary, *Armc5* KO mice manifested NTD, and a human genetic study revealed the association of *ARMC5* SNVs with MM risks. ARMC5 was part of a novel dominant POLR2A-specific E3 under a physiological condition in the absence of artificially-induced

DNA damage. *Armc5* KO or mutation caused an enlarged Pol II pool, which led to dysregulation of a subset but not all genes. Likely, the sum effect of some of the dysregulated genes resulted in increased NTD risks.

4.5. Materials and Methods

4.5.1. *In situ* hybridization

To localize *Armc5* mRNA, 1526-bp (starting from GATATC to the end) mouse *Armc5* cDNA (GenBank: BC032200, cDNA clone MGC: 36606) in pSPORT1 was employed as a template for sense and antisense riboprobe synthesis, with SP6 and T7 RNA polymerase for both ^{35}S-UTP and ^{35}S-CTP incorporation. Tissues from WT mice were frozen in -35 °C isopentane and kept at -80 °C until they were sectioned. X-ray autoradiography focused on 10 μm thick cryostat-cut sections. Briefly, overnight hybridization at 55 °C was followed by extensive washing and digestion with RNase to eliminate non-specifically bound probes. Anatomical level images of *in situ* hybridization were generated using X-Ray film autoradiography after four days' exposure.

4.5.2. Micro-CT whole-body bone imaging

The mice were euthanized by CO_2. The whole-body bone images were obtained by scanning the mice using Broker SkyScan1176 Micro-CT.

4.5.3. Collection of mouse fetuses

Fetuses were harvested for neural tubes (e9.5), for the assessment of anencephaly (e12.5), and for the generation of NPCs from CNS tissues (e13.5).

4.5.4. Neural tube isolation

The neural tubes were isolated from e9.5 mouse embryos under a dissecting microscope and digested with pancreatin (6 mg/ml in PBS) for 6 minutes at room temperature. Sticky lateral tissues were teased away and cleaned neural tubes were used in the experiments.

4.5.5. Generation of mouse NPCs

The brains from e13.5 mouse fetuses were separated at the cervical spinal cord level, and the ganglionic eminences were dissected and harvested. The harvested tissue pieces were collected in complete neural stem cell medium (NeuroCult$^{\text{TM}}$ NSC Basal Medium and NeuroCult$^{\text{TM}}$ NSC Proliferation Supplements at a 9:1 ratio; Stemcell Technologies) and dissociated thoroughly but gently by pressing the pipette tip to the bottom of the tube and pipetting five times to obtain a single-cell suspension. The cells were plated at a density of 2×10^5 cells/ml in complete NSC medium supplemented with 20 ng/ml EGF (Stemcell Technologies). Five to six days later, the neurospheres were treated with Accutase$^{\text{TM}}$ (Stemcell Technologies) and cultured for additional 5 – 6 days. The neurospheres of the second passage were used for experiments.

4.5.6. Immunofluorescence

E9.5 fetuses were fixed in PBS containing 4% paraformaldehyde at 4 °C overnight, and then sequentially soaked in PBS containing 30% sucrose at 4 °C for 24 hours followed by a mixture of 30% sucrose (in PBS) and OCT at 1:1 ratio at 4 °C for another 24 hours. The samples were then embedded in OCT and stored at -80 °C until use. Fetal WT and KO neural tubes at the level of hindbrains were cryosectioned (10 – 12 μm) transversely. The cryosections were first permeabilized with 0.3% Triton X-100 in PBS for 3 minutes and treated with blocking buffer (PBS containing 5% goat serum and 0.1% Tween 20) at room temperature for 1.5 hours. The sections were then reacted with rabbit anti-Ki-67 Ab (1:500 dilution; Abcam) at 4 °C overnight, followed by FITC-conjugated goat anti-rabbit Ab (1:1000 dilution; Millipore/Sigma) in the blocking buffer for two hours at room temperature. The sections were washed three times in PBS and mounted in ProLong Gold anti-fade (Invitrogen) containing DAPI.

For the quantification of apoptosis in e9.5 neural tubes, the cryosections described above were assessed by fluorescent TUNEL using *In situ* Cell Death Detection Kit (Roche) according to the manufacturer's instructions.

Fluorescent images were collected on an AxioPhot fluorescent microscope (Zeiss). The images were analyzed using the Cell Counter Plugin of the Image J software. Ki-67- or TUNEL-positive cells among total cells, which were visualized by DAPI staining in the neural folds were registered.

For immunofluorescent staining of NPCs, the cells were cultured on Poly-D-lysine and laminin pre-coated glass slips in the NeuroCultTM proliferation medium for one day. The cells were fixed with 4% paraformaldehyde in PBS and permeabilized with PBS containing 0.3% Triton X-100 for 3 minutes. The slips were then soaked in blocking buffer (PBS containing 5% goat serum and 0.1% Tween 20 at room temperature) for 1.5 hours and reacted with different first Abs (mouse anti-Nestin mAb, 4 μg/ml, Abcam; rabbit anti-Sox2 Ab, 1 μg/ml, Abcam; mouse anti-Myc mAb, 1 μg/ml, Cell Signaling Technologies). The slips were then incubated with secondary Abs (AlexaFluor 488-conjugated goat anti-mouse Ab; Invitrogen; rhodamine-conjugated goat anti-rabbit Ab; Jackson ImmunoResearch Laboratories) in the blocking buffer for 2 hours at room temperature. The slips were washed three times with PBS and mounted in ProLong gold anti-fade containing DAPI (Invitrogen).

For immunofluorescent staining of cytosolic and nuclear ARMC5, SK-N-SH neuroblastoma cells were cultured on CELLstart substrate (Invitrogen)-precoated glass slips overnight and transiently transfected for two days with plasmids expressing human ARMC5-HA (Genecopoeia) using Lipofectamine 3000 transfection reagent (Invitrogen). The procedure of immunofluorescent staining was the same as that for NPC staining except for that rabbit anti-HA mAb (Cell Signaling Technology), and rhodamine-conjugated goat anti-rabbit Ab (Jackson Laboratories) were used as the primary and secondary Abs, respectively.

4.5.7. NPC proliferation assay

NPCs were cultured in 96-well plates in complete NeuroCult^{TM} proliferation medium for 1 day. CellTiter 96 Aqueous One Solution (20 μl/well; Promega) was added to the wells. After an additional 2-hour culture, the absorbance of the wells at 490 nm was registered with an ELISA reader.

4.5.8. Flow cytometry

NPCs from KO and WT mice were cultured without EGF or with different concentrations of EGF for 20 hours. A single-cell suspension was obtained by treating the cells with Accutase. The cells were stained with Annexin-V (1:50 dilution; BD Bioscience) and analyzed by flow cytometry.

4.5.9. LC-MS/MS

HEK293 cells were cultured in DMEM medium supplemented with 10% fetal bovine serum and 2 mM glutamine and transfected with FLAG-tagged ARMC5- or ARMC5 Arg315Trp-expressing plasmids by using Jet Prime Transfection Reagent (PolyPlus). The transfected cells were incubated at 37 °C for 24 hours, washed with PBS, pelleted, and snap-frozen until use. Affinity purifications were performed in four independent replicate experiments as described previously[462]. The Speedvac-dried protein extracts were re-solubilized in 10 μl of 6M urea buffer reduced (45 mM DTT, 100 mM ammonium bicarbonate) for 30 minutes at 37 °C, and alkylated (100 mM iodoacetamide, 100 mM ammonium bicarbonate) for 20 minutes at 24°C. Proteins were digested in 10 μl of trypsin solution (5 ng/μl of trypsin, Promega; 50mM ammonium bicarbonate) at 37°C for 18 hours. The digests were acidified with trifluoroacetic acid and cleaned by the Oasis MCX 96-well Elution Plate (Waters). Peptides were identified by LC-MS/MS using HPLC coupled to an Orbitrap Fusion mass spectrometer (Thermo Scientific) through a Nanospray Flex Ion Source. MS/MS raw data were searched against the human SwissProt database (updated on April 24th, 2019) and X-Tandem using ProHits software[463]. Spectral counts were transferred in Perseus (Version 1.6.1.3)[464]. Proteins quantified in three out of four experiments for either WT ARMC5 or ARMC5 Arg315Trp were kept for further analysis. Spectral counts reported as 0 by X-Tandem were replaced by a randomly generated spectral count value normally distributed with a mean and S.D. equal to those of the lowest 20% spectral count values from the LC-MS/MS analysis. Spectral counts were normalized by the spectral count of the bait (ARMC5) to allow comparison between different purifications. WT ARMC5 and

ARMC5 Arg315Trp proteins were compared to the FLAG empty vector control samples and were labeled as high-confidence interactors when their p-value was under 0.05 and their spectral count ratio was over 1.5. Statistically significant differences between proteins from the affinity-purified WT ARMC5 and ARMC5 Arg315Trp samples were determined using a two-tailed t-test. They were subsequently adjusted for multiple testing using a Benjamini-Hochberg-based test[465]. FDR of 5% was adjusted using a 0 correction factor of 0.1. The level of differential interaction was considered statistically significant when the FDR was < 0.05 and its average spectral count fold-change between WT ARMC5 and ARMC5 Arg315Trp was $> \pm 2$.

4.5.10. Immunoprecipitation and immunoblotting

Human neuronal SK-N-SH cells were transfected with human ARMC5-HA-expressing plasmid or empty vectors. These human cells, mouse NPCs, or mouse e9.5 neural tubes were lysed in RIPA buffer (25 mM Tris, pH 7.6, 150 mM NaCl, 1% Nonidet P-40, 0.1% SDS) supplemented with protease inhibitors and phosphatase inhibitors (Roche Diagnostics). For immunoprecipitation, 0.5 mg of protein was incubated with mouse anti-HA mAb (clone HA-7; Sigma), mouse anti-ubiquitin mAb (clone F-11; Santa Cruz), mouse anti-POLR2A mAb (clone F-12; Santa Cruz Biotech), mouse anti-POLR2A mAb (clone 4H8; BioLegend) overnight, and then with protein-G pre-conjugated agarose beads for additional 2 hours at 4°C with rotary agitation. The beads were washed with lysis buffer three times and eluted in SDS-loading buffer. For immunoblotting, the lysates were resolved by 6 to 8% SDS-PAGE and transferred to nitrocellulose membranes. The membranes were blotted with mouse anti-HA-HRP mAb (clone 6E2Cell Signaling Technology), mouse anti-CUL-3 mAb (clone G-8; Santa Cruz), mouse anti-POLR2A mAb (clone 4H8; BioLegend), mouse anti-POLR2A mAb (clone 8WG16; BioLegend), rabbit anti-phospho-POLR2A-S2 mAb (clone E1Z3G; Cell Signaling Technology), rabbit anti-phospho-POLR2A-S5 mAb (clone D9N51; Cell Signaling Technology), mouse anti-ubiquitin mAb (clone F-11; Santa Cruz Biotech), rabbit anti-β-actin Ab (Cell Signaling Technology), mouse anti-K48-ubiquitin mAb (clone

Apu2; Millipore/Sigma), or mouse anti-K63-ubiquitin mAb (clone HWA4C4; Invitrogen), followed by HPR-conjugated horse anti-mouse IgG Ab (Cell Signaling Technology), or horse radish peroxidase (HPR)-conjugated goat anti-rabbit IgG Ab (Cell Signaling Technology). The signals were detected with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

4.5.11. RNA-seq

Total RNA of three biological replicates of KO and WT NPCs, respectively, was extracted by Rneasy kit (Qiagen). The total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies), and its integrity was assessed on a 2100 Bioanalyzer (Agilent Technologies). rRNA was depleted from 250 ng of total RNA using QIAseq Fast-Select (Human 96rxns; Qiagen). cDNA synthesis was achieved with the NEBNext RNA First-Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules (New England BioLabs). The remaining steps of library preparation were performed using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs), according to the manufacturer's instructions. Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using the Quant-iT^{TM} PicoGreen dsDNA Assay Kit (Invitrogen) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). The average size fragment was determined using a LabChip GX (PerkinElmer).

The libraries were normalized and pooled and then denatured in 0.05 N NaOH and neutralized using an HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S2 lane using Xp protocol as per the manufacturer's recommendations. The run was performed for 2×100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at a 1% level. Base-calling was performed with RTA v3. Program bcl2fastq2 v2.20 was then used to de-multiplex samples and generate fastq reads.

Reads were trimmed from the 3' end to have a phred score of at least 30. Illumina sequencing adapters were removed from the reads, and all reads were required to have a length of at least 32 bp. Trimming and clipping were performed using Trimmomatic[395].

Upstream processing of FastQ files was performed as described previously[396], using the ENSEMBL GRCm38 (Mus musculus) release 97 mouse reference genome sequences and annotations. Each readset was aligned using STAR[397], which creates a Binary Alignment Map file (.bam). Then, all readset BAM files from the same sample were merged into a single global BAM file using Picard.

The pairwise Pearson's correlation value per sample was performed. The correlation controlled the general transcripts expression consistency between samples. It could check sample mix-up or errors in name assignment. Thus, samples belonging to the same design group/condition were expected to show a higher level of correlation.

Gene expression was analyzed at the assembled transcript level using StringTie. Specifically, the count matrix was extracted from StringTie output using the auxiliary script prepDE.py as provided on the StringTie website.

It is generally beneficial to flag lowly expressed genes and filter them out before applying EDAseq, but importantly prior to differential expression analysis (to reduce the multiple testing burden). For this data set, we filtered genes that did not exceed 1 count per million in at least three samples. Out of the original transcripts, 47,059 transcripts were left after filtering. It is to be noted that one gene could have several different transcript isoforms due to alternative splicing or the use of varying initiation sites.

Each gene was tested for differential expression between WT and KO NPCs with an EdgeR LRT test. Due to the concern that the augmented Pol II pool caused by *Armc5* deletion might generally affect all gene transcription in the KO NPCs, the raw counts of each transcript were normalized by the ratio between the $\log_2$counts per million reads value of *Rn7sk* of a particular sample to the average *Rn7sk* $\log_2$counts per million reads value of across all samples. *Rn7sk* was transcribed by Pol III and was thus independent of the

putative influence of the Pol II pool size. This normalization was used instead of using edgeR::calcNormFactors, which uses a trimmed mean of M values normalization by default.

The heatmap was constructed using R pheatmap. The volcano plots and bar plots were produced using R v3.6.3. ggplot2. Based on a threshold for gene-level significance of 5% FDR, GO analysis of the RNA-seq data was performed using the Cytoscape v3.7.2 application ClueGO v2.5.6. The Uniprot Gene Ontology Annotations were used for the classification of the GO terms.

4.5.12. RT-qPCR

Total RNA from cells or tissues was extracted with Rneasy kit (Qiagen) and reverse-transcribed with iScriptTM cDNA Synthesis Kit (Bio-Rad Laboratories). The primer sequences are listed in **Table 4.S.4**. *Rn7sk* or *β-actin* was used as internal controls. The samples were first denatured at 95 °C for two minutes and then underwent 40 cycles of the following cycling condition: 95 °C for 15 seconds, 60 °C for 60 seconds, and finally with a melting step from 72 to 95 °C for 5 seconds. qPCR signals between 22 and 30 cycles were analyzed. Samples were assayed in triplicate, and the data were expressed as signal ratios of target mRNA/internal control mRNA.

4.5.13. ChIP-Seq

Three biological replicates of KO and WT NPCs, respectively, were washed with ice-cold PBS twice and resuspended in 1 ml PBS. They were crosslinked by adding 66.7 μl of 16% formaldehyde (1% final) at room temperature for 15 minutes. The reaction was quenched with 107 μl of 1.25 M glycine (0.125 M final) at room temperature for another 10 minutes in rotating tubes. The samples were centrifuged, and the pellets were washed twice with ice-cold PBS. The crosslinked pellets were suspended in 300 μl swelling buffer (25 mM HEPES, pH 7.5, 1.5 mM MgCl$_2$, 10 mM KCl, 0.1% NP-40) and incubated on ice for 20 minutes to release nuclei. The nuclei were harvested by centrifugation, resuspended in 200 μl ChIP sonication buffer (50 mM HEPES, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.1% Na-deoxycholate, 1%

Triton X-100, 0.1% SDS) and incubated on ice for 20 minutes. The nuclei were sonicated with a probe-based sonicator (model FB120, CL-18 probe; Fisher Scientific) at a 25% amplitude setting. The sonication was conducted using 30-second pulses at 30-second intervals for a total of 5 minutes. The sonicated nuclei were harvested by centrifugation. This sonicated chromatin was ready for immunoprecipitation.

To quantify chromatin and assess the degree of its fragmentation, we treated 5% of the sonicated nuclei (10 μl/sample) with 10 μg of RNase A for 15 minutes at 37 °C, followed by 20 μg of proteinase K for 30 minutes at 65 °C. They were quickly de-crosslinked for 5 minutes at 95 °C. DNA was extracted with the QIAquick PCR Purification Kit (Qiagen). DNA concentration was determined with a Nanodrop 1000 Fluorospectrometer. DNA fragment sizes were confirmed to be mainly within 100 to 800-bp range according to electrophoresis.

For immunoprecipitation, an equal amount (based on DNA measurements) of sonicated chromatin of different samples was reacted with mouse anti-POLR2A N-terminal domain mAb (clone D8L4Y, Cell Signaling Technology) (1:100) at 4 °C overnight, followed by 40 μl magnetic protein G beads (Bio-Rad) for another 2 hours at 4 °C. The beads were rinsed once with sonication buffer, once with wash buffer A (50 mM HEPES, pH 7.5, 500 mM NaCl, 1 mM EDTA, 0.1% Na-deoxycholate, 1% Triton X-100, 0.1% SDS), once with wash buffer B (20 mM Tris, pH 8.0, 250 mM LiCl, 1mM EDTA, 0.5% NP-40, 0.5% Na-deoxycholate) and then twice with TE buffer (10 mM Tris, pH 8.0, 1 mM EDTA). The chromatin was eluted with elution buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 1% SDS) at 65 °C for 10 minutes. The immunoprecipitated chromatins were de-crosslinked at 65 °C overnight with NaCl adjusted to 540 mM. The chromatins were then treated with 10-mg RNase A/sample at 37 °C for 1 hour, followed by 40 μg proteinase K per sample for 2 hours at 45 °C. DNA of the samples was purified with QIAquick PCR Purification kit (Qiagen) and quantified by the Bioanalyzer (Agilent).

Libraries were prepared robotically with 2 – 10 ng of fragmented DNA ranging from 100 to 300 bp in length, using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs), as per the manufacturer's recommendations. Adapters and PCR primers

were purchased from Integrated DNA Technologies. Size selection was carried out using SparQ beads (Qiagen) prior to PCR amplification (12 cycles). Libraries were quantified using the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems). Average fragment sizes were determined using a LabChip GX (PerkinElmer) instrument.

The libraries were normalized and pooled, and then denatured in 0.05 N NaOH and neutralized using HT1 buffer. The pool was loaded at 225 pM on an Illumina NovaSeq S4 lane (Illumina) using the Xp protocol as per the manufacturer's recommendations. The run was performed for 2×100 cycles (paired-end mode). A phiX library was used as a control and mixed with libraries at 1% level. The libraries were sequenced at 25 million reads per library. Base-calling was performed with RTA v3. Program bcl2fastq2 v2.20 was then used to demultiplex samples and generate fastq reads.

ChIP-seq reads were first trimmed for adapter sequences and low-quality score bases using Trimmomatic[395]. The resulting reads were mapped to the mouse reference genome (GRCm38 mm10) using BWA-MEM[403] in paired-end mode with default parameters. Only reads that had a unique alignment (mapping quality > 20) were retained, and PCR duplicates were marked using Picard tools. Peaks were called and annotated using MACS2[404] and HOMER[405] software suites, respectively.

To assess differences in Pol II occupancy patterns between WT and KO samples, we obtained ChIP-seq read counts within the following genomic regions using HOMER: the promoter region (from TSS (transcription start site) -400 bp to TSS +100 bp), the gene body (from TSS +100 bp to TES (transcription end site) -100 bp), the TES region (from TES -100 bp to TES +2,000 bp; also called the downstream region), the 5' untranslated region (5'-UTR), introns, 3'-UTR, enhancers (from TSS -5000 bp to TSS -400 bp), the region from -10,000 bp to TSS, the region from TSS to +10,000 bp, and the intergenic region. Since the POLR2A levels in the KO tissues were elevated, we speculated that there would be more Pol II association with the genes, hence a higher POLR2A ChIP signal in the KO promoter regions than in the WT counterparts. Therefore, genes that lacked POLR2A ChIP-seq signal in the KO tissues were filtered out, as these genes were believed to have no

signals in WT tissues neither. Raw counts were normalized using edgeR's trimmed mean of M algorithm[406] and were then transformed to $\log_2$ counts per million using the Voom function implemented in the Limma R package[407].

To construct the global metagene Pol II-binding profile, normalized read counts (Fragments Per Kilobase of transcript per Million Mapped reads of the full gene length plus 2,000-bp flanks (TSS -2,000 bp to TES +2,000 bp) were obtained from all genes that passed the filtering. Both flanks were divided into 20 equal-sized bins of 100 bp each. The gene bodies were scaled to 60 bins for the full gene length. FPKM was calculated from BAM input files using ngs.plot[408] with the following parameters: -G mm10 -R genebody -D ensembl -FL 200 -BOX 0 -SE 1 -VLN 0 -LWD 2 -WD 9. These global metagene Pol II binding profiles were only for visualization of differences in Pol II density, and inferential statistics was not conducted as per custom. The peak count versus distance (-10 kb to +10 kb from TSS) profile was generated from 51 equal-sized bins of 400 bp for this region of all genes that passed filtering. This profile was meant to give an overall view of the genomic location of all peaks, and inferential statistics was not conducted.

To test for differential Pol II density in WT and KO tissues, we used the R package Limma to fit a linear model on the $\log_2$CPM of the promoter region, the gene body, and the region downstream of the gene body. This model took into account the paired nature of our experiments. Nominal p-values were corrected for multiple testing using the Benjamini-Hochberg method[465] to obtain the FDR.

We calculated the pausing index for each gene by computing the ratio of Pol II signal density in the promoter region (from TSS -400 bp to TSS +100 bp) to signal density within the gene body (from TSS +100 bp to TES +2 kb) as follows: Pausing index (PI) = (Promoter region FPKM/L1)/(Genebody FPKM/L2). L1 is the length of the promoter region (always 500 bp), and L2 is the length of the gene body (variable).

Genome browser tracks were created with the HOMER makeUCSCfile command and bedGraphToBigWig utility from UCSC. Tracks were normalized so that each value represented the read count per base pair per 10 million reads. UCSC Genome Browser (http://genome.ucsc.edu/) was implemented for track visualization.

4.5.14. Nuclear run-on assay

Nuclear run-on assays were carried out according to a step-by-step protocol by Roberts et al.[466]. Briefly, nuclei from 4×10^6 KO or WT NPCs were collected and transcribed with Br-UTP and other NTPs. Nuclear RNA was extracted using the MEGAclear transcription clean-up kit (Life Technologies), and genomic DNA contamination was removed using the TURBO DNA-free kit (Life Technologies). Br-UTP-incorporated nascent transcripts were precipitated with anti-BrdU mAb (Santa Cruz Biotechnology), extracted, and reverse transcribed using a high-capacity cDNA reverse transcription kit (Invitrogen). qPCR was performed to quantify the nascent mRNA. To empirically determine the sensitivity of detecting nascent transcripts and the purity of Br-UTP-incorporated nascent transcripts over UTP-containing transcripts, before reverse transcription of the nuclear run-on reactions, we spiked the test samples with separately prepared control bacterial oligonucleotides with or without incorporated Br-UTP at known concentrations.

4.5.15. MM study population

A total of 511 subjects were selected for whole-exome sequencing from an MM study cohort enrolled from spina bifida clinics in five locations of North America between 1997 and 2010[467]. All the subjects were consented and enrolled in accordance with an institutional Internal Review Board at the University of Texas Health Science Center at Houston. In total, samples of 257 MM subjects of European descent comprising 140 females and 117 males, and 254 Mexican American MM subjects comprising 134 females and 120 males, were sequenced. Three hundred and sixty-five of the study subjects (over 70%) were born before January 1998, when the North American countries mandated folic acid fortification of food

crops. Sixty subjects were born in 1998, and eighty-six after the year 1998. Blood samples were collected from the subjects, and genomic DNA was extracted for the study.

4.5.16. Exome sequencing and variant annotation

Exome library probes were made from an in-house design based on TargetSeq (Invitrogen) with the addition of splice sites, UTRs, small non-coding RNAs (e.g., microRNAs), and a selection of miRNA binding sites, and 200-bp promoter regions. High-quality genomic DNA samples were processed using the exome library probes, and the captured DNA products were sequenced following the manufacturer's standard protocol for multiplexed sequencing using the P1 chip on the Ion Proton platform (Invitrogen). Quality of sequencing was maintained at 40 – 60 million reads/sample with read-length between 120 and 150 bases, and over 75% reads were on-target for all successfully sequenced samples. Other quality controls were implemented to map around 45,000 – 60,000 single-nucleotide variants (SNV) per sample with ∼50% heterozygote variants and the transition/transversion ratio around 2.5. Samples that failed to meet the above quality criteria were repeated or substituted by another subject's DNA.

Sequence data passed the above variant- and sample-quality filters were processed to call variants using Genome Analysis Toolkit HaplotypeCaller version 3.x following best-practice guidelines. Briefly, only variants designated a "PASS" by Variant Quality Score Recalibration and having a mapping quality score > 20 or inbreeding coefficient < -0.3, were retained for further analysis. Individual sample filters were used to ensure only high-fidelity variants with an alternate allele depth > 25%, a read depth > 10, and a genotype quality score > 20. The allele count, allele number, and allele frequency were recalculated for individual ethnicities after the filtering processes. Filtered high-quality SNVs were annotated using the non-synonymous SNV functional predictions database[468] with an in-house Python script for all currently functional prediction information publicly available. Further analyses were focused on single SNVs leading to stop-gained, stop-lost, non-synonymous, splice donor, and acceptor site changes in canonical transcripts.

4.5.17. Novel functional deleterious SNVs analysis

To analyze SNVs, we referred to AFs of variants observed in the non-Finnish European and Ad Mixed American populations of the genome aggregation database (gnomAD) Exome Controls[469]. Variants not observed in the non-Finnish European or Ad Mixed American gnomAD Exome Controls or having ethnic allele frequency = 0 were defined as novel SNVs (nSNVs). Datasets of non-Finnish Europeans or Ad Mixed Americans in gnomAD Exome Controls were downloaded for extracting alternate allele counts and total allele counts of all variants identified in MM subjects for comparison using the sample filters described previously[467]. For novel variants identified in subjects but not in gnomAD, we further verified that the loci were sequenced in gnomAD with $\geq 30\times$ coverage, and the corresponding variants were absent. Loci with $< 30\times$ coverage were considered as poor in quality and were discarded. These loci in the gnomAD controls were interpreted as having the reference alleles only, and the alternate allele frequency was considered zero.

nSNVs identified in the MM subjects were further verified by Sanger-sequencing. PCR primers franking 200 to 300 bases from the variants were designed to amplify the variant-containing loci from the MM subjects. The amplified loci were then sequenced.

Variants that had allele frequency in non-Finnish European or Ad Mixed American less than 0.01 were defined as rare, while allele frequency ≥ 0.01 was defined as common. Combined Annotation Dependent Depletion[430] (C-score) of variants was used as a model to predict deleteriousness. C-score is the -10 $\times$ log % rank of deleteriousness. A variant with a C-score of 13.01 is among the top 5% most deleterious variants, and a variant with a C-score of 20 among the top 1%.

For alternate allele counts between the MM subjects and gnomAD Exome Controls, odds ratios were calculated, and Fisher tests were performed.

Analysis of variants within *ARMC5* transcript (NM_001288767) for linkage disequilibrium (LD) was carried out using Idlink[470].

4.6. Data availability

The mouse RNA-seq and ChIP-seq datasets have been deposited to the Gene Expression Omnibus of NCBI (accession #GSE169350 and #GSE169582, respectively).

4.7. Ethics statement

All animal studies were approved by the Animal Protection Committee (Comité institutionnel d'intégration de la protection des animaux) of the CRCHUM. Human subjects were recruited with written consent to the research studies in accordance with an Institutional Review Board at the University of Texas Health Science Center at Houston.

4.8. Acknowledgments

4.9. Supplementary Figures

4.9.1. Supplementary Figure 1
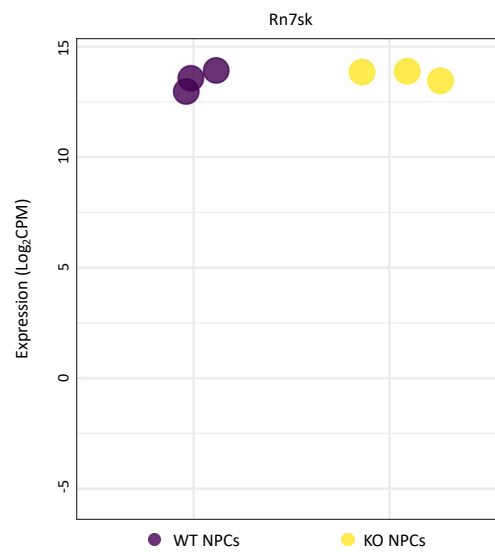
Fig. 4.S.1. *Rn7sk* expression in both KO and WT NPCs

4.10. Supplementary Tables

4.10.1. Supplementary Table 1

Table 4.S.1. Detailed parameters of differentially expressed transcripts in WT versus KO NPCs according to RNA-seq

| Gene name | Transcript id | logFC | FDR | Rawcounts | | | | | |
				KO1	KO2	KO3	WT1	WT2	WT3
Atxn2l	ENSMUST00000206577	-8.52	6.27E-13	62	23	28	10466	13651	12962
Sacs	ENSMUST00000119943	-15.02	1.23E-10	0	0	0	2383	7924	2675
Prdx6	ENSMUST00000156318	14.47	7.06E-10	2214	1428	4875	0	0	0
Tbkbp1	ENSMUST00000066078	-7.70	1.56E-09	1	1	1	121	325	291
H3f3a	ENSMUST00000081026	14.59	1.56E-09	12777	751	487	0	0	0
Lrrc47	ENSMUST00000030894	-7.62	1.81E-09	1	1	1	132	193	383
Pogk	ENSMUST00000128861	9.51	3.00E-09	764	385	623	0	1	1
Mrc2	ENSMUST00000100335	-8.68	3.00E-09	1	2	0	308	604	303
Xiap	ENSMUST00000055483	13.36	4.84E-09	1806	1120	1128	0	0	0
Ano10	ENSMUST00000214409	-13.28	4.97E-09	0	0	0	1267	1287	1074
Rars2	ENSMUST00000142045	-12.98	8.83E-09	0	0	0	1006	989	955
Pwwp3b	ENSMUST00000113045	10.54	9.31E-09	855	736	404	1	0	0
Itsn2	ENSMUST00000217981	12.78	1.02E-08	999	840	818	0	0	0
Lrif1	ENSMUST00000106736	-9.52	1.02E-08	0	1	0	252	416	219
Rps10	ENSMUST00000025052	-12.80	1.42E-08	0	0	0	1626	475	484
Zfp451	ENSMUST00000140163	-10.24	1.54E-08	0	0	1	462	560	530
Cep170	ENSMUST00000195433	-12.26	1.56E-08	0	0	0	784	689	275
Nudcd3	ENSMUST00000135631	-11.64	1.58E-08	0	0	0	332	431	429
Sucla2	ENSMUST00000022706	11.87	1.58E-08	575	615	273	0	0	0
Erlin1	ENSMUST00000112028	-12.90	1.58E-08	0	0	0	1273	1281	187
Spg20	ENSMUST00000118118	12.51	1.58E-08	734	329	1155	0	0	0
Slc25a53	ENSMUST00000145842	-9.28	1.62E-08	0	1	0	265	273	230
Stard5	ENSMUST00000075418	-11.57	1.62E-08	0	0	0	434	277	400
Ncaph	ENSMUST00000175885	-11.58	1.69E-08	0	0	0	236	584	351
Ears2	ENSMUST00000033159	-7.01	1.84E-08	1	1	1	71	171	223
Zfp661	ENSMUST00000077422	11.63	1.87E-08	398	282	505	0	0	0
Cux1	ENSMUST00000176172	-11.25	1.93E-08	0	0	0	456	207	205
Dis3	ENSMUST00000228643	11.33	1.96E-08	539	372	140	0	0	0
Pxdn	ENSMUST00000122328	5.98	1.96E-08	168	126	180	2	2	3
Spg20	ENSMUST00000117341	11.48	1.96E-08	378	230	463	0	0	0
Adam22	ENSMUST00000115388	12.16	2.00E-08	345	1257	245	0	0	0
Hacl1	ENSMUST00000127204	-10.66	2.04E-08	0	0	0	181	216	200
Anks6	ENSMUST00000229609	-10.64	2.30E-08	0	0	0	206	162	218
Tex2	ENSMUST00000207249	-10.63	3.00E-08	0	0	0	204	227	144
Ppara	ENSMUST00000109423	10.55	4.45E-08	245	143	185	0	0	0
P4ha1	ENSMUST00000092512	-11.41	5.93E-08	0	0	0	740	104	158
4930402H24Rik	ENSMUST00000138758	11.08	5.93E-08	132	144	527	0	0	0

Gene name	Transcript id	logFC	FDR	Rawcounts					
				KO1	KO2	KO3	WT1	WT2	WT3
Svbp	ENSMUST00000106345	5.21	7.51E-08	255	197	132	7	5	3
Terc	ENSMUST00000082862	-7.01	7.72E-08	1	1	1	110	288	75
Gm45140	ENSMUST00000204419	6.53	9.13E-08	723	931	1164	8	15	7
Btaf1	ENSMUST00000099494	5.50	1.04E-07	8118	3090	4125	106	67	145
Gtf2ird1	ENSMUST00000111244	-10.34	1.18E-07	0	0	0	135	166	185
Pcdh8	ENSMUST00000195355	10.43	1.18E-07	107	200	196	0	0	0
Skil	ENSMUST00000118470	-5.96	2.29E-07	26	1	3	1019	90	259
9130023H24Rik	ENSMUST00000078816	-6.75	2.93E-07	7	22	8	1266	1064	1159
Cox7a2	ENSMUST00000215933	-4.68	3.13E-07	17	9	8	314	233	273
Ssh1	ENSMUST00000112298	10.17	4.31E-07	113	171	142	0	0	0
Zeb2	ENSMUST00000200844	10.03	6.83E-07	141	129	122	0	0	0
Shank3	ENSMUST00000066545	4.85	1.35E-06	800	593	563	19	24	22
B130024G19Rik	ENSMUST00000190320	9.78	1.91E-06	158	87	95	0	0	0
Ncald	ENSMUST00000168992	5.63	2.11E-06	132	107	80	2	3	1
Exosc4	ENSMUST00000230512	-6.90	3.45E-06	5	0	3	281	267	256
Dbp	ENSMUST00000211513	-9.50	5.27E-06	0	0	0	100	75	88
Rbm6	ENSMUST00000194436	-8.58	5.56E-06	0	0	2	268	204	259
Dnah9	ENSMUST00000080665	5.32	7.20E-06	138	154	1238	5	3	25
Sema6d	ENSMUST00000077847	5.29	8.12E-06	1220	1001	1370	31	13	60
Vav2	ENSMUST00000185188	-4.28	1.18E-05	23	19	24	405	306	574
Tial1	ENSMUST00000033135	4.84	2.11E-05	2596	2043	1030	52	60	68
Armc5	ENSMUST00000044660	-4.73	2.17E-05	84	119	43	1964	2110	1796
Smcr8	ENSMUST00000102667	-4.48	2.90E-05	92	69	66	1689	1797	1417
Ahi1	ENSMUST00000105525	-4.88	1.26E-04	325	108	128	3417	5877	5224
Ralgps2	ENSMUST00000063199	-8.03	2.20E-04	4	6	0	883	341	918
Alg2	ENSMUST00000136912	3.97	2.50E-04	640	862	523	25	75	36
Ercc6l2	ENSMUST00000067821	3.40	3.22E-04	358	219	158	28	16	21
Gfpt1	ENSMUST00000113658	-4.47	5.35E-04	137	159	80	4813	2000	1827
Pld1	ENSMUST00000148827	3.90	5.64E-04	282	158	129	7	12	17
Gm10167	ENSMUST00000150411	4.03	6.90E-04	1964	1588	905	59	116	86
Cenpk	ENSMUST00000224500	-6.76	7.91E-04	0	10	4	369	509	258
Mcc	ENSMUST00000089874	-3.62	8.74E-04	23	14	4	170	143	107
Pclo	ENSMUST00000182915	5.33	9.05E-04	119	193	720	0	8	33
Chp1	ENSMUST00000137585	-3.95	1.19E-03	8	16	3	141	129	84
Cbs	ENSMUST00000118504	-3.45	1.68E-03	7	11	7	73	89	106
Glrx2	ENSMUST00000145571	-3.32	1.72E-03	9	14	34	107	154	256
Acbd5	ENSMUST00000227809	3.65	2.08E-03	127	92	89	8	4	13
Dnajc14	ENSMUST00000219508	-3.82	2.60E-03	46	8	3	247	167	85
Slx1b	ENSMUST00000144897	2.88	2.87E-03	222	226	95	23	28	17
Oxr1	ENSMUST00000229769	-3.41	4.25E-03	17	31	9	232	168	142
Pogk	ENSMUST00000169324	3.66	4.25E-03	8725	3373	6267	404	503	420
Pole3	ENSMUST00000128100	-9.00	4.41E-03	0	0	7	714	231	585
Elmo2	ENSMUST00000103088	4.48	4.42E-03	355	993	841	35	37	19
Chchd10	ENSMUST00000219839	-2.81	5.13E-03	19	31	37	123	201	287
Rnf2	ENSMUST00000187048	-4.05	7.60E-03	23	17	27	184	178	1432

277

Table 4.S.1 continued from previous page

Gene name	Transcript id	logFC	FDR	Rawcounts					
				KO1	KO2	KO3	WT1	WT2	WT3
Aldh7a1	ENSMUST00000066208	3.30	8.09E-03	2618	1977	1205	181	251	132
Slc25a53	ENSMUST00000154484	2.98	1.20E-02	296	262	712	36	31	93
Ajuba	ENSMUST00000054487	-4.22	1.20E-02	7	25	4	283	151	108
Rps13	ENSMUST00000205490	3.13	1.21E-02	6466	2921	3031	596	406	321
Slc8a1	ENSMUST00000234923	-3.34	1.30E-02	18	29	56	223	386	329
Gm6311	ENSMUST00000073790	-3.58	1.39E-02	424	220	91	2331	2333	2485
Nisch	ENSMUST00000168206	-3.03	1.46E-02	822	662	594	5933	4063	6657
Tmem123	ENSMUST00000052865	-2.91	1.47E-02	1759	1308	892	8895	9089	9939
Cd151	ENSMUST00000177840	-5.84	1.48E-02	71	0	10	750	154	776
Pkp4	ENSMUST00000184332	-3.14	1.86E-02	64	63	63	647	344	725
BC030867	ENSMUST00000133930	2.83	1.93E-02	277	129	91	20	18	23
Pla2g4a	ENSMUST00000111926	6.45	1.94E-02	272	638	73	5	0	3
Actn4	ENSMUST00000127210	4.67	2.19E-02	743	573	56	7	80	2
8430429K09Rik	ENSMUST00000146456	-5.72	2.21E-02	18	7	0	207	284	285
Phf8	ENSMUST00000112666	-3.15	2.21E-02	40	28	77	353	370	432
5330417C22Rik	ENSMUST00000106625	2.70	2.30E-02	87	125	416	8	20	89
Ehmt2	ENSMUST00000097342	5.20	2.45E-02	11796	9227	5787	1331	59	161
Mark2	ENSMUST00000025921	-4.98	2.70E-02	3	12	38	348	440	311
Arpp19	ENSMUST00000166549	-3.04	2.88E-02	181	131	67	948	889	985
Gm10320	ENSMUST00000170205	-3.05	2.97E-02	42	35	16	262	157	308
Ifi44	ENSMUST00000029671	5.10	3.29E-02	191	136	413	14	0	18
Glrx3	ENSMUST00000064404	-2.91	3.39E-02	210	128	107	1037	1189	911
Tmem176b	ENSMUST00000166247	3.24	3.39E-02	1689	1055	512	163	59	102
Slx1b	ENSMUST00000084586	-2.56	3.39E-02	50	42	54	209	287	361
Mroh1	ENSMUST00000162204	3.08	3.67E-02	189	179	126	37	11	15
Farsb	ENSMUST00000068333	-2.97	3.89E-02	26	20	22	75	253	267
Fam172a	ENSMUST00000091459	4.38	4.23E-02	1645	1977	2072	143	20	191
Fam172a	ENSMUST00000163257	-2.86	4.50E-02	21	14	23	114	187	115
Pak6	ENSMUST00000099557	-2.60	4.73E-02	162	28	13	697	192	92

Note: RNA-seq of WT and KO NPCs was performed, and differentially expressed transcripts with FDR < 0.05 are listed. Detailed parameters of each transcript are presented. For the calculation of fold change (FC), FDR and p-values, the rawcounts were normalized by the ratio between the $\log_2$CPM value of Rn7sk of a particular sample to the average of Rn7sk $\log_2$CPM (counts per million reads) values across all samples.

4.10.2. Supplementary Table 2

Table 4.S.2. GO analysis of dysregulated genes in terms of biological process

GO ID	GO Terms: Biological process	*p*-Value	Associated Genes Found
GO:0006903	vesicle targeting	0.00	Ahi1, Alg2, Clasp2, Pclo
GO:0009395	phospholipid catabolic process	0.04	Pla2g4a, Pld1, Prdx6
GO:0043967	histone H4 acetylation	0.05	Hcfc2, Kansl3, Kmt2a
GO:0051187	cofactor catabolic process	0.05	Blvra, Prdx6, Pxdn
GO:0099558	maintenance of synapse structure	0.04	Chchd10, Pclo, Shank3
GO:0019210	kinase inhibitor activity	0.05	Chp1, Smcr8, Spred2
GO:0004860	protein kinase inhibitor activity	0.05	Chp1, Smcr8, Spred2
GO:0044088	regulation of vacuole organization	0.04	5330417C22Rik, Ehmt2, Smcr8
GO:2000785	regulation of autophagosome assembly	0.03	5330417C22Rik, Ehmt2, Smcr8
GO:0006338	chromatin remodeling	0.00	Baz1a, H3f3a, Hcfc2, Ino80, Pole3, Smarca2, Smarcc2
GO:0043044	ATP-dependent chromatin remodeling	0.05	Ino80, Smarca2, Smarcc2
GO:0045454	cell redox homeostasis	0.03	Glrx2, Glrx3, Prdx6
GO:0009055	electron transfer activity	0.02	Cox7a2, Glrx2, Glrx3
GO:0016279	protein-lysine N-methyltransferase activity	0.05	Ehmt2, Kmt2a, Rbbp5
GO:0042054	histone methyltransferase activity	0.05	Ehmt2, Kmt2a, Rbbp5
GO:0018024	histone-lysine N-methyltransferase activity	0.05	Ehmt2, Kmt2a, Rbbp5
GO:0043038	amino acid activation	0.05	Ears2, Farsb, Rars2
GO:0043039	tRNA aminoacylation	0.05	Ears2, Farsb, Rars2
GO:0006418	tRNA aminoacylation for protein translation	0.04	Ears2, Farsb, Rars2
GO:0031047	gene silencing by RNA	0.04	Ajuba, Cnot1, Fam172a, Pum1
GO:0016441	posttranscriptional gene silencing	0.04	Ajuba, Cnot1, Pum1
GO:0035194	posttranscriptional gene silencing by RNA	0.05	Ajuba, Cnot1, Pum1
GO:0035195	gene silencing by miRNA	0.04	Ajuba, Cnot1, Pum1
GO:0048864	stem cell development	0.02	Alg2, Fam172a, Rab11fip3, Sema6d, Zeb2
GO:0014031	mesenchymal cell development	0.02	Alg2, Fam172a, Rab11fip3, Sema6d, Zeb2
GO:0014033	neural crest cell differentiation	0.02	Alg2, Fam172a, Rab11fip3, Sema6d, Zeb2
GO:0014032	neural crest cell development	0.02	Alg2, Fam172a, Rab11fip3, Sema6d, Zeb2
GO:0001755	neural crest cell migration	0.04	Rab11fip3, Sema6d, Zeb2
GO:0016331	morphogenesis of embryonic epithelium	1.00	PCDH8,ZEB2,SHANK3
GO:0043065	positive regulation of apoptotic process	1.00	SKIL, PLA2G4A, gadd45b
GO:0007018	microtubule-based movement	1.00	DNAH9
GO:0001775	cell activation	1.00	CD151, PLA2G4A, Tgfb1
GO:0006357	regulation of transcription by RNA polymerase II	1.00	ZFP451, ZFP661, CUX1, PPARA, GTF2IRD1, SKIL, ZEB2, DBP, CENPK, RNF2, AJUBA, EHMT2, MAFA
GO:0006955	immune response	1.00	TBKBP1,PLA2G4A,IFI44
GO:0000122	negative regulation of transcription by RNA polymerase II	1.00	ZFP451, PPARA, GTF2IRD1, SKIL, ZEB2, RNF2, AJUBA, EHMT2, Irf8
GO:0051726	regulation of cell cycle	1.00	SKIL, PKP4, CDKN1A

4.10.3. Supplementary Table 3

Table 4.S.3. Statistical analysis of normalized POLR2A read counts of different gene regions in WT and KO NPCs

Gene	Region	FC	log$_2$FC	*p-value*	FDR
TSS					
Taf4b	Taf4b_TSS	1.77	0.83	1.03E-05	3.39E-02
Arrdc2	Arrdc2_TSS	2.83	1.50	1.24E-05	3.39E-02
Dzip3	Dzip3_TSS	1.99	0.99	2.98E-05	6.14E-02
Chchd2	Chchd2_TSS	2.20	1.14	3.03E-05	6.14E-02
Rrp1b	Rrp1b_TSS	2.58	1.37	3.21E-05	6.14E-02
Rdm1	Rdm1_TSS	2.02	1.01	3.79E-05	6.53E-02
Rab11fip3	Rab11fip3_TSS	1.60	0.68	5.50E-05	7.85E-02
Vmac	Vmac_TSS	2.32	1.22	6.90E-05	7.85E-02
Gtf3c6	Gtf3c6_TSS	1.61	0.69	6.90E-05	7.85E-02
Hectd4	Hectd4_TSS	1.67	0.74	7.94E-05	7.93E-02
Tmsb4x	Tmsb4x_TSS	1.98	0.99	8.23E-05	7.93E-02
Mphosph9	Mphosph9_TSS	1.73	0.79	8.88E-05	8.03E-02
Gm10719	Gm10719_TSS	15.37	3.94	9.14E-05	8.03E-02
Tmco6	Tmco6_TSS	1.77	0.82	9.20E-05	8.03E-02
Qsox2	Qsox2_TSS	2.55	1.35	9.54E-05	8.03E-02
Tm7sf3	Tm7sf3_TSS	1.90	0.92	1.03E-04	8.08E-02
Teddm2	Teddm2_TSS	1.79	0.84	1.16E-04	8.67E-02
Hist1h4a	Hist1h4a_TSS	1.58	0.66	1.22E-04	8.75E-02
Upf3a	Upf3a_TSS	1.54	0.62	1.27E-04	8.94E-02
Rgs4	Rgs4_TSS	1.89	0.92	1.34E-04	9.06E-02
Ep400	Ep400_TSS	1.61	0.69	1.65E-04	9.84E-02
Emc8	Emc8_TSS	2.77	1.47	1.66E-04	9.84E-02
Nefm	Nefm_TSS	1.88	0.91	1.69E-04	9.85E-02
GENEBODY					
Hist2h2aa2	Hist2h2aa2_GENEBODY	4.89	2.29	1.05E-07	3.62E-03
Tex14	Tex14_GENEBODY	0.60	-0.73	2.86E-07	4.93E-03
Arc	Arc_GENEBODY	2.01	1.01	1.48E-06	1.70E-02
Gadd45b	Gadd45b_GENEBODY	2.24	1.17	3.81E-06	3.28E-02
Hist1h4n	Hist1h4n_GENEBODY	2.09	1.06	6.70E-06	3.39E-02
Cdkn1a	Cdkn1a_GENEBODY	1.44	0.52	8.16E-06	3.39E-02
Baiap2	Baiap2_GENEBODY	1.40	0.48	1.11E-05	3.39E-02
Rhob	Rhob_GENEBODY	1.54	0.62	1.14E-05	3.39E-02
Dgat2	Dgat2_GENEBODY	1.62	0.70	1.25E-05	3.39E-02
Hist1h1a	Hist1h1a_GENEBODY	2.35	1.24	1.28E-05	3.39E-02
Pcdh8	Pcdh8_GENEBODY	2.27	1.19	2.07E-05	5.11E-02
Nefm	Nefm_GENEBODY	1.55	0.63	2.33E-05	5.35E-02
Srsf7	Srsf7_GENEBODY	1.43	0.52	3.55E-05	6.44E-02
Jund	Jund_GENEBODY	1.47	0.56	4.30E-05	7.06E-02
Slc20a1	Slc20a1_GENEBODY	1.42	0.50	5.13E-05	7.85E-02

Table 4.S.3 continued from previous page

Gene	Region	FC	log$_2$FC	*p-value*	FDR
Klf4	Klf4_GENEBODY	1.85	0.88	6.15E-05	7.85E-02
Hist1h4c	Hist1h4c_GENEBODY	1.76	0.81	6.39E-05	7.85E-02
Cebpa	Cebpa_GENEBODY	1.79	0.84	6.44E-05	7.85E-02
Hist1h3a	Hist1h3a_GENEBODY	2.25	1.17	6.78E-05	7.85E-02
Hist1h3i	Hist1h3i_GENEBODY	2.17	1.12	6.96E-05	7.85E-02
Hist1h2ai	Hist1h2ai_GENEBODY	2.20	1.14	7.07E-05	7.85E-02
Hist1h3h	Hist1h3h_GENEBODY	2.58	1.37	7.67E-05	7.93E-02
Ttyh1	Ttyh1_GENEBODY	0.71	-0.49	9.73E-05	8.03E-02
Adcyap1r1	Adcyap1r1_GENEBODY	0.76	-0.40	9.79E-05	8.03E-02
Mafa	Mafa_GENEBODY	2.44	1.28	1.03E-04	8.08E-02
Rgs2	Rgs2_GENEBODY	1.69	0.76	1.10E-04	8.39E-02
Ier3	Ier3_GENEBODY	1.84	0.88	1.22E-04	8.75E-02
Ier2	Ier2_GENEBODY	1.44	0.53	1.33E-04	9.06E-02
Hist1h2an	Hist1h2an_GENEBODY	2.39	1.26	1.44E-04	9.42E-02
Tob1	Tob1_GENEBODY	1.50	0.59	1.47E-04	9.42E-02
Hist1h3f	Hist1h3f_GENEBODY	1.46	0.55	1.48E-04	9.42E-02
Mat2a	Mat2a_GENEBODY	1.37	0.45	1.54E-04	9.62E-02
Rplp1	Rplp1_GENEBODY	1.57	0.65	1.57E-04	9.67E-02
TES					
Slx1b	Slx1b_TES	2.10	1.07	1.02E-05	3.39E-02
Hist1h3a	Hist1h3a_TES	1.50	0.58	7.54E-05	7.93E-02
Armh1	Armh1_TES	2.42	1.28	8.29E-05	7.93E-02

Note: A linear model was used to fit the log$_2$CPM of the promoter region, gene body, and region downstream of gene body of WT and KO NPCs. This model took into account the paired nature of our experiment. Nominal p-values were corrected for multiple testing using the Benjamini-Hochberg method to obtain the FDR. All the regions of genes with FDR <0.1 were included.

4.10.4. Supplementary Table 4

Table 4.S.4. RT-qPCR primer sequences

Name	5'→3'
Primers for nuclear run-on RT-qPCR assays	
Dnah9, Forward	CCT GCC CCA GCT GTC CCT TAG T
Dnah9, Reverse	CCT TGC ACC CAG ACC TTC GCT G
Ifi44, Forward	GGT CCT CTG AGC CGA CCA GTC A
Ifi44, Reverse	AGG ACC CAG CAG CAG AAC TCG T
Irf8, Forward	AAG AGC CCC TGC TGA GGT CAG G
Irf8, Reverse	TGC TGA GGA CTG GAC CCA GAG C
Tgfb1i1, Forward	ATG AGC GTG TGA GTG CAC CAG C
Tgfb1i1, Reverse	CAG CGA GCA TCA CGG AAC AGG G
ACTB, Forward	AGC TCA TTG TAG AAG GTG TGG
ACTB, Reverse	GGC ATG GGT CAG AAG GAT TC
GAPDH, Forward	AAT CCC ATC ACC ATC TTC CAG
GAPDH, Reverse	GAG CCA CAC CAT CCT AGT TG
RPL10. Forward	TTA TCA TGT CCA TCC GCA CC
RPL10, Reverse	GCA TTG AAC TTG GTG AAG CC
RPLP0, Forward	CGC AGC CAA TAG ACA GGA G
RPLP0, Reverse	GCG CGT GCC TTT TAT AAT GC
UBC, Forward	GCC TTA GAA CCC CAG TAT CAG
UBC, Reverse	AAG AAA ACC AGT GCC CTA GAG
Primers of mouse mRNA RT-qPCR	
7skRNA, Forward	TCA CCC CAT TGA TCG CCA GGG T
7skRNA, Reverse	CAC ATG GAG CGG TGA GGG AGG A
Polr2a, Forward	CAC TGT CAT CAC CCC TGA CC
Polr2a, Reverse	ATA CTG GCT GTT TCC CCT GC
Cdkn1a, Forward	TGC CCA AGG TCT ACC TGA GCC C
Cdkn1a, Reverse	AGT GCA AGA CAG CGA CAA GGC C
Dnah9, Forward	AGC CAG TCT CAG ATG CCA TAG A
Dnah9, Reverse	CCA GGA GTT CAG CCA TTC CC
Gadd45b, Forward	ACG CGG TTC AGA AGA TGC AGG C
Gadd45b, Reverse	GGA GGC ACA AGA CCA CGC TGT C
Mafa, Forward	ATT CTG GAG AGC GAG AAG TGC CAG

Table 4.S.4 continued from previous page	
Name	**5'→3'**
Mafa, Reverse	CGC CAA CTT CTC GTA TTT CTC CTT
Pcdh8, Forward	CAT GCA GAG TGG ACT GTG GGC G
Pcdh8, Reverse	TAC GTT GGG TCC GGC ACA GGA T
Folh1, Forward	GCC ACC ATA CAG TGC CTT CT
Folh1, Reverse	TGG CAA TCA CAA TCT TCC CAG A
Primers of human mRNA RT-qPCR	
MAFA, Forward	CTC CTC GCT CAT TCG CTC TG
MAFA, Reverse	GGA CCA ACA CGC AGG TAA GT
FOLH1, Forward	AAC TGG ACC CCA GGT CTG GAG C
FOLH1, Reverse	CCT GTG CTG CTG CTC TAC TGC G
CDKN1a, Forward	TGG CTC CTT CCC ATC GCT GTC A
CDKN1a, Reverse	GGT GAG GCC CCT TCA AAG TGC C
GADD45B, Forward	AGG CCC GAG ACC TGC ATT GTC T
GADD45B, Reverse	TCT TCG CAG TAG CTG CCA CCT
IFI44, Forward	ACG CTG GTG TGG TAC ATG TGG C
IFI44, Reverse	GCT TGG ACC TCA CAG CTC ACA
IRF8, Forward	TGC CTG TGG AGG GGA CAG TCA G
IRF8, Reverse	ACG TCC CCT GGT CCT CTC CTC T
PCDH8, Forward	ACG ACA GCG ATT CCG ACA TCA G
PCDH8, Reverse	GCT CCA GCA GCG GTC AGA GTG

Chapter 5

Discussion

In this project, we discovered that ARMC5-CUL3-RBX1 formed an E3 complex targeting RPB1 and controlling its degradation. *Armc5* KO/mutation caused abnormally enlarged Pol II pool size due to compromised RPB1 degradation, leading to the pathogenesis of PBMAH, NTDs, and dysfunctional T cell immune responses.

5.1. The Pol II pool size and transcriptome

5.1.1. The Pol II pool size

Pol II homeostasis is critical in cell biology. Recent studies by Tufegdžić Vidaković et al.[264] and Nakazawa et al.[263] showed that the Pol II pool size could influence the overall transcriptome. The Pol II pool size is largely decided by the abundance of the largest Pol II subunit, RPB1, which harbors the DNA template-binding jaw and catalytic activity. Thus, RPB1 degradation is critical for the Pol II pool size.

A single ubiquitination site, K1268, is necessary and sufficient for RPB1 proteasomal degradation in response to UV irradiation[263,264]. The failed degradation of K1268R mutant cells results in an enlarged Pol II pool upon acute DNA damage, which causes activation of certain genes and long-term transcription defects persistently[264].

On the other hand, a short-term depletion of RPB1 in human DLD-1 colorectal cancer cells causes a rapid change of the nascent transcriptome[471], and more than 90% of the ~600

affected genes were downregulated, most of them being involved in transcription, chromatin binding, and cell cycling regulation[471]. In some infectious disease, nsP2 proteins from the alphavirus, such as *Sindbis virus*, *Semliki Forest virus*, and *Chikungunya viruses*, inhibit cellular transcription by rapid degradation of Rpb1. Complete degradation of Rpb1 occurs after six hours of alphavirus infection. The depletion of the Rpb1 pool, in this case, results in slowing down or even turning off the entire cellular transcriptional machinery[472].

In the case of enlarged Pol II pool due to *Armc5* KO (our study) in the adrenal glands and *RPB1* K1268R mutation in yeasts, the larger pool caused predominantly augmented upregulation in a large number of genes, although its effect on NPCs did not manifest bias to upregulation. The large pool affects a different set of genes in different tissues. Probably due to such cell-type specificity, the enlarged Pol II due to *Armc5* deficiency causes distinct diseases in different organs, e.g., PBMAH in the adrenal glands where cells slowly overgrow, NTDs and the compromised T cells where cells showed reduced growth, and meningioma, a slow-growing tumor in most cases. Therefore, Pol II pool size plays a vital role in cell biology and pathogenesis, and in many cases, its effect is not universal to all genes.

5.1.2. How does an enlarged Pol II pool change gene transcription?

How does an enlarged or reduced Pol II pool preferentially affect a subset of genes? This is a question needs answers.

The transcription is finely regulated through the coordinated interaction among Pol II, regulatory elements in DNA, and many transcription factors. The core promoter region contains various DNA elements. The TATA box is the most well-known one that is recognized by the TBP subunit of the TFIID complex. However, it only exists in $10 - 20\%$ of metazoan genes.

We did a preliminary TATA box analysis with the tools provided by Eukaryotic Promoter Database. The results are summarized in **Table 5.1**. Among 1481 promoters from the upregulated (FDR < 0.05) genes (excluding predicted genes or miRNA) in the RNA-seq result of *Armc5* KO adrenal glands, 216 of them (14.5%) contain TATA boxes. There is

about 18% of the 1000 randomly selected genes (FDR > 0.05) that contain TATA boxes. These percentages are in a range compatible with that of TATA box-containing genes in the mouse genome. However, when we set a stricter threshold ($\log_2$FC > 2 and FDR < 0.01) of RNA-seq results, a significantly higher number of promoters (49 out of 116 genes; 42.2%) contain TATA boxes. Furthermore, among the genes with both upregulated mRNA according to RNA-seq (FDR < 0.05; Log_2FC > 2) and Pol II density (FDR < 0.05) at the TSS according to ChIP-seq, 31 out of 45 genes contain TATA boxes in their promoters, and this represents an even higher percentage (68.9%).

Table 5.1. The TATA box analysis of upregulated genes from *Armc5* KO adrenal glands

Gene datasets	Total promoter number	The number of promoters containing TATA box	Percentage
RNA-seq ↑ (FDR < 0.05)	1481	216	14.5% (216/1481)
RNA-seq ↑ (FDR < 0.01, logFC > 2)	116	49	42.2% (49/116)
RNA-seq ↑ and ChIP-Seq ↑ (FDR < 0.05)	45	31	68.9% (31/45)

Note: The analyzed genes were selected from RNA-seq and ChIP-seq results of *Armc5* WT/KO adrenal glands. The promoter region was set to -499 to +100. The TATA box was predicted by Eukaryotic Promoter Database (`https://epd.epfl.ch/EPDnew_select.php`).

The step-wisely increased stringency in selecting the genes as described above probably eliminates false-negative ones in both RNA-seq and ChIP-seq, and the selected ones at the highest stringency (FDR < 0.05, Log_2FC > 2 for RNA-seq and FDR < 0.05 for ChIP-seq) are likely the true upregulated genes with increased Pol II density in their promoters. This implies that the presence of a TATA box in its promoter is a condition favoring the upregulation of the gene when there is an enlarged Pol II pool.

However, the question still remains as to why only a subset of TATA box-containing genes were increased instead of all. We can speculate that other regulatory elements are needed in addition to the TATA box in response to an enlarged Pol II pool. More DNA motifs, such as BREu, BREd, MTE, DPE, and DCE, within the core promoter, have been revealed, as mentioned in Introduction. In addition, the TFIIA response element (IIARE)

could also enhance the activity of the TATA box-containing promoter[473]. These motifs that are often present near the TATA box may contribute to the gene responses to an enlarged Pol II pool. Tissue-specific enhancers also play roles in determining whether a TATA-box containing gene can be expressed, not to mention being upregulated. We could select some confirmed highly upregulated genes, such as *Mafa*, *Pcdh8*, and *Actn2*, in the KO tissues as gene models in the next phase of investigation to fully elucidate the mechanisms by which an enlarged Pol II modulates gene expression.

5.2. The effector genes causing PBMAH

The accumulated RPB1 protein causes the dysfunction of Pol II. *Rpb1* is not an oncogene that leads to tumorigenesis of PBMAH or meningiomas. More likely, the abnormal Pol II pool affects a subset of genes in different tissues. Those downstream genes could be the real culprits for these diseases.

As mentioned in Introduction (**Figure 1.3**), adrenocortical tumors are highly related to the cAMP-PKA signaling pathways. Under a physiological condition, PKA remains in a resting state as an inactive tetramer comprising two regulatory subunits and two catalytic subunits. Adenylyl cyclase catalyzes the convention of ATP to cAMP. The higher level of cAMP activates PKA by releasing its catalytic subunits, which phosphorylates the transcription factor cAMP response-element binding protein (CREB). Phosphodiesterases (PDEs) inhibit cAMP-PKA signaling by degrading cAMP to AMP[474].

According to the RNA-seq of *Armc5* KO adrenal glands, we found that several genes involved in the cAMP-PKA pathway were upregulated. The different isoforms of mouse tetramer PKA subunits are encoded by *Prkar1a*, *Prkar1b*, *Prkar2a*, *Prkar2b*, *Prkaca*, *Prkacb*, *Prkaa1*, *Prkaa2*, *Prkab1*, and *Prkab2*. *Prkar1b*, which encodes a regulatory subunit, is significantly increased.

Among the genes encode PDE4, PDE7, and PDE8, which are able to degrade cAMP, *Pde4a*, *Pde4b*, *Pde4d*, and *Pde8b* are increased notably. These proteins act as negative

regulators in the cAMP-PKA signaling, resulting in the downregulation of CREB-related transcription pathways.

In addition, *Tfcp2l1*, *Pcdh8*, *Mafa*, and *Taf4b*, which are genes related to the regulation of proliferation and are normally silenced in normal adrenal glands, are dramatically upregulated.

Conceivably, we believe that the collective effect of these dysregulated effector genes found in the KO adrenal glands, regardless of the dysregulation being direct or indirect due to the enlarged Pol II, drives the adrenocortical cells to benign slow-growing tumors instead of fast-growing malignant ones. Additional investigation of these effector genes in the pathogenesis of PBMAH will be warranted in the next phase of study.

5.3. Diseases associated with accumulated RPB1

As we know now, the accumulation of RPB1 caused by the defect of ARMC5 is highly related to PBMAH, NTDs, and meningiomas[1,5,19]. The elevated level of RPB1 is related to other diseases as well, although it is not sure whether it is caused by the failed degradation of RPB1 or increased *RPB1* transcription.

Here are some examples. In lung adenocarcinoma, RPB1 protein increased in the cancer tissue compared to the normal adjacent tissues[475]. RPB1 was also accumulated and mislocalized in the cytoplasm along with the pathological Tau protein in Alzheimer's disease. The failed degradation of misfolded RPB1 or the failed assembly of Pol II complex leads to the mislocalization of RPB1[476]. The abnormal Pol II pool size could fundamentally affect the normal function of neuron cells. In addition, α-amanitin, a toxin from mushrooms, directly interacts with Pol II, resulting in the stalling of Pol II and decreased expression of all mRNA[477].

Enhancing the activity of ARMC5 may reduce the pathological Pol II and serve as a potential therapeutic strategy to treat these diseases.

5.4. The degrons in RPB1 recognized by ARMC5

A specific feature of the substrate that selectively recognized by E3s is termed "degron", which is defined as the minimal element within the substrate sufficient to serve as a degradation signal. Although degrons can locate anywhere in the protein sequence, the studies of degrons are still focusing on the N-terminus and the C-terminus degron pathways.

Recent studies show that RING E3s recognize substrates by different degron pathways. The acetylated N-terminus residues can serve as a degron element for MARCH6 E3 and Doa10 E3[478,479]. The Pro/N-degron pathway and the Gly/N-degron pathway refer to a Proline after Methionine (MP-) and a Glycine after Methionine (MG-), which can be recognized by the GID E3[480] and ZYG11B-CUL2 E3[481], respectively.

The study of the C-degron pathway remains in its infancy now. Generally, the C-degron signal is targeted by the recognition unit in the E3 complex, such as ARMC5 in CUL3 E3, KLHDC in CUL2 E3, and TRPC4AP in CUL4 E3. These adaptors contain tandem repeat domains to facilitate degron recognition. Kelch repeats in KLHLDC2 prefer the Gly/C-degron signal[482], while TPR repeats in APPBP2 recognize RxxGx/C-degron motifs[483]. Although there is no study about the degron pattern targeted by Armadillo repeats, TRPC4AP, an adaptor protein containing Armadillo-like repeats, requires arginine at the C-terminus (-Rxx motif) as a degron signal for CUL4 E3[484]. This suggests the existence of a degron pattern which can be recognized by the Armadillo repeats in ARMC5.

The N-terminus of RPB1 protein sequence is "MHGGGPPS-", which does not show a clear pattern fitting the known N-degron pathways, except the acetylation of its first methionine possibly serving as the signal. The C-terminus protein sequence of RPB1 with or without the CTD is "-MTVTRRTL" or "-HQVVPCLP". Thus, we can speculate that the Rxx/C-degron at the end of CTD or an unknown C-degron before CTD may serve as the degradation signal to ARMC5-CUL3-RBX1 E3. Moreover, the degrons of RPB1 may exist anywhere within the protein, as long as they locate at the accessible sites for recognition or modification.

5.5. The PROTACs hijack the ubiquitin-proteasome system

5.5.1. Proteolysis-Targeting Chimeras

The Proteolysis-Targeting Chimeras (PROTACs) technology hijacks the UPS to degrade specific proteins. This approach employs a heterobifunctional molecule that consists of two subunits, one each binding to an E3 and the targeted protein, connected with a linker, to direct the ubiquitination of a "neo-substrate". Besides, some small molecules are also able to play the same role as PROTACs to glue the protein of interest to the E3. The neo-substrate is then ubiquitinated and degraded by proteasome. The schematic diagram is shown in **Figure 5.1**.
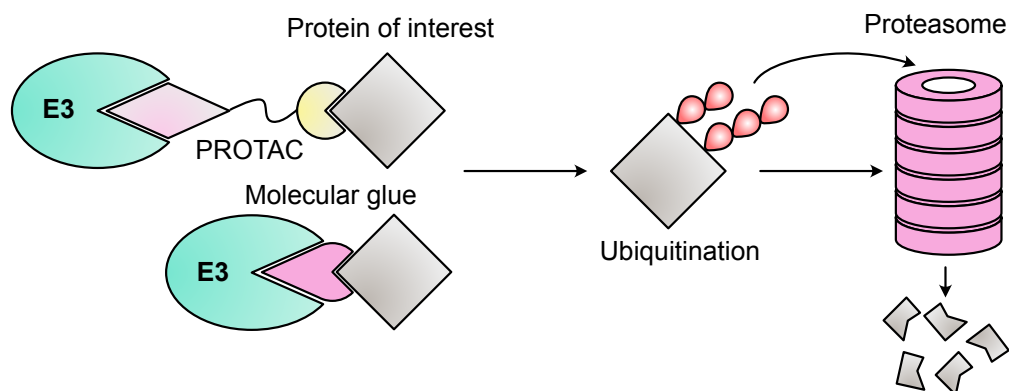

Fig. 5.1. The Proteolysis-Targeting Chimeras (PROTACs) technology. The PROTACs and molecular glues hijack the UPS to degrade the protein of interest.

Beginning with a conceptual study of artificially directing methionine aminopeptidase 2 (METAP2) protein degradation by the SCFβ-TRCP-CUL1 E3 in the cell lysate[485], this technology has dramatically advanced over the past twenty years and has been applied in cells, animal models, and clinical trials[486].

Molecular glues are initially developed from thalidomide, which was a drug to treat sleep disorders in the 1950s. It has now been proved its association to CRBN-DDB1-CUL4 E3 complex. PROTACs or other molecular glues based on thalidomide or its analog pomalido-mide have been generated to direct the protein of interest for degradation[487,488,489,490,491].

However, the discovery of molecular glues mostly relies on serendipity. PROTACs are more rational designed molecules based on the protein sequence and structure. With a good E3 recognition subunit designed to associate to the VHL-ElonginB-ElonginC-CUL2 E3 complex, many VHL-based PROTACs have been developed, targeting on neo-substrates, including androgen receptor[492], RIPK2[493], TBK1[494], FAK[495], TRIM24[496], and SMARCA2/4[497].

Although many potential PROTACs against dozens of disease-associated proteins have been discovered recently, most of them are similarly designed based on CRBN-DDB1-CUL4 E3 and VHL-ElonginB-ElonginC-CUL2 E3. There are more than 600 E3s in the human genome, only fewer than 10 E3s have been used in the PROTACs technology.

PROTACs function in cells or tissues only if the target E3s are expressed there. Although VHL is ubiquitously expressed in many tissues, it is still undetectable in the brain, muscle, gastrointestinal tract, and blood cells. This limits the application of VHL-based PROTACs treating the diseases in these organs. Other potential E3s need to be explored to expand the application of PROTACs.

5.5.2. The Armadillo domain expands the PROTACs toolbox

With the discovery of more E3s, the toolbox of PROTACs is rapidly expanding. PROTACs hijacking MDM E3 for the degradation of androgen receptor[498] and BRD4 protein[499] have been developed. Out of the BTB domain-containing protein family, Kelch-like ECH-associated protein 1 (KEAP1) has been exploited for PROTACs. A proof-of-concept experiment applying a peptide-based PROTAC hijacking KEAP1-CUL3 E3 to degrade Tau protein, has been developed. It consists of a short binder of KEAP1, a peptide from β-tubulin that interacts with Tau, and a cell-penetrating peptide (poly-D-arginine)[500].

The expression of different E3s or E3 subunits has cell- or tissue-specificity. As mentioned, VHL is not expressed everywhere. FBX16 is specifically expressed in the caudate and cerebral cortex; KLHL40 and KLHL41 are only expressed in the testis and skeletal muscles, respectively. MDM2 is wildly expressed in human tissues. ARMC5 is highly expressed in adrenal glands, thyroid, tonsils, lymphoid tissues, brain, spinal cord, and ureter

in mice and humans[349,8]. The tissue specificity of certain E3s would be an advantage for therapeutic applications, as they do not degrade proteins in the tissues where these E3s are not expressed. The strategy of designing PROTACs based on ARMC5-CUL3 E3 could lead to tissue-specific degradation of disease-related proteins, such as Tau protein in Alzheimer's disease[501].

Not only the spatial specificity, but E3s also have temporal expression under certain circumstances. Some E3s are only expressed at specific stages of the cell cycle or can be triggered by infection or diseases. ARMC5 is one of them, as it is upregulated after T cell activation. Such temporal expression could be used to develop PROTACs to degrade certain proteins in a time-specific or disease-specific fashion.

As we now know better about PROTACs, we realize that the protein-protein interaction between the neo-substrate and E3 ligase is essential for a well-designed PROTAC drug. Because of the dimerization or oligomerization of BTB domain-containing proteins, CUL3-based E3s have the advantage of amplifying the degradation signal by recruiting substrates with multiple recognition units.

However, there are also some disadvantages when applying the BTB domain-containing E3s for PROTACs. Many of the BTB domain-containing proteins have been linked to proteolytic and non-proteolytic ubiquitination signals, which may complicate their utility for protein degradation.

ARMC5 contains seven Armadillo repeats and one BTB domain. Each Armadillo repeat forms one α-helix, and the Armadillo domain forms a groove that is suitable for protein-protein interaction. There is no study applying the Armadillo domain as the E3 recognition subunit in PROTACs yet. However, two stapled α-helical peptides (SAHPA1 and SAHPA2) have been designed to fit into the narrow groove formed by the Armadillo repeats in β-catenin[502]. These two peptides disrupted the interaction between β-catenin and Axin protein. Inspired by this discovery, a similar sequence of peptide could be designed for the Armadillo repeats in ARMC5, that acts as a PROTACs subunit recruiting ARMC5-CUL3-RBX1 E3. On the other end of the designed peptide, a peptide linked to

some disease-associated protein in ARMC5-expressed tissues would be promising for drug discovery.

5.6. Future plan

Several logical and important follow-up questions remain unanswered after the completion of my Ph.D. project, and we hope to address them in future studies.

5.6.1. To identify the ubiquitination sites in RPB1 targeted by ARMC5-CUL3 E3

The ubiquitination sites of RPB1 targeted by the novel E3 could help us to know better about the preferred lysine residues of the ARMC5-CUL3 E3 ligase. Besides, Some *RPB1* recurrent mutations in humans are highly related to certain diseases, such as a distinct subset of meningiomas[503] and a neurodevelopmental syndrome with infantile-onset hypotonia[504,505]. Knowing better about the function of each site in RPB1 will be helpful when exploring the pathogenesis of these genetic diseases.

As mentioned in Introduction, there are more than 40 lysine sites in RPB1, which could be potentially ubiquitinated. These lysine sites were identified by the di-GLY antibody enrichment followed by LC-MS/MS analysis. Different conditions, such as Bortezomib (proteasome inhibitor), MLN4942 (CRLs inhibitor), cycloheximide (the translational inhibitor), UV and ionizing radiation were applied in these studies[261,262,506,507,508]. These potential ubiquitination sites are summarized in **Figure 5.2**. Thirty-one lysine sites were detected without artificially induced DNA damage, while some lysine sites were only ubiquitinated upon DNA damage. With so many possible ubiquitination sites in RPB1, we can presume that the ubiquitination pattern of RPB1 would be very complicated.

Based on our Co-IP results, we know ARMC5 binds to the main body of RPB1 but not the CTD (**Chapter 2, Fig. 2**). it is possible that the ARMC5-related ubiquitination sites are also located in this part, which is at the proximity of RPB1.
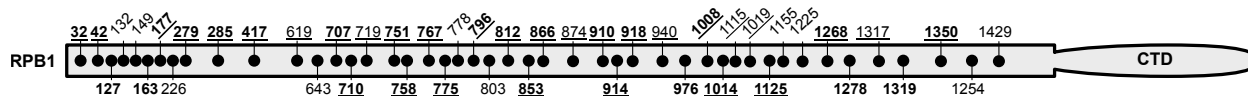
32 42 132 149 177 279 285 417 619 707 719 751 767 778 796 812 866 874 910 918 940 1008 1115 1019 1155 1225 1268 1317 1350 1429

RPB1 ●●●●●●●● ● ● ●●●● ● ●● ●●●●●● ●● ● ●●● ● ● ●●●●● ●● ● ● ● ● ● ● **CTD**

127 163 226 643 710 758 775 803 853 914 976 1014 1125 1278 1319 1254

Fig. 5.2. The potential ubiquitination sites in RPB1. These lysine sites in RPB1 were detected by di-Gly antibody enrichment followed by LC-MS/MS analysis by independent studies with DNA damage[506,507,508] or without DNA damage[261,262]. The bold font indicates that the sites were detected in more than two independent studies. The underlined font indicates that these sites were detected without inducing DNA damage.

To conduct our own investigation of the ubiquitination sites targeted by ARMC5-CUL3-RBX1 E3, we can detect the ubiquitome in WT and KO MEFs or T cells with quantitative MS using the stable isotope labeling with amino acids in cell culture (SILAC) strategy[509]. Since the proteolytic ubiquitination signal is weak under physiological conditions, proteasome inhibitors, such as Bortezomib or MG132, could be added to amplify the signal.

5.6.2. To identify other substrates targeted by ARMC5-CUL3 E3

An E3 normally has more than one substrate. It can act on any protein in its vicinity. The job of the substrate recognition unit is to pull a protein close enough to the E3. Therefore, any protein interacting with ARMC5 could be the potential substrates of ARMC5-CUL3-RBX1 E3. We identified ARMC5 binding proteins by Y2H and LC-MS/MS analysis of ARMC5 precipitates from HEK293 cells. A group of DNA-binding proteins and RNA-binding proteins, such as DHX9, MCM4/6, and RUVBL1/2, are in the list. These proteins are probably associated with Pol II, which is, in turn, interacting with ARMC5. If these proteins are close enough to the catalytic domain of ARMC5-CUL3-RBX1 E3, they could get ubiquitinated. In addition, the other neighbor subunits of RPB1 in Pol II, such as RPB2, RPB5, and RPB9 could be ubiquitinated as well.

Another group of ARMC5-binding partners is heat shock protein (HSP). HSPs act as chaperons to help protein folding and guide misfolded proteins to degradation[510]. HSPs might be required when assembling the multi-unit E3, and they could also be ubiquitinated by E3. So far, there are two known HSP-specific E3s. ChIP is the one directing HSP70 to proteasome degradation[511]. Parkin monoubiquitinates HSP70 at multiple sites, which does

not alter its turnover but regulates its signaling[512]. ARMC5-CUL3-RBX1 might be a novel E3 for HSPs controlling both their turnover and function.

Several other E3s, such as TRIM28 and HUWE1, are found to be associated with ARMC5 as well. HUWE1 is a HECT E3 controlling the degradation of many proteins, such as p53 and Myc[513,514]. TRIM28 is a RING E3 responsible for the sumoylation of PCNA and NPM1/B23[515,516]. It can be recruited to the chromatin to control gene expression and enhance the repair machinery to DNA damage[517]. Different E3s could cooperate together to conduct the ubiquitination mission. For example, ARI-1, an RBR E3, cooperates with CUL1 E3 sequentially monoubiquitinates the substrate and builds polyubiquitin chains on the substrates[518]. Therefore, ARMC5 could cooperate with TRIM28 and HUWE1 to ubiquitinate the substrates in a more complicated way.

Moreover, the degradation of E3 ligases can be achieved by self-ubiquitination or targeted by other E3s. Through E3-E3 interaction, E3s are possibly controlling the turnover of each other, thus timely terminating the ubiquitination process. It will be interesting to study the cooperation process of these E3 pairs.

We will assess these candidate substrates according to their protein levels in the KO tissues and according to *in vitro* ubiquitination assays. To comprehensively identify the substrates of this ARMC5-CUL3-RBX1 complex will allow us to fully understand the biological and pathogenic significance of this E3.

References

[1] Assié G, Libé R, Espiard S, Rizk-Rabin M, Guimier A, Luscap W, et al. *ARMC5* Mutations in Macronodular Adrenal Hyperplasia with Cushing's Syndrome. New England Journal of Medicine. 2013 Nov;369(22):2105-14.

[2] Bourdeau I, Oble S, Magne F, Levesque I, Caceres-Gorriti KY, Nolet S, et al. ARMC5 Mutations in a Large French-Canadian Family with Cortisol-Secreting Beta-Adrenergic/Vasopressin Responsive Bilateral Macronodular Adrenal Hyperplasia. Eur J Endocrinol. 2016 Jan;174(1):85-96.

[3] Gagliardi L, Schreiber AW, Hahn CN, Feng J, Cranston T, Boon H, et al. ARMC5 Mutations Are Common in Familial Bilateral Macronodular Adrenal Hyperplasia. J Clin Endocrinol Metab. 2014 Sep;99(9):E1784-92.

[4] Jojima T, Kogai T, Iijima T, Kato K, Sagara M, Kezuka A, et al. Genetic Alteration of ARMC5 in a Patient Diagnosed with Meningioma and Primary Macronodular Adrenal Hyperplasia: A Case Report. European Journal of Endocrinology. 2020 Dec;183(6):K7-K12.

[5] Ferreira MJ, Pedro J, Salazar D, Costa C, Aragão Rodrigues J, Costa MM, et al. ARMC5 Primary Bilateral Macronodular Adrenal Hyperplasia Associated with a Meningioma: A Family Report. Case Reports in Endocrinology. 2020 Sep;2020:e8848151.

[6] Peifer M, Sweeton D, Casey M, Wieschaus E. Wingless Signal and Zeste-white 3 Kinase Trigger Opposing Changes in the Intracellular Distribution of Armadillo. Development. 1994 Feb;120(2):369-80.

[7] Krek W. BTB Proteins as Henchmen of Cul3-based Ubiquitin Ligases. Nature Cell Biology. 2003 Nov;5(11):950-1.

[8] Hu Y, Lao L, Mao J, Jin W, Luo H, Charpentier T, et al. Armc5 Deletion Causes Developmental Defects and Compromises T-cell Immune Responses. Nature Communications. 2017 Feb;8(1).

[9] Ensembl. ARMC5 (ENSG00000140691) - Homo_sapiens;. `http://useast.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000140691`. Website.

[10] Berthon A, Faucz FR, Espiard S, Drougat L, Bertherat J, Stratakis CA. Age-Dependent Effects of Armc5 Haploinsufficiency on Adrenocortical Function. Human molecular genetics. 2017;26(18):3495-507.

[11] Lacroix A, Feelders RA, Stratakis CA, Nieman LK. Cushing's Syndrome. The Lancet. 2015 Aug;386(9996):913-27.

[12] Strohm M, Reincke M, Theiß M, Diehl KL, Allolio B. Bilaterale massive makronoduläre Nebennieren-hyperplasie: Eine seltene Ursache des adrenalen Cushing-Syndroms. DMW - Deutsche Medizinische Wochenschrift. 1994;119(6):180-4.

[13] Lieberman SA, Eccleshall TR, Feldman D. ACTH-independent Massive Bilateral Adrenal Disease (AIMBAD): A Subtype of Cushing's Syndrome with Major Diagnostic and Therapeutic Implications. European Journal of Endocrinology. 1994 Jul;131(1):67-73.

[14] Faucz FR, Zilbermint M, Lodish MB, Szarek E, Trivellin G, Sinaii N, et al. Macronodular Adrenal Hyperplasia Due to Mutations in an Armadillo Repeat Containing 5 (ARMC5) Gene: A Clinical and Genetic Investigation. J Clin Endocrinol Metab. 2014 Jun;99(6):E1113-9.

[15] Newell-Price J. Cushing's Syndrome. Clinical Medicine. 2008 Apr;8(2):204-8.

[16] Espiard S, Drougat L, Libe R, Assie G, Perlemoine K, Guignat L, et al. ARMC5 Mutations in a Large Cohort of Primary Macronodular Adrenal Hyperplasia: Clinical and Functional Consequences. J Clin Endocrinol Metab. 2015 Jun;100(6):E926-35.

[17] Albiger NM, Regazzo D, Rubin B, Ferrara AM, Rizzati S, Taschin E, et al. A Multicenter Experience on the Prevalence of ARMC5 Mutations in Patients with Primary Bilateral Macronodular Adrenal Hyperplasia: From Genetic Characterization to Clinical Phenotype. Endocrine. 2016;55(3):959-68.

[18] Alencar GA, Lerario AM, Nishi MY, Mariani BM, Almeida MQ, Tremblay J, et al. ARMC5 Mutations Are a Frequent Cause of Primary Macronodular Adrenal Hyperplasia. J Clin Endocrinol Metab. 2014 Aug;99(8):E1501-9.

[19] Elbelt U, Trovato A, Kloth M, Gentz E, Finke R, Spranger J, et al. Molecular and Clinical Evidence for an ARMC5 Tumor Syndrome: Concurrent Inactivating Germline and Somatic Mutations Are Associated with Both Primary Macronodular Adrenal Hyperplasia and Meningioma. J Clin Endocrinol Metab. 2015 Jan;100(1):E119-28.

[20] Qian Z, Liang C, Jiang-Ping G, Wen-Hua Y, Nan J, Kang C, et al. Whole-Genome Sequencing Revealed Armadillo Repeat Containing 5 (ARMC5) Mutation in a Chinese Family with ACTH-independent Macronodular Adrenal Hyperplasia. Endocrine Journal. 2017:317.

[21] Correa R, Zilbermint M, Berthon A, Espiard S, Batsis M, Papadakis GZ, et al. The ARMC5 Gene Shows Extensive Genetic Variance in Primary Macronodular Adrenocortical Hyperplasia. Eur J Endocrinol. 2015 Oct;173(4):435-40.

[22] Suzuki S, Tatsuno I, Oohara E, Nakayama A, Komai E, Shiga A, et al. Germline Deletion of Armc5 in Familial Primary Macronodular Adrenal Hyperplasia. Endocr Pract. 2015 Oct;21(10):1152-60.

[23] Holly E, Ioanna T, Treena C, Stylianos T, Ashley BG. Do Patients with Incidentally Discovered Bilateral Adrenal Nodules Represent an Early Form of ARMC5-mediated Bilateral Macronodular Hyperplasia? Endocrine. 2016;53(3):801-8.

[24] Mulatero P, Schiavi F, Williams TA, Monticone S, Barbon G, Opocher G, et al. ARMC5 Mutation Analysis in Patients with Primary Aldosteronism and Bilateral Adrenal Lesions. J Hum Hypertens. 2016 Jun;30(6):374-8.

[25] Damjanovic SS, Antic JA, Elezovic-Kovacevic VI, Dundjerovic DM, Milicevic IT, Beleslin-Cokic BB, et al. ARMC5 Alterations in Patients With Sporadic Neuroendocrine Tumors and Multiple Endocrine Neoplasia Type 1 (MEN1). The Journal of Clinical Endocrinology & Metabolism. 2020 Dec;105(12):e4531-42.

[26] Zhang F, Lin X, Yu X. Primary Macronodular Adrenal Hyperplasia (PMAH) Can Be Generated by a New *ARMC5* Germline Variant (c.52C>T (p.Gln18X)). Endocrine Journal. 2020;67(12):1179-86.

[27] He WT, Wang X, Song W, Song XD, Lu YJ, Lv YK, et al. A Novel Nonsense Mutation in ARMC5 Causes Primary Bilateral Macronodular Adrenocortical Hyperplasia. BMC medical genomics. 2021 May;14(1):126.

[28] Lacroix A. ACTH-independent Macronodular Adrenal Hyperplasia. Best Practice & Research Clinical Endocrinology & Metabolism. 2009 Apr;23(2):245-59.

[29] Hofland J, Hofland LJ, van Koetsveld PM, Steenbergen J, de Herder WW, van Eijck CH, et al. ACTH-independent Macronodular Adrenocortical Hyperplasia Reveals Prevalent Aberrant in Vivo and in Vitro Responses to Hormonal Stimuli and Coupling of Arginine-Vasopressin Type 1a Receptor to 11β-Hydroxylase. Orphanet Journal of Rare Diseases. 2013 Sep;8:142.

[30] Candida Barisson Villares Fragoso M, Pontes Cavalcante I, Meneses Ferreira A, Marinho de Paula Mariani B, Ferini Pacicco Lotfi C. Genetics of Primary Macronodular Adrenal Hyperplasia. La Presse Médicale. 2018 Jul;47(7, Part 2):e139-49.

[31] Weinstein LS, Shenker A, Gejman PV, Merino MJ, Friedman E, Spiegel AM. Activating Mutations of the Stimulatory G Protein in the McCune-Albright Syndrome. The New England Journal of Medicine. 1991 Dec;325(24):1688-95.

[32] Boyce AM, Florenzano P, de Castro LF, Collins MT. Fibrous Dysplasia/McCune-Albright Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJ, Mirzaa G, et al., editors. GeneReviews®. Seattle (WA): University of Washington, Seattle; 1993. p. 1.

[33] Carney JA, Young WF, Stratakis CA. Primary Bimorphic Adrenocortical Disease: Cause of Hypercortisolism in McCune-Albright Syndrome. The American Journal of Surgical Pathology. 2011 Sep;35(9):1311-26.

[34] Kirschner LS, Carney JA, Pack SD, Taymans SE, Giatzakis C, Cho YS, et al. Mutations of the Gene Encoding the Protein Kinase A Type I-alpha Regulatory Subunit in Patients with the Carney Complex. Nature Genetics. 2000 Sep;26(1):89-92.

[35] Bertherat J, Groussin L, Sandrini F, Matyakhina L, Bei T, Stergiopoulos S, et al. Molecular and Functional Analysis of PRKAR1A and Its Locus (17q22-24) in Sporadic Adrenocortical Tumors: 17q Losses, Somatic Mutations, and Protein Kinase A Expression and Activity. Cancer Research. 2003 Sep;63(17):5308-19.

[36] Forlino A, Vetro A, Garavelli L, Ciccone R, London E, Stratakis CA, et al. PRKACB and Carney Complex. The New England Journal of Medicine. 2014 Mar;370(11):1065-7.

[37] Libé R, Fratticci A, Coste J, Tissier F, Horvath A, Ragazzon B, et al. Phosphodiesterase 11A (PDE11A) and Genetic Predisposition to Adrenocortical Tumors. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008 Jun;14(12):4016-24.

[38] Vezzosi D, Libé R, Baudry C, Rizk-Rabin M, Horvath A, Levy I, et al. Phosphodiesterase 11A (PDE11A) Gene Defects in Patients with Acth-Independent Macronodular Adrenal Hyperplasia (AIMAH): Functional Variants May Contribute to Genetic Susceptibility of Bilateral Adrenal Tumors. The Journal of Clinical Endocrinology and Metabolism. 2012 Nov;97(11):E2063-9.

[39] Rothenbuhler A, Horvath A, Libé R, Faucz FR, Fratticci A, Sanson MLR, et al. Identification of Novel Genetic Variants in Phosphodiesterase 8B (PDE8B), a cAMP-specific Phosphodiesterase Highly Expressed in the Adrenal Cortex, in a Cohort of Patients with Adrenal Tumours. Clinical Endocrinology. 2012;77(2):195-9.

[40] Gatta-Cherifi B, Chabre O, Murat A, Niccoli P, Cardot-Bauters C, Rohmer V, et al. Adrenal Involvement in MEN1. Analysis of 715 Cases from the Groupe d'etude Des Tumeurs Endocrines Database. European Journal of Endocrinology. 2012 Feb;166(2):269-79.

[41] Burgess JR, Harle RA, Tucker P, Parameswaran V, Davies P, Greenaway TM, et al. Adrenal Lesions in a Large Kindred with Multiple Endocrine Neoplasia Type 1. Archives of Surgery (Chicago, Ill: 1960). 1996 Jul;131(7):699-702.

[42] Cetta F, Curia MC, Montalto G, Gori M, Cama A, Battista P, et al. Thyroid Carcinoma Usually Occurs in Patients with Familial Adenomatous Polyposis in the Absence of Biallelic Inactivation of the Adenomatous Polyposis Coli Gene. The Journal of Clinical Endocrinology and Metabolism. 2001 Jan;86(1):427-32.

[43] Marchesa P, Fazio VW, Church JM, McGannon E. Adrenal Masses in Patients with Familial Adenomatous Polyposis. Diseases of the Colon and Rectum. 1997 Sep;40(9):1023-8.

[44] Shuch Brian, Ricketts Christopher J , Vocke Cathy D , Valera Vladimir A , Chen Clara C , Gautam Rabi, et al. Adrenal Nodular Hyperplasia in Hereditary Leiomyomatosis and Renal Cell Cancer. Journal of Urology. 2013 Feb;189(2):430-5.

[45] Mitchell LE. Epidemiology of Neural Tube Defects. American Journal of Medical Genetics Part C: Seminars in Medical Genetics. 2005;135C(1):88-94.

[46] Copp AJ, Greene ND. Genetics and Development of Neural Tube Defects. The Journal of Pathology. 2010;220(2):217-30.

[47] Smithells RW, Sheppard S, Schorah CJ. Vitamin Deficiencies and Neural Tube Defects. Archives of Disease in Childhood. 1976 Dec;51(12):944-50.

[48] Correa A, Botto L, Liu Y, Mulinare J, Erickson JD. Do Multivitamin Supplements Attenuate the Risk for Diabetes-Associated Birth Defects? Pediatrics. 2003 May;111(Supplement 1):1146-51.

[49] Wlodarczyk BJ, Palacios AM, George TM, Finnell RH. Antiepileptic Drugs and Pregnancy Outcomes. American Journal of Medical Genetics Part A. 2012 Aug;158A(8):2071-90.

[50] Missmer SA, Suarez L, Felkner M, Wang E, Merrill AH, Rothman KJ, et al. Exposure to Fumonisins and the Occurrence of Neural Tube Defects along the Texas-Mexico Border. Environmental Health Perspectives. 2006 Feb;114(2):237-41.

[51] GROUP MVSR. Prevention of Neural Tube Defects: Results of the Medical Research Council Vitamin Study. The Lancet. 1991 Jul;338(8760):131-7.

[52] Berry RJ, Li Z, Erickson JD, Li S, Moore CA, Wang H, et al. Prevention of Neural-Tube Defects with Folic Acid in China. China-U.S. Collaborative Project for Neural Tube Defect Prevention. The New England Journal of Medicine. 1999 Nov;341(20):1485-90.

[53] Burren KA, Savery D, Massa V, Kok RM, Scott JM, Blom HJ, et al. Gene-Environment Interactions in the Causation of Neural Tube Defects: Folate Deficiency Increases Susceptibility Conferred by Loss of Pax3 Function. Human Molecular Genetics. 2008 Dec;17(23):3675-85.

[54] Yang P, Li X, Xu C, Eckert RL, Reece EA, Zielke HR, et al. Maternal Hyperglycemia Activates an ASK1-FoxO3a-caspase 8 Pathway That Leads to Embryonic Neural Tube Defects. Science Signaling. 2013 Aug;6(290):ra74.

[55] Reece EA. Diabetes-Induced Birth Defects: What Do We Know? What Can We Do? Current Diabetes Reports. 2012 Feb;12(1):24-32.

[56] Harris MJ, Juriloff DM. Mouse Mutants with Neural Tube Closure Defects and Their Role in Understanding Human Neural Tube Defects. Birth Defects Research Part A, Clinical and Molecular Teratology. 2007 Mar;79(3):187-210.

[57] Christensen KE, Deng L, Leung KY, Arning E, Bottiglieri T, Malysheva OV, et al. A Novel Mouse Model for Genetic Variation in 10-Formyltetrahydrofolate Synthetase Exhibits Disturbed Purine Synthesis with Impacts on Pregnancy and Embryonic Development. Human Molecular Genetics. 2013 Sep;22(18):3705-19.

[58] Chowdary D, Streck D, Schwalb MN, Dermody JJ. High Incidence of Two Methylenetetrahydrofolate Reductase Mutations (C677T and A1298C) in Hispanics. Genetic Testing. 2003;7(3):255-7.

[59] Beaudin AE, Abarinov EV, Noden DM, Perry CA, Chu S, Stabler SP, et al. Shmt1 and de Novo Thymidylate Biosynthesis Underlie Folate-Responsive Neural Tube Defects in Mice. The American Journal of Clinical Nutrition. 2011 Apr;93(4):789-98.

[60] Dunlevy LPE, Chitty LS, Burren KA, Doudney K, Stojilkovic-Mikic T, Stanier P, et al. Abnormal Folate Metabolism in Foetuses Affected by Neural Tube Defects. Brain: A Journal of Neurology. 2007 Apr;130(Pt 4):1043-9.

[61] Czeizel AE, Bártfai Z, Bánhidy F. Primary Prevention of Neural-Tube Defects and Some Other Congenital Abnormalities by Folic Acid and Multivitamins: History, Missed Opportunity and Tasks. Therapeutic Advances in Drug Safety. 2011 Aug;2(4):173-88.

[62] Blencowe H, Cousens S, Modell B, Lawn J. Folic Acid to Reduce Neonatal Mortality from Neural Tube Disorders. International Journal of Epidemiology. 2010 Apr;39 Suppl 1:i110-21.

[63] Molloy AM, Kirke PN, Troendle JF, Burke H, Sutton M, Brody LC, et al. Maternal Vitamin B12 Status and Risk of Neural Tube Defects in a Population with High Neural Tube Defect Prevalence and No Folic Acid Fortification. Pediatrics. 2009 Mar;123(3):917-23.

[64] Sotres-Alvarez D, Siega-Riz AM, Herring AH, Carmichael SL, Feldkamp ML, Hobbs CA, et al. Maternal Dietary Patterns Are Associated with Risk of Neural Tube and Congenital Heart Defects. American Journal of Epidemiology. 2013 Jun;177(11):1279-88.

[65] Czeizel AE, Dudás I, Vereczkey A, Bánhidy F. Folate Deficiency and Folic Acid Supplementation: The Prevention of Neural-Tube Defects and Congenital Heart Defects. Nutrients. 2013 Nov;5(11):4760-75.

[66] Leung KY, De Castro SCP, Savery D, Copp AJ, Greene NDE. Nucleotide Precursors Prevent Folic Acid-Resistant Neural Tube Defects in the Mouse. Brain. 2013 Sep;136(9):2836-41.

[67] Momb J, Lewandowski JP, Bryant JD, Fitch R, Surman DR, Vokes SA, et al. Deletion of Mthfd1l Causes Embryonic Lethality and Neural Tube and Craniofacial Defects in Mice. Proceedings of the National Academy of Sciences of the United States of America. 2013 Jan;110(2):549-54.

[68] Özkaynak E, Finley D, Varshavsky A. The Yeast Ubiquitin Gene: Head-to-Tail Repeats Encoding a Polyubiquitin Precursor Protein. Nature. 1984 Dec;312(5995):663-6.

[69] Ozkaynak E, Finley D, Solomon MJ, Varshavsky A. The Yeast Ubiquitin Genes: A Family of Natural Gene Fusions. The EMBO Journal. 1987 May;6(5):1429-39.

[70] Redman KL, Rechsteiner M. Identification of the Long Ubiquitin Extension as Ribosomal Protein S27a. Nature. 1989 Mar;338(6214):438-40.

[71] Schulman BA, Harper JW. Ubiquitin-like Protein Activation by E1 Enzymes: The Apex for Downstream Signalling Pathways. Nature Reviews Molecular Cell Biology. 2009 May;10(5):319-31.

[72] McGrath JP, Jentsch S, Varshavsky A. UBA 1: An Essential Yeast Gene Encoding Ubiquitin-Activating Enzyme. The EMBO Journal. 1991 Jan;10(1):227-36.

[73] Jin J, Li X, Gygi SP, Harper JW. Dual E1 Activation Systems for Ubiquitin Differentially Regulate E2 Enzyme Charging. Nature. 2007 Jun;447(7148):1135-8.

[74] Chiu YH, Sun Q, Chen ZJ. E1-L2 Activates Both Ubiquitin and FAT10. Molecular Cell. 2007 Sep;27(6):1014-23.

[75] Pelzer C, Kassner I, Matentzoglu K, Singh RK, Wollscheid HP, Scheffner M, et al. UBE1L2, a Novel E1 Enzyme Specific for Ubiquitin. Journal of Biological Chemistry. 2007 Aug;282(32):23010-4.

[76] Lin Y, Hwang WC, Basavappa R. Structural and Functional Analysis of the Human Mitotic-Specific Ubiquitin-Conjugating Enzyme, UbcH10. The Journal of Biological Chemistry. 2002 Jun;277(24):21913-21.

[77] Brzovic PS, Klevit RE. Ubiquitin Transfer from the E2 Perspective: Why Is UbcH5 So Promiscuous? Cell Cycle. 2006 Dec;5(24):2867-73.

[78] Jin L, Williamson A, Banerjee S, Philipp I, Rape M. Mechanism of Ubiquitin-Chain Formation by the Human Anaphase-Promoting Complex. Cell. 2008 May;133(4):653-65.

[79] Windheim M, Peggie M, Cohen P. Two Different Classes of E2 Ubiquitin-Conjugating Enzymes Are Required for the Mono-Ubiquitination of Proteins and Elongation by Polyubiquitin Chains with a Specific Topology. The Biochemical Journal. 2008 Feb;409(3):723-9.

[80] Rotin D, Kumar S. Physiological Functions of the HECT Family of Ubiquitin Ligases. Nature Reviews Molecular Cell Biology. 2009 Jun;10(6):398-409.

[81] Li W, Bengtson MH, Ulbrich A, Matsuda A, Reddy VA, Orth A, et al. Genome-Wide and Functional Annotation of Human E3 Ubiquitin Ligases Identifies MULAN, a Mitochondrial E3 That Regulates

the Organelle's Dynamics and Signaling. PLoS ONE. 2008 Jan;3(1):e1487.

[82] Enchev RI, Schulman BA, Peter M. Protein Neddylation: Beyond Cullin–RING Ligases. Nature Reviews Molecular Cell Biology. 2015 Jan;16(1):30-44.

[83] Goldenberg SJ, Cascio TC, Shumway SD, Garbutt KC, Liu J, Xiong Y, et al. Structure of the Cand1-Cul1-Roc1 Complex Reveals Regulatory Mechanisms for the Assembly of the Multisubunit Cullin-Dependent Ubiquitin Ligases. Cell. 2004 Nov;119(4):517-28.

[84] Zheng J, Yang X, Harrell JM, Ryzhikov S, Shim EH, Lykke-Andersen K, et al. CAND1 Binds to Unneddylated CUL1 and Regulates the Formation of SCF Ubiquitin E3 Ligase Complex. Molecular Cell. 2002 Dec;10(6):1519-26.

[85] Saha A, Deshaies RJ. Multimodal Activation of the Ubiquitin Ligase SCF by Nedd8 Conjugation. Molecular Cell. 2008 Oct;32(1):21-31.

[86] Schulman BA, Carrano AC, Jeffrey PD, Bowen Z, Kinnucan ER, Finnin MS, et al. Insights into SCF Ubiquitin Ligases from the Structure of the Skp1-Skp2 Complex. Nature. 2000 Nov;408(6810):381-6.

[87] Zheng N, Schulman BA, Song L, Miller JJ, Jeffrey PD, Wang P, et al. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF Ubiquitin Ligase Complex. Nature. 2002 Apr;416(6882):703-9.

[88] Jin J, Cardozo T, Lovering RC, Elledge SJ, Pagano M, Harper JW. Systematic Analysis and Nomenclature of Mammalian F-box Proteins. Genes & Development. 2004 Nov;18(21):2573-80.

[89] Shi L, Du D, Peng Y, Liu J, Long J. The Functional Analysis of Cullin 7 E3 Ubiquitin Ligases in Cancer. Oncogenesis. 2020 Oct;9(10):1-12.

[90] Kershaw NJ, Babon JJ. VHL: Cullin-g the Hypoxic Response. Structure. 2015 Mar;23(3):435-6.

[91] Jin J, Arias EE, Chen J, Harper JW, Walter JC. A Family of Diverse Cul4-Ddb1-Interacting Proteins Includes Cdt2, Which Is Required for S Phase Destruction of the Replication Factor Cdt1. Molecular Cell. 2006 Sep;23(5):709-21.

[92] Scrima A, Koníčková R, Czyzewski BK, Kawasaki Y, Jeffrey PD, Groisman R, et al. Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2 Complex. Cell. 2008 Dec;135(7):1213-23.

[93] Higa LA, Zhang H. Stealing the Spotlight: CUL4-DDB1 Ubiquitin Ligase Docks WD40-repeat Proteins to Destroy. Cell Division. 2007 Feb;2(1):5.

[94] Singer JD, Gurian-West M, Clurman B, Roberts JM. Cullin-3 Targets Cyclin E for Ubiquitination and Controls S Phase in Mammalian Cells. Genes & Development. 1999 Sep;13(18):2375-87.

[95] Angers S, Thorpe CJ, Biechele TL, Goldenberg SJ, Zheng N, MacCoss MJ, et al. The KLHL12–Cullin-3 Ubiquitin Ligase Negatively Regulates the Wnt–β-Catenin Pathway by Targeting Dishevelled for Degradation. Nature Cell Biology. 2006 Apr;8(4):348-57.

[96] Furukawa M, Xiong Y. BTB Protein Keap1 Targets Antioxidant Transcription Factor Nrf2 for Ubiquitination by the Cullin 3-Roc1 Ligase. Molecular and Cellular Biology. 2005 Jan;25(1):162-71.

[97] Boyden LM, Choi M, Choate KA, Nelson-Williams CJ, Farhi A, Toka HR, et al. Mutations in Kelch-like 3 and Cullin 3 Cause Hypertension and Electrolyte Abnormalities. Nature. 2012 Feb;482(7383):98-102.

[98] Steklov M, Pandolfi S, Baietti MF, Batiuk A, Carai P, Najm P, et al. Mutations in LZTR1 Drive Human Disease by Dysregulating RAS Ubiquitination. Science. 2018 Dec;362(6419):1177-82.

[99] Lee YR, Yuan WC, Ho HC, Chen CH, Shih HM, Chen RH. The Cullin 3 Substrate Adaptor KLHL20 Mediates DAPK Ubiquitination to Control Interferon Responses. The EMBO Journal. 2010 May;29(10):1748-61.

[100] Taguchi K, Motohashi H, Yamamoto M. Molecular Mechanisms of the Keap1–Nrf2 Pathway in Stress Response and Cancer Evolution. Genes to Cells. 2011 Feb;16(2):123-40.

[101] Amar M, Pramod AB, Yu NK, Herrera VM, Qiu LR, Moran-Losada P, et al. Autism-Linked Cullin3 Germline Haploinsufficiency Impacts Cytoskeletal Dynamics and Cortical Neurogenesis through RhoA Signaling. Molecular Psychiatry. 2021 Mar:1-28.

[102] Furukawa M, He YJ, Borchers C, Xiong Y. Targeting of Protein Ubiquitination by BTB–Cullin 3–Roc1 Ubiquitin Ligases. Nature Cell Biology. 2003 Nov;5(11):1001-7.

[103] Gingerich DJ, Gagne JM, Salter DW, Hellmann H, Estelle M, Ma L, et al. Cullins 3a and 3b Assemble with Members of the Broad Complex/Tramtrack/Bric-a-Brac (BTB) Protein Family to Form Essential Ubiquitin-Protein Ligases (E3s) in Arabidopsis*,. Journal of Biological Chemistry. 2005 May;280(19):18810-21.

[104] Werner A, Iwasaki S, McGourty CA, Medina-Ruiz S, Teerikorpi N, Fedrigo I, et al. Cell-Fate Determination by Ubiquitin-Dependent Regulation of Translation. Nature. 2015 Sep;525(7570):523-7.

[105] Arama E, Bader M, Rieckhof GE, Steller H. A Ubiquitin Ligase Complex Regulates Caspase Activation During Sperm Differentiation in Drosophila. PLOS Biology. 2007 Sep;5(10):e251.

[106] Escamilla CO, Filonova I, Walker AK, Xuan ZX, Holehonnur R, Espinosa F, et al. Kctd13 Deletion Reduces Synaptic Transmission via Increased RhoA. Nature. 2017 Nov;551(7679):227-31.

[107] Clague MJ, Urbé S, Komander D. Breaking the Chains: Deubiquitylating Enzyme Specificity Begets Function. Nature Reviews Molecular Cell Biology. 2019 Jun;20(6):338.

[108] Clague MJ, Barsukov I, Coulson JM, Liu H, Rigden DJ, Urbé S. Deubiquitylases From Genes to Organism. Physiological Reviews. 2013 Jul;93(3):1289-315.

[109] Kaiser SE, Riley BE, Shaler TA, Trevino RS, Becker CH, Schulman H, et al. Protein Standard Absolute Quantification (PSAQ) Method for the Measurement of Cellular Ubiquitin Pools. Nature Methods. 2011 Aug;8(8):691-6.

[110] Voges D, Zwickl P, Baumeister W. The 26S Proteasome: A Molecular Machine Designed for Controlled Proteolysis. Annual Review of Biochemistry. 1999 Jun;68(1):1015-68.

[111] Saeki Y. Ubiquitin Recognition by the Proteasome. The Journal of Biochemistry. 2017 Feb;161(2):113-24.

[112] Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, et al. Role of Histone H2A Ubiquitination in Polycomb Silencing. Nature. 2004 Oct;431(7010):873-8.

[113] Wood A, Krogan NJ, Dover J, Schneider J, Heidt J, Boateng MA, et al. Bre1, an E3 Ubiquitin Ligase Required for Recruitment and Substrate Selection of Rad6 at a Promoter. Molecular Cell. 2003 Jan;11(1):267-74.

[114] Zheng S, Li D, Lu Z, Liu G, Wang M, Xing P, et al. Bre1-Dependent H2B Ubiquitination Promotes Homologous Recombination by Stimulating Histone Eviction at DNA Breaks. Nucleic Acids Research. 2018 Nov;46(21):11326-39.

[115] Ulrich HD, Walden H. Ubiquitin Signalling in DNA Replication and Repair. Nature Reviews Molecular Cell Biology. 2010 Jul;11(7):479-89.

[116] Zeng W, Sun L, Jiang X, Chen X, Hou F, Adhikari A, et al. Reconstitution of the RIG-I Pathway Reveals a Signaling Role of Unanchored Polyubiquitin Chains in Innate Immunity. Cell. 2010 Apr;141(2):315-30.

[117] Kirkpatrick DS, Denison C, Gygi SP. Weighing in on Ubiquitin: The Expanding Role of Mass-Spectrometry-Based Proteomics. Nature Cell Biology. 2005 Aug;7(8):750-7.

[118] Kondapalli C, Kazlauskaite A, Zhang N, Woodroof HI, Campbell DG, Gourlay R, et al. PINK1 Is Activated by Mitochondrial Membrane Potential Depolarization and Stimulates Parkin E3 Ligase Activity by Phosphorylating Serine 65. Open Biology. 2012;2(5):120080.

[119] Ohtake F, Saeki Y, Sakamoto K, Ohtake K, Nishikawa H, Tsuchiya H, et al. Ubiquitin Acetylation Inhibits Polyubiquitin Chain Elongation. EMBO reports. 2015 Feb;16(2):192-201.

[120] Tsuchiya H, Burana D, Ohtake F, Arai N, Kaiho A, Komada M, et al. Ub-ProT Reveals Global Length and Composition of Protein Ubiquitylation in Cells. Nature Communications. 2018 Feb;9(1):524.

[121] Richard Y, Michael R. The Increasing Complexity of the Ubiquitin Code. Nature Cell Biology. 2016;18(6).

[122] Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, et al. A Multiubiquitin Chain Is Confined to Specific Lysine in a Targeted Short-Lived Protein. Science. 1989 Mar;243(4898):1576-83.

[123] Scheffner M, Huibregtse JM, Vierstra RD, Howley PM. The HPV-16 E6 and E6-AP Complex Functions as a Ubiquitin-Protein Ligase in the Ubiquitination of P53. Cell. 1993 Nov;75(3):495-505.

[124] Bennett EJ, Rush J, Gygi SP, Harper JW. Dynamics of Cullin-RING Ubiquitin Ligase Network Revealed by Systematic Quantitative Proteomics. Cell. 2010 Dec;143(6):951-65.

[125] Brown NG, VanderLinden R, Watson ER, Weissmann F, Ordureau A, Wu KP, et al. Dual RING E3 Architectures Regulate Multiubiquitination and Ubiquitin Chain Elongation by APC/C. Cell. 2016 Jun;165(6):1440-53.

[126] Xu Y, Anderson DE, Ye Y. The HECT Domain Ubiquitin Ligase HUWE1 Targets Unassembled Soluble Proteins for Degradation. Cell Discovery. 2016 Dec;2(1):16040.

[127] Thrower JS. Recognition of the Polyubiquitin Proteolytic Signal. The EMBO Journal. 2000 Jan;19(1):94-102.

[128] Lu Y, Wang W, Kirschner MW. Specificity of the Anaphase-Promoting Complex: A Single-Molecule Study. Science. 2015 Apr;348(6231):1248737-7.

[129] Peth A, Uchiki T, Goldberg AL. ATP-Dependent Steps in the Binding of Ubiquitin Conjugates to the 26S Proteasome That Commit to Degradation. Molecular Cell. 2010 Nov;40(4):671-81.

[130] Kim HT, Kim KP, Lledias F, Kisselev AF, Scaglione KM, Skowyra D, et al. Certain Pairs of Ubiquitin-conjugating Enzymes (E2s) and Ubiquitin-Protein Ligases (E3s) Synthesize Nondegradable Forked Ubiquitin Chains Containing All Possible Isopeptide Linkages. Journal of Biological Chemistry. 2007 Jun;282(24):17375-86.

[131] Saeki Y, Kudo T, Sone T, Kikuchi Y, Yokosawa H, Toh-e A, et al. Lysine 63-Linked Polyubiquitin Chain May Serve as a Targeting Signal for the 26S Proteasome. The EMBO Journal. 2009 Feb;28(4):359-71.

[132] Nathan JA, Tae Kim H, Ting L, Gygi SP, Goldberg AL. Why Do Cellular Proteins Linked to K63-polyubiquitin Chains Not Associate with Proteasomes? The EMBO Journal. 2013 Feb;32(4):552-65.

[133] Raasi S, Orlov I, Fleming KG, Pickart CM. Binding of Polyubiquitin Chains to Ubiquitin-associated (UBA) Domains of HHR23A. Journal of Molecular Biology. 2004 Aug;341(5):1367-79.

[134] Xu M, Skaug B, Zeng W, Chen ZJ. A Ubiquitin Replacement Strategy in Human Cells Reveals Distinct Mechanisms of IKK Activation by TNFα and IL-1β. Molecular Cell. 2009 Oct;36(2):302-14.

[135] Wang C, Deng L, Hong M, Akkaraju GR, Inoue Ji, Chen ZJ. TAK1 Is a Ubiquitin-Dependent Kinase of MKK and IKK. Nature. 2001 Jul;412(6844):346-51.

[136] Wang G, Long J, Gao Y, Zhang W, Han F, Xu C, et al. SETDB1-mediated Methylation of Akt Promotes Its K63-linked Ubiquitination and Activation Leading to Tumorigenesis. Nature Cell Biology. 2019 Feb;21(2):214-25.

[137] Gack MU, Shin YC, Joo CH, Urano T, Liang C, Sun L, et al. TRIM25 RING-finger E3 Ubiquitin Ligase Is Essential for RIG-I-mediated Antiviral Activity. Nature. 2007 Apr;446(7138):916-20.

[138] Cunningham CN, Baughman JM, Phu L, Tea JS, Yu C, Coons M, et al. USP30 and Parkin Homeostatically Regulate Atypical Ubiquitin Chains on Mitochondria. Nature Cell Biology. 2015 Feb;17(2):160-9.

[139] Doil C, Mailand N, Bekker-Jensen S, Menard P, Larsen DH, Pepperkok R, et al. RNF168 Binds and Amplifies Ubiquitin Conjugates on Damaged Chromosomes to Allow Accumulation of Repair Proteins. Cell. 2009 Feb;136(3):435-46.

[140] Silva GM, Finley D, Vogel C. K63 Polyubiquitination Is a New Modulator of the Oxidative Stress Response. Nature Structural & Molecular Biology. 2015 Feb;22(2):116-23.

[141] Deng L, Wang C, Spencer E, Yang L, Braun A, You J, et al. Activation of the IκB Kinase Complex by TRAF6 Requires a Dimeric Ubiquitin-Conjugating Enzyme Complex and a Unique Polyubiquitin Chain. Cell. 2000 Oct;103(2):351-61.

[142] Wertz IE, O'Rourke KM, Zhou H, Eby M, Aravind L, Seshagiri S, et al. De-Ubiquitination and Ubiquitin Ligase Domains of A20 Downregulate NF-κB Signalling. Nature. 2004 Aug;430(7000):694-9.

[143] Liu Y, Fallon L, Lashuel HA, Liu Z, Lansbury PT. The UCH-L1 Gene Encodes Two Opposing Enzymatic Activities That Affect α-Synuclein Degradation and Parkinson's Disease Susceptibility. Cell. 2002 Oct;111(2):209-18.

[144] Tao M, Scacheri PC, Marinis JM, Harhaj EW, Matesic LE, Abbott DW. ITCH K63-Ubiquitinates the NOD2 Binding Protein, RIP2, to Influence Inflammatory Signaling Pathways. Current Biology. 2009 Aug;19(15):1255-63.

[145] Vina-Vilaseca A, Sorkin A. Lysine 63-Linked Polyubiquitination of the Dopamine Transporter Requires WW3 and WW4 Domains of Nedd4-2 and UBE2D Ubiquitin-conjugating Enzymes*. Journal of Biological Chemistry. 2010 Mar;285(10):7645-56.

[146] Xia ZP, Sun L, Chen X, Pineda G, Jiang X, Adhikari A, et al. Direct Activation of Protein Kinases by Unanchored Polyubiquitin Chains. Nature. 2009 Sep;461(7260):114-9.

[147] Ikeda F, Deribe YL, Skånland SS, Stieglitz B, Grabbe C, Franz-Wachtel M, et al. SHARPIN Forms a Linear Ubiquitin Ligase Complex Regulating NF-κB Activity and Apoptosis. Nature. 2011 Mar;471(7340):637-41.

[148] Kirisako T, Kamei K, Murata S, Kato M, Fukumoto H, Kanie M, et al. A Ubiquitin Ligase Complex Assembles Linear Polyubiquitin Chains. The EMBO Journal. 2006 Oct;25(20):4877-87.

[149] Keusekotten K, Elliott PR, Glockner L, Fiil BK, Damgaard RB, Kulathu Y, et al. OTULIN Antagonizes LUBAC Signaling by Specifically Hydrolyzing Met1-Linked Polyubiquitin. Cell. 2013 Jun;153(6):1312-26.

[150] Inn KS, Gack MU, Tokunaga F, Shi M, Wong LY, Iwai K, et al. Linear Ubiquitin Assembly Complex Negatively Regulates RIG-I- and TRIM25-Mediated Type I Interferon Induction. Molecular Cell. 2011

Feb;41(3):354-65.

[151] Rivkin E, Almeida SM, Ceccarelli DF, Juang YC, MacLean TA, Srikumar T, et al. The Linear Ubiquitin-Specific Deubiquitinase Gumby Regulates Angiogenesis. Nature. 2013 Jun;498(7454):318-24.

[152] Gerlach B, Cordier SM, Schmukle AC, Emmerich CH, Rieser E, Haas TL, et al. Linear Ubiquitination Prevents Inflammation and Regulates Immune Signalling. Nature. 2011 Mar;471(7340):591-6.

[153] Emmerich CH, Ordureau A, Strickson S, Arthur JSC, Pedrioli PGA, Komander D, et al. Activation of the Canonical IKK Complex by K63/M1-linked Hybrid Ubiquitin Chains. Proceedings of the National Academy of Sciences. 2013 Sep;110(38):15247-52.

[154] Emmerich CH, Bakshi S, Kelsall IR, Ortiz-Guerrero J, Shpiro N, Cohen P. Lys63/Met1-hybrid Ubiquitin Chains Are Commonly Formed during the Activation of Innate Immune Signalling. Biochemical and Biophysical Research Communications. 2016 Jun;474(3):452-61.

[155] Wertz IE, Newton K, Seshasayee D, Kusam S, Lam C, Zhang J, et al. Phosphorylation and Linear Ubiquitin Direct A20 Inhibition of Inflammation. Nature. 2015 Dec;528(7582):370-5.

[156] Williamson A, Wickliffe KE, Mellone BG, Song L, Karpen GH, Rape M. Identification of a Physiological E2 Module for the Human Anaphase-Promoting Complex. Proceedings of the National Academy of Sciences. 2009 Oct;106(43):18213-8.

[157] Garnett MJ, Mansfeld J, Godwin C, Matsusaka T, Wu J, Russell P, et al. UBE2S Elongates Ubiquitin Chains on APC/C Substrates to Promote Mitotic Exit. Nature Cell Biology. 2009 Nov;11(11):1363-9.

[158] Wu T, Merbl Y, Huo Y, Gallop JL, Tzur A, Kirschner MW. UBE2S Drives Elongation of K11-linked Ubiquitin Chains by the Anaphase-Promoting Complex. Proceedings of the National Academy of Sciences of the United States of America. 2010 Jan;107(4):1355-60.

[159] Matsumoto ML, Wickliffe KE, Dong KC, Yu C, Bosanac I, Bustos D, et al. K11-Linked Polyubiquitination in Cell Cycle Control Revealed by a K11 Linkage-Specific Antibody. Molecular Cell. 2010 Aug;39(3):477-84.

[160] Meyer HJ, Rape M. Enhanced Protein Degradation by Branched Ubiquitin Chains. Cell. 2014 May;157(4):910-21.

[161] Yau RG, Doerner K, Castellanos ER, Haakonsen DL, Werner A, Wang N, et al. Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control. Cell. 2017 Nov;171(4):918-33.e20.

[162] Leto DE, Morgens DW, Zhang L, Walczak CP, Elias JE, Bassik MC, et al. Genome-Wide CRISPR Analysis Identifies Substrate-Specific Conjugation Modules in ER-Associated Degradation. Molecular Cell. 2019 Jan;73(2):377-89.e11.

[163] Kaiho-Soma A, Akizuki Y, Igarashi K, Endo A, Shoda T, Kawase Y, et al. TRIP12 Promotes Small-Molecule-Induced Degradation through K29/K48-branched Ubiquitin Chains. Molecular Cell. 2021 Apr;81(7):1411-24.e7.

[164] Ohtake F, Tsuchiya H, Saeki Y, Tanaka K. K63 Ubiquitylation Triggers Proteasomal Degradation by Seeding Branched Ubiquitin Chains. Proceedings of the National Academy of Sciences. 2018 Feb;115(7):E1401-8.

[165] Ohtake F, Saeki Y, Ishido S, Kanno J, Tanaka K. The K48-K63 Branched Ubiquitin Chain Regulates NF-κB Signaling. Molecular Cell. 2016 Oct;64(2):251-66.

[166] Gatti M, Pinato S, Maiolica A, Rocchio F, Prato MG, Aebersold R, et al. RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. Cell Reports. 2015 Jan;10(2):226-38.

[167] Fei C, Li Z, Li C, Chen Y, Chen Z, He X, et al. Smurf1-Mediated Lys29-Linked Nonproteolytic Polyubiquitination of Axin Negatively Regulates Wnt/β-Catenin Signaling. Molecular and Cellular Biology. 2013 Oct;33(20):4095-105.

[168] Huang H, Jeon Ms, Liao L, Yang C, Elly C, Yates JR, et al. K33-Linked Polyubiquitination of T Cell Receptor-ζ Regulates T Cell Activation by Modulating Its Endocytosis-Independent Phosphorylation. Immunity. 2010 Jul;33(1):60-70.

[169] Al-Hakim AK, Zagorska A, Chapman L, Deak M, Peggie M, Alessi DR. Control of AMPK-related Kinases by USP9X and Atypical Lys29/Lys33-linked Polyubiquitin Chains. Biochemical Journal. 2008 Mar;411(2):249-60.

[170] Yuan WC, Lee YR, Lin SY, Chang LY, Tan YP, Hung CC, et al. K33-Linked Polyubiquitination of Coronin 7 by Cul3-KLHL20 Ubiquitin E3 Ligase Regulates Protein Trafficking. Molecular Cell. 2014 May;54(4):586-600.

[171] Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, et al. Architecture of RNA Polymerase II and Implications for the Transcription Mechanism. Science. 2000 Apr;288(5466):640-9.

[172] Cramer P, Bushnell DA, Kornberg RD. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science. 2001 Jun;292(5523):1863-76.

[173] Armache KJ, Kettenberger H, Cramer P. Architecture of Initiation-Competent 12-Subunit RNA Polymerase II. Proceedings of the National Academy of Sciences. 2003 Jun;100(12):6964-8.

[174] Bushnell DA, Kornberg RD. Complete, 12-Subunit RNA Polymerase II at 4.1-Å Resolution: Implications for the Initiation of Transcription. Proceedings of the National Academy of Sciences. 2003 Jun;100(12):6969-73.

[175] Aibara S, Schilbach S, Cramer P. Structures of Mammalian RNA Polymerase II Pre-Initiation Complexes. Nature. 2021 Apr:1-8.

[176] Chen X, Qi Y, Wu Z, Wang X, Li J, Zhao D, et al. Structural Insights into Preinitiation Complex Assembly on Core Promoters. Science. 2021 Apr.

[177] Chen X, Yin X, Li J, Wu Z, Qi Y, Wang X, et al. Structures of the Human Mediator and Mediator-bound Preinitiation Complex. Science. 2021 May.

[178] Farnung L, Vos SM, Cramer P. Structure of Transcribing RNA Polymerase II-nucleosome Complex. Nature Communications. 2018 Dec;9(1):5432.

[179] Rengachari S, Schilbach S, Aibara S, Dienemann C, Cramer P. Structure of Human Mediator–RNA Polymerase II Pre-Initiation Complex. Nature. 2021 Apr:1-8.

[180] Wild T, Cramer P. Biogenesis of Multisubunit RNA Polymerases. Trends in Biochemical Sciences. 2012 Mar;37(3):99-105.

[181] Boulon S, Pradet-Balade B, Verheggen C, Molle D, Boireau S, Georgieva M, et al. HSP90 and Its R2TP/Prefoldin-like Cochaperone Are Involved in the Cytoplasmic Assembly of RNA Polymerase II. Molecular Cell. 2010 Sep;39(6):912-24.

[182] Calera MR, Zamora-Ramos C, Araiza-Villanueva MG, Moreno-Aguilar CA, Peña-Gómez SG, Castellanos-Terán F, et al. Parcs/Gpn3 Is Required for the Nuclear Accumulation of RNA Polymerase II. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research. 2011 Oct;1813(10):1708-16.

[183] Forget D, Lacombe AA, Cloutier P, Al-Khoury R, Bouchard A, Lavallée-Adam M, et al. The Protein Interaction Network of the Human Transcription Machinery Reveals a Role for the Conserved GTPase RPAP4/GPN1 and Microtubule Assembly in Nuclear Import and Biogenesis of RNA Polymerase II. Molecular & Cellular Proteomics. 2010 Dec;9(12):2827-39.

[184] Lynch CJ, Bernad R, Calvo I, Nóbrega-Pereira S, Ruiz S, Ibarz N, et al. The RNA Polymerase II Factor RPAP1 Is Critical for Mediator-Driven Transcription and Cell Identity. Cell Reports. 2018 Jan;22(2):396-410.

[185] Carré C, Shiekhattar R. Human GTPases Associate with RNA Polymerase II To Mediate Its Nuclear Import. Molecular and Cellular Biology. 2011 Oct;31(19):3953-62.

[186] Czeko E, Seizl M, Augsberger C, Mielke T, Cramer P. Iwr1 Directs RNA Polymerase II Nuclear Import. Molecular Cell. 2011 Apr;42(2):261-6.

[187] Chapman RD, Heidemann M, Hintermair C, Eick D. Molecular Evolution of the RNA Polymerase II CTD. Trends in Genetics. 2008 Jun;24(6):289-96.

[188] West ML, Corden JL. Construction and Analysis of Yeast RNA Polymerase II CTD Deletion and Substitution Mutations. Genetics. 1995 Aug;140(4):1223-33.

[189] Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. Cell

Reports. 2016 Jun;15(10):2147-58.

[190] Schwer B, Shuman S. Deciphering the RNA Polymerase II CTD Code in Fission Yeast. Molecular Cell. 2011 Jul;43(2):311-8.

[191] Corden JL. RNA Polymerase II C-Terminal Domain: Tethering Transcription to Transcript and Template. Chemical Reviews. 2013 Nov;113(11):8423-55.

[192] Heidemann M, Hintermair C, Voß K, Eick D. Dynamic Phosphorylation Patterns of RNA Polymerase II CTD during Transcription. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2013 Jan;1829(1):55-62.

[193] Lu L, Fan D, Hu CW, Worth M, Ma ZX, Jiang J. Distributive O-GlcNAcylation on the Highly Repetitive C-Terminal Domain of RNA Polymerase II. Biochemistry. 2016 Feb;55(7):1149-58.

[194] Li H, Zhang Z, Wang B, Zhang J, Zhao Y, Jin Y. Wwp2-Mediated Ubiquitination of the RNA Polymerase II Large Subunit in Mouse Embryonic Pluripotent Stem Cells. Molecular and Cellular Biology. 2007 Aug;27(15):5296-305.

[195] Dias JD, Rito T, Torlai Triglia E, Kukalev A, Ferrai C, Chotalia M, et al. Methylation of RNA Polymerase II Non-Consensus Lysine Residues Marks Early Transcription in Mammalian Cells. eLife. 2015 Dec;4:e11215.

[196] Voss K, Forné I, Descostes N, Hintermair C, Schüller R, Maqbool MA, et al. Site-Specific Methylation and Acetylation of Lysine Residues in the C-terminal Domain (CTD) of RNA Polymerase II. Transcription. 2015;6(5):91-101.

[197] Zhao DY, Gish G, Braunschweig U, Li Y, Ni Z, Schmitges FW, et al. SMN and Symmetric Arginine Dimethylation of RNA Polymerase II C-terminal Domain Control Termination. Nature. 2016 Jan;529(7584):48-53.

[198] Schüller R, Forné I, Straub T, Schreieck A, Texier Y, Shah N, et al. Heptad-Specific Phosphorylation of RNA Polymerase II CTD. Molecular Cell. 2016 Jan;61(2):305-14.

[199] Suh H, Ficarro SB, Kang UB, Chun Y, Marto JA, Buratowski S. Direct Analysis of Phosphorylation Sites on the Rpb1 C-Terminal Domain of RNA Polymerase II. Molecular Cell. 2016 Jan;61(2):297-304.

[200] Bonnet F, Vigneron M, Bensaude O, Dubois MF. Transcription-Independent Phosphorylation of the RNA Polymerase IIC-terminal Domain (CTD) Involves ERK Kinases (MEK1/2). Nucleic Acids Research. 1999 Nov;27(22):4399-404.

[201] Bellier S, Chastant S, Adenot P, Vincent M, Renard JP, Bensaude O. Nuclear Translocation and Carboxyl-Terminal Domain Phosphorylation of RNA Polymerase II Delineate the Two Phases of Zygotic Gene Activation in Mammalian Embryos. Embo Journal. 1997 Oct;16(20):6250-62.

[202] Jenkins HL, Spencer CA. RNA Polymerase II Holoenzyme Modifications Accompany Transcription Reprogramming in Herpes Simplex Virus Type 1-Infected Cells. Journal of Virology. 2001 Oct;75(20):9872-84.

[203] Harlen KM, Churchman LS. The Code and beyond: Transcription Regulation by the RNA Polymerase II Carboxy-Terminal Domain. Nature Reviews Molecular Cell Biology. 2017 Apr;18(4):263-73.

[204] Hahn S, Buratowski S, Sharp PA, Guarente L. Isolation of the Gene Encoding the Yeast TATA Binding Protein TFIID: A Gene Identical to the SPT15 Suppressor of Ty Element Insertions. Cell. 1989 Sep;58(6):1173-81.

[205] Horikoshi M, Wang CK, Fujii H, Cromlish JA, Weil PA, Roeder RG. Cloning and Structure of a Yeast Gene Encoding a General Transcription Initiation Factor TFIID That Binds to the TATA Box. Nature. 1989 Sep;341(6240):299-303.

[206] Kuras L, Kosa P, Mencia M, Struhl K. TAF-Containing and TAF-independent Forms of Transcriptionally Active TBP in Vivo. Science (New York, NY). 2000 May;288(5469):1244-8.

[207] Basehoar AD, Zanton SJ, Pugh BF. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. Cell. 2004 Mar;116(5):699-709.

[208] Nikolov DB, Hu SH, Lin J, Gasch A, Hoffmann A, Horikoshi M, et al. Crystal Structure of TFIID TATA-box Binding Protein. Nature. 1992 Nov;360(6399):40-6.

[209] Yang C, Bolotin E, Jiang T, Sladek FM, Martinez E. Prevalence of the Initiator over the TATA Box in Human and Yeast Genes and Identification of DNA Motifs Enriched in Human TATA-less Core Promoters. Gene. 2007 Mar;389(1):52-65.

[210] Chalkley GE, Verrijzer CP. DNA Binding Site Selection by RNA Polymerase II TAFs: A TAFII250–TAFII150 Complex Recognizes the Initiator. The EMBO Journal. 1999 Sep;18(17):4835-45.

[211] Burke TW, Kadonaga JT. The Downstream Core Promoter Element, DPE, Is Conserved from Drosophila to Humans and Is Recognized by TAFII60 of Drosophila. Genes & Development. 1997 Nov;11(22):3020-31.

[212] Theisen JWM, Lim CY, Kadonaga JT. Three Key Subregions Contribute to the Function of the Downstream RNA Polymerase II Core Promoter. Molecular and Cellular Biology. 2010 Jul;30(14):3471-9.

[213] Lim CY, Santoso B, Boulay T, Dong E, Ohler U, Kadonaga JT. The MTE, a New Core Promoter Element for Transcription by RNA Polymerase II. Genes & Development. 2004 Jul;18(13):1606-17.

[214] Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. Functional Characterization of Core Promoter Elements: The Downstream Core Element Is Recognized by TAF1. Molecular and

Cellular Biology. 2005 Nov;25(21):9674-86.

[215] Flores O, Maldonado E, Reinberg D. Factors Involved in Specific Transcription by Mammalian RNA Polymerase II: Factors IIE and IIF Independently Interact with RNA Polymerase II. Journal of Biological Chemistry. 1989 May;264(15):8913-21.

[216] Maxon ME, Goodrich JA, Tjian R. Transcription Factor IIE Binds Preferentially to RNA Polymerase IIa and Recruits TFIIH: A Model for Promoter Clearance. Genes & Development. 1994 Mar;8(5):515-24.

[217] Gibbons BJ, Brignole EJ, Azubel M, Murakami K, Voss NR, Bushnell DA, et al. Subunit Architecture of General Transcription Factor TFIIH. Proceedings of the National Academy of Sciences. 2012 Feb;109(6):1949-54.

[218] Max T, Søgaard M, Svejstrup JQ. Hyperphosphorylation of the C-terminal Repeat Domain of RNA Polymerase II Facilitates Dissociation of Its Complex with Mediator *. Journal of Biological Chemistry. 2007 May;282(19):14113-20.

[219] Kim YJ, Björklund S, Li Y, Sayre MH, Kornberg RD. A Multiprotein Mediator of Transcriptional Activation and Its Interaction with the C-terminal Repeat Domain of RNA Polymerase II. Cell. 1994 May;77(4):599-608.

[220] Koh SS, Ansari AZ, Ptashne M, Young RA. An Activator Target in the RNA Polymerase II Holoenzyme. Molecular Cell. 1998 May;1(6):895-904.

[221] Lu H, Flores O, Weinmann R, Reinberg D. The Nonphosphorylated Form of RNA Polymerase II Preferentially Associates with the Preinitiation Complex. Proceedings of the National Academy of Sciences. 1991 Nov;88(22):10004-8.

[222] Jeronimo C, Robert F. Kin28 Regulates the Transient Association of Mediator with Core Promoters. Nature Structural & Molecular Biology. 2014 May;21(5):449-55.

[223] Topisirovic I, Svitkin YV, Sonenberg N, Shatkin AJ. Cap and Cap-Binding Proteins in the Control of Gene Expression. WIREs RNA. 2011;2(2):277-98.

[224] Ho CK, Shuman S. Distinct Roles for CTD Ser-2 and Ser-5 Phosphorylation in the Recruitment and Allosteric Activation of Mammalian mRNA Capping Enzyme. Molecular Cell. 1999 Mar;3(3):405-11.

[225] Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH. A Highly Purified RNA Polymerase II Elongation Control System*. Journal of Biological Chemistry. 2001 Nov;276(45):42601-9.

[226] Fujinaga K, Irwin D, Huang Y, Taube R, Kurosu T, Peterlin BM. Dynamics of Human Immunodeficiency Virus Transcription: P-TEFb Phosphorylates RD and Dissociates Negative Effectors from the Transactivation Response Element. Molecular and Cellular Biology. 2004 Jan;24(2):787-95.

[227] Yamada T, Yamaguchi Y, Inukai N, Okamoto S, Mura T, Handa H. P-TEFb-Mediated Phosphorylation of hSpt5 C-Terminal Repeats Is Critical for Processive Transcription Elongation. Molecular Cell. 2006 Jan;21(2):227-37.

[228] Chen FX, Woodfin AR, Gardini A, Rickels RA, Marshall SA, Smith ER, et al. PAF1, a Molecular Regulator of Promoter-Proximal Pausing by RNA Polymerase II. Cell. 2015 Aug;162(5):1003-15.

[229] Zhang W, Prakash C, Sum C, Gong Y, Li Y, Kwok JJT, et al. Bromodomain-Containing Protein 4 (BRD4) Regulates RNA Polymerase II Serine 2 Phosphorylation in Human CD4+ T Cells*. Journal of Biological Chemistry. 2012 Dec;287(51):43137-55.

[230] Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, et al. RNA Polymerase Stalling at Developmental Control Genes in the Drosophila Melanogaster Embryo. Nature Genetics. 2007 Dec;39(12):1512-6.

[231] Shilatifard A, Conaway RC, Conaway JW. The RNA Polymerase II Elongation Complex. Annual Review of Biochemistry. 2003 Jun;72(1):693-715.

[232] Hoeijmakers JHJ. Genome Maintenance Mechanisms for Preventing Cancer. Nature. 2001 May;411(6835):366-74.

[233] Lindahl T. Instability and Decay of the Primary Structure of DNA. Nature. 1993 Apr;362(6422):709-15.

[234] Brégeon D, Doetsch PW. Transcriptional Mutagenesis: Causes and Involvement in Tumour Development. Nature Reviews Cancer. 2011 Mar;11(3):218-27.

[235] Brueckner F, Hennecke U, Carell T, Cramer P. CPD Damage Recognition by Transcribing RNA Polymerase II. Science. 2007 Feb;315(5813):859-62.

[236] Donahue BA, Fuchs RP, Reines D, Hanawalt PC. Effects of Aminofluorene and Acetylaminofluorene DNA Adducts on Transcriptional Elongation by RNA Polymerase II. The Journal of Biological Chemistry. 1996 May;271(18):10588-94.

[237] Brooks PJ, Wise DS, Berry DA, Kosmoski JV, Smerdon MJ, Somers RL, et al. The Oxidative DNA Lesion 8,5'-(S)-Cyclo-2'-Deoxyadenosine Is Repaired by the Nucleotide Excision Repair Pathway and Blocks Gene Expression in Mammalian Cells. The Journal of Biological Chemistry. 2000 Jul;275(29):22355-62.

[238] Wang W, Xu J, Chong J, Wang D. Structural Basis of DNA Lesion Recognition for Eukaryotic Transcription-Coupled Nucleotide Excision Repair. DNA Repair. 2018 Nov;71:43-55.

[239] Mullenders L. DNA Damage Mediated Transcription Arrest: Step Back to Go Forward. DNA Repair. 2015 Dec;36:28-35.

[240] Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding Nucleotide Excision Repair and Its Roles in Cancer and Ageing. Nature Reviews Molecular Cell Biology. 2014 Jul;15(7):465-81.

[241] Carrillo Oesterreich F, Preibisch S, Neugebauer KM. Global Analysis of Nascent RNA Reveals Transcriptional Pausing in Terminal Exons. Molecular Cell. 2010 Nov;40(4):571-81.

[242] Oesterreich FC, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. Cell. 2016 Apr;165(2):372-81.

[243] Kastner B, Will CL, Stark H, Lührmann R. Structural Insights into Nuclear Pre-mRNA Splicing in Higher Eukaryotes. Cold Spring Harbor Perspectives in Biology. 2019 Nov;11(11):a032417.

[244] Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, et al. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. Cell. 2015 Apr;161(3):526-40.

[245] Gu B, Eick D, Bensaude O. CTD Serine-2 Plays a Critical Role in Splicing and Termination Factor Recruitment to RNA Polymerase II in Vivo. Nucleic Acids Research. 2013 Feb;41(3):1591-603.

[246] Eaton JD, West S. Termination of Transcription by RNA Polymerase II: BOOM! Trends in Genetics. 2020 Sep;36(9):664-75.

[247] Mayer A, Heidemann M, Lidschreiber M, Schreieck A, Sun M, Hintermair C, et al. CTD Tyrosine Phosphorylation Impairs Termination Factor Recruitment to RNA Polymerase II. Science. 2012 Jun;336(6089):1723-5.

[248] Houseley J, Tollervey D. The Many Pathways of RNA Degradation. Cell. 2009 Feb;136(4):763-76.

[249] Porrua O, Libri D. Transcription Termination and the Control of the Transcriptome: Why, Where and How to Stop. Nature Reviews Molecular Cell Biology. 2015 Mar;16(3):190-202.

[250] Vasiljeva L, Kim M, Mutschler H, Buratowski S, Meinhart A. The Nrd1–Nab3–Sen1 Termination Complex Interacts with the Ser5-phosphorylated RNA Polymerase II C-terminal Domain. Nature Structural & Molecular Biology. 2008 Aug;15(8):795-804.

[251] Kim M, Krogan NJ, Vasiljeva L, Rando OJ, Nedea E, Greenblatt JF, et al. The Yeast Rat1 Exonuclease Promotes Transcription Termination by RNA Polymerase II. Nature. 2004 Nov;432(7016):517-22.

[252] Licatalosi DD, Geiger G, Minet M, Schroeder S, Cilli K, McNeil JB, et al. Functional Interaction of Yeast Pre-mRNA 3′ End Processing Factors with RNA Polymerase II. Molecular Cell. 2002 May;9(5):1101-11.

[253] Ansari A, Hampsey M. A Role for the CPF 3′-End Processing Machinery in RNAP II-dependent Gene Looping. Genes & Development. 2005 Dec;19(24):2969-78.

[254] O'Sullivan JM, Tan-Wong SM, Morillon A, Lee B, Coles J, Mellor J, et al. Gene Loops Juxtapose Promoters and Terminators in Yeast. Nature Genetics. 2004 Sep;36(9):1014-8.

[255] Perkins KJ, Lusic M, Mitar I, Giacca M, Proudfoot NJ. Transcription-Dependent Gene Looping of the HIV-1 Provirus Is Dictated by Recognition of Pre-mRNA Processing Signals. Molecular Cell. 2008 Jan;29(1):56-68.

[256] Yun K, So JS, Jash A, Im SH. Lymphoid Enhancer Binding Factor 1 Regulates Transcription through Gene Looping. The Journal of Immunology. 2009 Oct;183(8):5129-37.

[257] O'Reilly D, Greaves DR. Cell-Type-Specific Expression of the Human CD68 Gene Is Associated with Changes in Pol II Phosphorylation and Short-Range Intrachromosomal Gene Looping. Genomics. 2007 Sep;90(3):407-15.

[258] Singh I, Lee SH, Sperling AS, Samur MK, Tai YT, Fulciniti M, et al. Widespread Intronic Polyadenylation Diversifies Immune Cell Transcriptomes. Nature Communications. 2018 Apr;9(1):1716.

[259] Krebs AR, Imanci D, Hoerner L, Gaidatzis D, Burger L, Schübeler D. Genome-Wide Single-Molecule Footprinting Reveals High RNA Polymerase II Turnover at Paused Promoters. Molecular Cell. 2017 Aug;67(3):411-22.e4.

[260] Steurer B, Janssens RC, Geverts B, Geijer ME, Wienholz F, Theil AF, et al. Live-Cell Analysis of Endogenous GFP-RPB1 Uncovers Rapid Turnover of Initiating and Promoter-Paused RNA Polymerase II. Proceedings of the National Academy of Sciences of the United States of America. 2018 Apr.

[261] Kim W, Bennett EJ, Huttlin EL, Guo A, Li J, Possemato A, et al. Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome. Molecular Cell. 2011 Oct;44(2):325-40.

[262] Wagner SA, Beli P, Weinert BT, Nielsen ML, Cox J, Mann M, et al. A Proteome-wide, Quantitative Survey of In Vivo Ubiquitylation Sites Reveals Widespread Regulatory Roles. Molecular & Cellular Proteomics. 2011 Oct;10(10):M111.013284.

[263] Nakazawa Y, Hara Y, Oka Y, Komine O, van den Heuvel D, Guo C, et al. Ubiquitination of DNA Damage-Stalled RNAPII Promotes Transcription-Coupled Repair. Cell. 2020 Mar;180(6):1228-44.e24.

[264] Tufegdžić Vidaković A, Mitter R, Kelly GP, Neumann M, Harreman M, Rodríguez-Martínez M, et al. Regulation of the RNAPII Pool Is Integral to the DNA Damage Response. Cell. 2020 Mar;180(6):1245-61.e21.

[265] Beltrao P, Albanèse V, Kenner LR, Swaney DL, Burlingame A, Villén J, et al. Systematic Functional Prioritization of Protein Posttranslational Modifications. Cell. 2012 Jul;150(2):413-25.

[266] Huibregtse JM, Yang JC, Beaudenon SL. The Large Subunit of RNA Polymerase II Is a Substrate of the Rsp5 Ubiquitin-Protein Ligase. Proceedings of the National Academy of Sciences. 1997 Apr;94(8):3656-61.

[267] Beaudenon SL, Huacani MR, Wang G, McDonnell DP, Huibregtse JM. Rsp5 Ubiquitin-Protein Ligase Mediates DNA Damage-Induced Degradation of the Large Subunit of RNA Polymerase II in *Saccharomyces Cerevisiae*. Molecular and Cellular Biology. 1999 Oct;19(10):6972-9.

[268] Harreman M, Taschner M, Sigurdsson S, Anindya R, Reid J, Somesh B, et al. Distinct Ubiquitin Ligases Act Sequentially for RNA Polymerase II Polyubiquitylation. Proceedings of the National Academy of

Sciences of the United States of America. 2009;106(49):20705-10.

[269] Kee Y, Lyon N, Huibregtse JM. The Rsp5 Ubiquitin Ligase Is Coupled to and Antagonized by the Ubp2 Deubiquitinating Enzyme. The EMBO Journal. 2005 Jul;24(13):2414-24.

[270] Daulny A, Geng F, Muratani M, Geisinger JM, Salghetti SE, Tansey WP. Modulation of RNA Polymerase II Subunit Composition by Ubiquitylation. Proceedings of the National Academy of Sciences. 2008 Dec;105(50):19649-54.

[271] Kvint K, Uhler JP, Taschner MJ, Sigurdsson S, Erdjument-Bromage H, Tempst P, et al. Reversal of RNA Polymerase II Ubiquitylation by the Ubiquitin Protease Ubp3. Molecular cell. 2008;30(4):498-506.

[272] Starita LM, Horwitz AA, Keogh MC, Ishioka C, Parvin JD, Chiba N. BRCA1/BARD1 Ubiquitinate Phosphorylated RNA Polymerase II. Journal of Biological Chemistry. 2005 Jul;280(26):24498-505.

[273] Kleiman FE, Wu-Baer F, Fonseca D, Kaneko S, Baer R, Manley JL. BRCA1/BARD1 Inhibition of mRNA 3′ Processing Involves Targeted Degradation of RNA Polymerase II. Genes & Development. 2005 May;19(10):1227-37.

[274] Bregman DB, Halaban R, van Gool AJ, Henning KA, Friedberg EC, Warren SL. UV-induced Ubiquitination of RNA Polymerase II: A Novel Modification Deficient in Cockayne Syndrome Cells. Proceedings of the National Academy of Sciences. 1996 Oct;93(21):11586-90.

[275] Anindya R, Aygün O, Svejstrup JQ. Damage-Induced Ubiquitylation of Human RNA Polymerase II by the Ubiquitin Ligase Nedd4, but Not Cockayne Syndrome Proteins or BRCA1. Molecular Cell. 2007 Nov;28(3):386-97.

[276] Mayne LV, Lehmann AR. Failure of RNA Synthesis to Recover after UV Irradiation: An Early Defect in Cells from Individuals with Cockayne's Syndrome and Xeroderma Pigmentosum. Cancer Research. 1982 Apr;42(4):1473-8.

[277] Kamura T, Burian D, Yan Q, Schmidt SL, Lane WS, Querido E, et al. MUF1, A Novel Elongin BC-interacting Leucine-rich Repeat Protein That Can Assemble with Cul5 and Rbx1 to Reconstitute a Ubiquitin Ligase. Journal of Biological Chemistry. 2001 Aug;276(32):29748-53.

[278] Mikhaylova O, Ignacak ML, Barankiewicz TJ, Harbaugh SV, Yi Y, Maxwell PH, et al. The von Hippel-Lindau Tumor Suppressor Protein and Egl-9-Type Proline Hydroxylases Regulate the Large Subunit of RNA Polymerase II in Response to Oxidative Stress. Molecular and Cellular Biology. 2008 Apr;28(8):2701-17.

[279] Kuznetsova AV, Meller J, Schnell PO, Nash JA, Ignacak ML, Sanchez Y, et al. Von Hippel–Lindau Protein Binds Hyperphosphorylated Large Subunit of RNA Polymerase II through a Proline Hydroxylation Motif and Targets It for Ubiquitination. Proceedings of the National Academy of Sciences of the United States of America. 2003 Mar;100(5):2706-11.

[280] Ribar B, Prakash L, Prakash S. Requirement of ELC1 for RNA Polymerase II Polyubiquitylation and Degradation in Response to DNA Damage in Saccharomyces Cerevisiae. Molecular and cellular biology. 2006;26(11):3999-4005.

[281] Ribar B, Prakash L, Prakash S. ELA1 and CUL3 Are Required along with ELC1 for RNA Polymerase II Polyubiquitylation and Degradation in DNA-damaged Yeast Cells. Molecular and cellular biology. 2007;27(8):3211-6.

[282] Yasukawa T, Kamura T, Kitajima S, Conaway RC, Conaway JW, Aso T. Mammalian Elongin A Complex Mediates DNA-damage-induced Ubiquitylation and Degradation of Rpb1. The EMBO Journal. 2008 Dec;27(24):3256-66.

[283] Sun J, Keim CD, Wang J, Kazadi D, Oliver PM, Rabadan R, et al. E3-Ubiquitin Ligase Nedd4 Determines the Fate of AID-associated RNA Polymerase II in B Cells. Genes & Development. 2013 Aug;27(16):1821-33.

[284] van der Weegen Y, de Lint K, van den Heuvel D, Nakazawa Y, Mevissen TET, van Schie JJM, et al. ELOF1 Is a Transcription-Coupled DNA Repair Factor That Directs RNA Polymerase II Ubiquitylation. Nature Cell Biology. 2021 Jun;23(6):595-607.

[285] Mokuda S, Nakamichi R, Matsuzaki T, Ito Y, Sato T, Miyata K, et al. Wwp2 Maintains Cartilage Homeostasis through Regulation of Adamts5. Nature Communications. 2019 Jun;10(1):2429.

[286] Kurihara T, Kubota Y, Ozawa Y, Takubo K, Noda K, Simon MC, et al. Von Hippel-Lindau Protein Regulates Transition from the Fetal to the Adult Circulatory System in Retina. Development (Cambridge, England). 2010 May;137(9):1563-71.

[287] Nusslein-Volhard C, Wieschaus E. Mutations Affecting Segment Number and Polarity in Drosophila. Nature. 1980 Oct;287(5785):795-801.

[288] Wieschaus E, Riggleman R. Autonomous Requirements for the Segment Polarity Gene Armadillo during Drosophila Embryogenesis. Cell. 1987 Apr;49(2):177-84.

[289] Ozawa M, Baribault H, Kemler R. The Cytoplasmic Domain of the Cell Adhesion Molecule Uvomorulin Associates with Three Independent Proteins Structurally Related in Different Species. Embo j. 1989 Jun;8(6):1711-7.

[290] Huber AH, Nelson WJ, Weis WI. Three-Dimensional Structure of the Armadillo Repeat Region of β-Catenin. Cell. 1997 Sep;90(5):871-82.

[291] Tewari R, Bailes E, Bunting KA, Coates JC. Armadillo-Repeat Protein Functions: Questions for Little Creatures. Trends in Cell Biology. 2010 Aug;20(8):470-81.

[292] Hatzfeld M. The Armadillo Family of Structural Proteins. Int Rev Cytol. 1999;186:179-224.

[293] Conti E, Uy M, Leighton L, Blobel G, Kuriyan J. Crystallographic Analysis of the Recognition of a Nuclear Localization Signal by the Nuclear Import Factor Karyopherin α. Cell. 1998 Jul;94(2):193-204.

[294] Hatzfeld M, Haffner C, Schulze K, Vinzens U. The Function of Plakophilin 1 in Desmosome Assembly and Actin Filament Organization. Journal of Cell Biology. 2000 Apr;149(1):209-22.

[295] Keil R, Hatzfeld M. The Armadillo Protein P0071 Is Involved in Rab11-dependent Recycling. Journal of Cell Science. 2014 Jan;127(1):60-71.

[296] Roberts DM, Pronobis MI, Poulton JS, Kane EG, Peifer M. Regulation of Wnt Signaling by the Tumor Suppressor Adenomatous Polyposis Coli Does Not Require the Ability to Enter the Nucleus or a Particular Cytoplasmic Localization. Mol Biol Cell. 2012 Jun;23(11):2041-56.

[297] Zhao G, Li G, Schindelin H, Lennarz WJ. An Armadillo Motif in Ufd3 Interacts with Cdc48 and Is Involved in Ubiquitin Homeostasis and Protein Degradation. Proceedings of the National Academy of Sciences. 2009 Sep;106(38):16197-202.

[298] Pipino C, Di Tomo P, Mandatori D, Cianci E, Lanuti P, Cutrona MB, et al. Calcium Sensing Receptor Activation by Calcimimetic R-568 in Human Amniotic Fluid Mesenchymal Stem Cells: Correlation with Osteogenic Differentiation. Stem Cells Dev. 2014 Dec;23(24):2959-71.

[299] Tan F, Qian C, Tang K, Abd-Allah SM, Jing N. Inhibition of Transforming Growth Factor Beta (TGF-beta) Signaling Can Substitute for Oct4 Protein in Reprogramming and Maintain Pluripotency. J Biol Chem. 2015 Feb;290(7):4500-11.

[300] Hillesheim A, Nordhoff C, Boergeling Y, Ludwig S, Wixler V. Beta-Catenin Promotes the Type I IFN Synthesis and the IFN-dependent Signaling Response but Is Suppressed by Influenza A Virus-Induced RIG-I/NF-kappaB Signaling. Cell Commun Signal. 2014;12:29.

[301] Liu CC, Pearson C, Bu G. Cooperative Folding and Ligand-Binding Properties of LRP6 Beta-Propeller Domains. J Biol Chem. 2009 May;284(22):15299-307.

[302] Bujalowski PJ, Nicholls P, Barral JM, Oberhauser AF. Thermally-Induced Structural Changes in an Armadillo Repeat Protein Suggest a Novel Thermosensor Mechanism in a Molecular Chaperone. FEBS Letters. 2015 Jan;589(1):123-30.

[303] Xie C, Jiang G, Fan C, Zhang X, Zhang Y, Miao Y, et al. ARMC8alpha Promotes Proliferation and Invasion of Non-Small Cell Lung Cancer Cells by Activating the Canonical Wnt Signaling Pathway. Tumour Biol. 2014 Sep;35(9):8903-11.

[304] Riese J, Yu X, Munnerlyn A, Eresh S, Hsu SC, Grosschedl R, et al. LEF-1, a Nuclear Factor Coordinating Signaling Inputs from Wingless and Decapentaplegic. Cell. 1997 Mar;88(6):777-87.

[305] Iguchi H, Urashima Y, Inagaki Y, Ikeda Y, Okamura M, Tanaka T, et al. SOX6 Suppresses Cyclin D1 Promoter Activity by Interacting with Beta-Catenin and Histone Deacetylase 1, and Its down-Regulation Induces Pancreatic Beta-Cell Proliferation. J Biol Chem. 2007 Jun;282(26):19052-61.

[306] Kim S, Ishidate T, Sharma R, Soto MC, Conte D, Mello CC, et al. Wnt and CDK-1 Regulate Cortical Release of WRM-1/Beta-Catenin to Control Cell Division Orientation in Early Caenorhabditis Elegans Embryos. Proc Natl Acad Sci U S A. 2013 Mar;110(10):E918-27.

[307] Xu Y, Banerjee D, Huelsken J, Birchmeier W, Sen JM. Deletion of Beta-Catenin Impairs T Cell Development. Nat Immunol. 2003 Dec;4(12):1177-82.

[308] Mucenski ML, Wert SE, Nation JM, Loudy DE, Huelsken J, Birchmeier W, et al. Beta-Catenin Is Required for Specification of Proximal/Distal Cell Fate during Lung Morphogenesis. J Biol Chem. 2003 Oct;278(41):40231-8.

[309] Soshnikova N, Zechner D, Huelsken J, Mishina Y, Behringer RR, Taketo MM, et al. Genetic Interaction between Wnt/Beta-Catenin and BMP Receptor Signaling during Formation of the AER and the Dorsal-Ventral Axis in the Limb. Genes Dev. 2003 Aug;17(16):1963-8.

[310] Zhao T, Gan Q, Stokes A, Lassiter RN, Wang Y, Chan J, et al. Beta-Catenin Regulates Pax3 and Cdx2 for Caudal Neural Tube Closure and Elongation. Development. 2014 Jan;141(1):148-57.

[311] Dao DY, Jonason JH, Zhang Y, Hsu W, Chen D, Hilton MJ, et al. Cartilage-Specific Beta-Catenin Signaling Regulates Chondrocyte Maturation, Generation of Ossification Centers, and Perichondrial Bone Formation during Skeletal Development. J Bone Miner Res. 2012 Aug;27(8):1680-94.

[312] Guo X, Day TF, Jiang X, Garrett-Beal L, Topol L, Yang Y. Wnt/Beta-Catenin Signaling Is Sufficient and Necessary for Synovial Joint Formation. Genes Dev. 2004 Oct;18(19):2404-17.

[313] Kim AC, Reuter AL, Zubair M, Else T, Serecky K, Bingham NC, et al. Targeted Disruption of Beta-Catenin in Sf1-expressing Cells Impairs Development and Maintenance of the Adrenal Cortex. Development. 2008 Aug;135(15):2593-602.

[314] Simcha I, Geiger B, Yehuda-Levenberg S, Salomon D, Ben-Ze'ev A. Suppression of Tumorigenicity by Plakoglobin: An Augmenting Effect of N-cadherin. J Cell Biol. 1996 Apr;133(1):199-209.

[315] NCBI. Armc5 Armadillo Repeat Containing 5 [Mus Musculus (House Mouse)]; 2020. `https://www.ncbi.nlm.nih.gov/gene/233912`. Website.

[316] NCBI. ARMC5 Armadillo Repeat Containing 5 [Homo Sapiens (Human)]; 2021. `https://www.ncbi.nlm.nih.gov/gene/79798`. Website.

[317] Bardwell VJ, Treisman R. The POZ Domain: A Conserved Protein-Protein Interaction Motif. Genes Dev. 1994 Jul;8(14):1664-77.

[318] Ahmad KF, Engel CK, Prive GG. Crystal Structure of the BTB Domain from PLZF. Proceedings of the National Academy of Sciences. 1998 Oct;95(21):12123-8.

[319] Zollman S, Godt D, Privé GG, Couderc JL, Laski FA. The BTB Domain, Found Primarily in Zinc Finger Proteins, Defines an Evolutionarily Conserved Family That Includes Several Developmentally Regulated Genes in Drosophila. Proc Natl Acad Sci U S A. 1994 Oct;91(22):10717-21.

[320] Sarkar S, Kalia V, Haining WN, Konieczny BT, Subramaniam S, Ahmed R. Functional and Genomic Profiling of Effector CD8 T Cell Subsets with Distinct Memory Fates. The Journal of Experimental Medicine. 2008 Mar;205(3):625-40.

[321] Formstecher E, Aresta S, Collura V, Hamburger A, Meil A, Trehin A, et al. Protein Interaction Mapping: A Drosophila Case Study. Genome Research. 2005 Mar;15(3):376-84.

[322] Marcinkiewicz M. BetaAPP and Furin mRNA Concentrates in Immature Senile Plaques in the Brain of Alzheimer Patients. J Neuropathol Exp Neurol. 2002 Sep;61(9):815-29.

[323] Luo H, Charpentier T, Wang X, Qi S, Han B, Wu T, et al. Efnb1 and Efnb2 Proteins Regulate Thymocyte Development, Peripheral T Cell Differentiation, and Antiviral Immune Responses and Are Essential for Interleukin-6 (IL-6) Signaling. The Journal of Biological Chemistry. 2011 Dec;286(48):41135-52.

[324] Terra R, Wang X, Hu Y, Charpentier T, Lamarre A, Zhong M, et al. To Investigate the Necessity of STRA6 Upregulation in T Cells during T Cell Immune Responses. PLOS ONE. 2013;8(12):e82808.

[325] Lacasse P, Denis J, Lapointe R, Leclerc D, Lamarre A. Novel Plant Virus-Based Vaccine Induces Protective Cytotoxic T-lymphocyte-mediated Antiviral Immunity through Dendritic Cell Maturation. Journal of Virology. 2008 Jan;82(2):785-94.

[326] Han B, Moore PA, Wu J, Luo H. Overexpression of Human Decoy Receptor 3 in Mice Results in a Systemic Lupus Erythematosus-like Syndrome. Arthritis Rheum. 2007 Nov;56(11):3748-58.

[327] Luo H, Yu G, Tremblay J, Wu J. EphB6-null Mutation Results in Compromised T Cell Function. The Journal of Clinical Investigation. 2004 Dec;114(12):1762-73.

[328] Bright JJ, Du C, Coon M, Sriram S, Klaus SJ. Prevention of Experimental Allergic Encephalomyelitis via Inhibition of IL-12 Signaling and IL-12-mediated Th1 Differentiation: An Effect of the Novel Anti-Inflammatory Drug Lisofylline. J Immunol. 1998 Dec;161(12):7015-22.

[329] Butterfield RJ, Blankenhorn EP, Roper RJ, Zachary JF, Doerge RW, Teuscher C. Identification of Genetic Loci Controlling the Characteristics and Severity of Brain and Spinal Cord Lesions in Experimental Allergic Encephalomyelitis. Am J Pathol. 2000 Aug;157(2):637-45.

[330] Vojtek AB, Hollenberg SM. Ras-Raf Interaction: Two-Hybrid Analysis. Methods in enzymology. 1995;255:331-42.

[331] Beranger F, Aresta S, de Gunzburg J, Camonis J. Getting More from the Two-Hybrid System: N-terminal Fusions to LexA Are Efficient and Sensitive Baits for Two-Hybrid Studies. Nucleic Acids Research. 1997 May;25(10):2035-6.

[332] Bartel PL, Chien CT, Sternglanz R, Fields S. Using the Two-Hybrid System to Detect Protein-Protein Interactions. Cellular interactions in development: A practical approach. 1993;18.

[333] Fromont-Racine M, Rain JC, Legrain P. Toward a Functional Analysis of the Yeast Genome through Exhaustive Two-Hybrid Screens. Nature Genetics. 1997 Jul;16(3):277-82.

[334] Villicaña C, Cruz G, Zurita M. The Basal Transcription Machinery as a Target for Cancer Therapy. Cancer Cell International. 2014 Feb;14(1):18.

[335] Young RA. Rna Polymerase Ii. Annual Review of Biochemistry. 1991 Jun;60(1):689-715.

[336] Edenberg ER, Downey M, Toczyski D. Polymerase Stalling during Replication, Transcription and Translation. Current Biology. 2014 May;24(10):R445-52.

[337] Noe Gonzalez M, Blears D, Svejstrup JQ. Causes and Consequences of RNA Polymerase II Stalling during Transcript Elongation. Nature Reviews Molecular Cell Biology. 2020 Nov;22(1):1-19.

[338] van den Heuvel D, van der Weegen Y, Boer DEC, Ogi T, Luijsterburg MS. Transcription-Coupled DNA Repair: From Mechanism to Human Disorder. Trends in Cell Biology. 2021 May;31(5):359-71.

[339] Lans H, Hoeijmakers JHJ, Vermeulen W, Marteijn JA. The DNA Damage Response to Transcription Stress. Nature Reviews Molecular Cell Biology. 2019 Sep;20(12):1-19.

[340] Karakasili E, Burkert-Kautzsch C, Kieser A, Sträßer K. Degradation of DNA Damage-Independently Stalled RNA Polymerase II Is Independent of the E3 Ligase Elc1. Nucleic Acids Research. 2015 Feb;43(4):2486.

[341] Gillette TG, Gonzalez F, Delahodde A, Johnston SA, Kodadek T. Physical and Functional Association of RNA Polymerase II and the Proteasome. Proceedings of the National Academy of Sciences. 2004 Apr;101(16):5904-9.

[342] Svejstrup JQ. Rescue of Arrested RNA Polymerase II Complexes. Journal of cell science. 2003;116(Pt 3):447-51.

[343] Caron P, Pankotai T, Wiegant WW, Tollenaere MAX, Furst A, Bonhomme C, et al. WWP2 Ubiquitylates RNA Polymerase II for DNA-PK-dependent Transcription Arrest and Repair at DNA Breaks. Genes & Development. 2019 May.

[344] Iconomou M, Saunders DN. Systematic Approaches to Identify E3 Ligase Substrates. Biochemical Journal. 2016;473(22):4083-101.

[345] Stewart MD, Ritterhoff T, Klevit RE, Brzovic PS. E2 Enzymes: More than Just Middle Men. Cell Research. 2016 Apr;26(4):423-40.

[346] Uchida C, Kitagawa M. RING-, HECT-, and RBR-type E3 Ubiquitin Ligases: Involvement in Human Cancer. Current Cancer Drug Targets. 2016;16(2):157-74.

[347] Chew EH, Hagen T. Substrate-Mediated Regulation of Cullin Neddylation. Journal of Biological Chemistry. 2007 Jun;282(23):17032-40.

[348] Andérica-Romero AC, González-Herrera IG, Santamaría A, Pedraza-Chaverri J. Cullin 3 as a Novel Target in Diverse Pathologies. Redox Biology. 2013 Jul;1:366-72.

[349] Berthon A, Faucz F, Bertherat J, Stratakis CA. Analysis of ARMC5 Expression in Human Tissues. Molecular and Cellular Endocrinology. 2017 Feb;441:140-5.

[350] De Venanzi A, Alencar GA, Bourdeau I, Fragoso MCBV, Lacroix A. Primary Bilateral Macronodular Adrenal Hyperplasia. Current Opinion in Endocrinology, Diabetes and Obesity. 2014 Jun;21(3):177-84.

[351] Kyo C, Usui T, Kosugi R, Torii M, Yonemoto T, Ogawa T, et al. ARMC5 Alterations in Primary Macronodular Adrenal Hyperplasia (PMAH) and the Clinical State of Variant Carriers. Journal of the Endocrine Society. 2019 Oct;3(10):1837-46.

[352] Benderdour M, Martel-Pelletier J, Pelletier JPP, Kapoor M, Zunzunegui MVV, Fahmi H. Cellular Aging, Senescence and Autophagy Processes in Osteoarthritis. Current aging science. 2015;8(2):147-57.

[353] Bourdeau I, Antonini SR, Lacroix A, Kirschner LS, Matyakhina L, Lorang D, et al. Gene Array Analysis of Macronodular Adrenal Hyperplasia Confirms Clinical Heterogeneity and Identifies Several Candidate Genes as Molecular Mediators. Oncogene. 2004 Feb;23(8):1575-85.

[354] Antonini SR, Baldacchino V, Tremblay J, Hamet P, Lacroix A. Expression of ACTH Receptor Pathway Genes in Glucose-Dependent Insulinotrophic Peptide (GIP)-Dependent Cushing's Syndrome. Clin Endocrinol (Oxf). 2006 Jan;64(1):29-36.

[355] Cavalcante IP, Clauser E, Vaczlavik A, Drougat L, Lotfi C, Fragoso M, et al. Cullin 3 Is a Partner of Armadillo Repeat Containing 5 (ARMC5), the Product of the Gene Responsible for Primary Bilateral Macronodular Adrenal Hyperplasia. Annales d'Endocrinologie. 2018 Jun;79(3):184-5.

[356] Figueroa P, Gusmaroli G, Serino G, Habashi J, Ma L, Shen Y, et al. Arabidopsis Has Two Redundant Cullin3 Proteins That Are Essential for Embryo Development and That Interact with RBX1 and BTB Proteins to Form Multisubunit E3 Ubiquitin Ligase Complexes in Vivo. The Plant Cell. 2005 Apr;17(4):1180-95.

[357] Anandapadamanaban M, Kyriakidis NC, Csizmók V, Wallenhammar A, Espinosa AC, Ahlner A, et al. E3 Ubiquitin-Protein Ligase TRIM21-mediated Lysine Capture by UBE2E1 Reveals Substrate-Targeting Mode of a Ubiquitin-Conjugating E2. Journal of Biological Chemistry. 2019 Jul;294(30):11404-19.

[358] Jishage M, Yu X, Shi Y, Ganesan SJ, Chen WY, Sali A, et al. Architecture of Pol II(G) and Molecular Mechanism of Transcription Regulation by Gdown1. Nature Structural & Molecular Biology. 2018 Sep;25(9):859-67.

[359] Plechanovová A, Jaffray EG, Tatham MH, Naismith JH, Hay RT. Structure of a RING E3 Ligase and Ubiquitin-Loaded E2 Primed for Catalysis. Nature. 2012 Sep;489(7414):115-20.

[360] Scott DC, Sviderskiy VO, Monda JK, Lydeard JR, Cho SE, Harper JW, et al. Structure of a RING E3 Trapped in Action Reveals Ligation Mechanism for the Ubiquitin-like Protein NEDD8. Cell. 2014 Jun;157(7):1671-84.

[361] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly Accurate Protein Structure Prediction with AlphaFold. Nature. 2021 Jul:1-7.

[362] Egloff S, Murphy S. Cracking the RNA Polymerase II CTD Code. Trends in Genetics. 2008 Jun;24(6):280-8.

[363] Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, et al. Ring1-Mediated Ubiquitination of H2A Restrains Poised RNA Polymerase II at Bivalent Genes in Mouse ES Cells. Nature Cell Biology. 2007 Dec;9(12):1428-35.

[364] Gagliardi L, Hotu C, Casey G, Braund WJ, Ling KH, Dodd T, et al. Familial Vasopressin-Sensitive ACTH-independent Macronodular Adrenal Hyperplasia (VPs-AIMAH): Clinical Studies of Three Kindreds. Clinical Endocrinology. 2009 Jun;70(6):883-91.

[365] Lacroix A, Tremblay J, Rousseau G, Bouvier M, Hamet P. Propranolol Therapy for Ectopic β-Adrenergic Receptors in Adrenal Cushing's Syndrome. New England Journal of Medicine. 1997 Nov;337(20):1429-34.

[366] Cavalcante IP, Nishi M, Zerbini MCN, Almeida MQ, Brondani VB, Botelho M, et al. The Role of ARMC5 in Human Cell Cultures from Nodules of Primary Macronodular Adrenocortical Hyperplasia (PMAH). Mol Cell Endocrinol. 2018 Jan;460:36-46.

[367] Walker JJ, Spiga F, Gupta R, Zhao Z, Lightman SL, Terry JR. Rapid Intra-Adrenal Feedback Regulation of Glucocorticoid Synthesis. Journal of The Royal Society Interface. 2014 Nov;12(102):20140875-5.

[368] Somesh BP, Reid J, Liu WFF, Søgaard TMM, Erdjument-Bromage H, Tempst P, et al. Multiple Mechanisms Confining RNA Polymerase II Ubiquitylation to Polymerases Undergoing Transcriptional Arrest. Cell. 2005;121(6):913-23.

[369] Somesh BP, Sigurdsson S, Saeki H, Erdjument-Bromage H, Tempst P, Svejstrup JQ. Communication between Distant Sites in RNA Polymerase II through Ubiquitylation Factors and the Polymerase CTD. Cell. 2007;129(1):57-68.

[370] Phatnani HP, Greenleaf AL. Phosphorylation and Functions of the RNA Polymerase II CTD. Genes & Development. 2006 Nov;20(21):2922-36.

[371] Lee DF, Kuo HP, Liu M, Chou CK, Xia W, Du Y, et al. KEAP1 E3 Ligase-Mediated Downregulation of NF-κB Signaling by Targeting IKKβ. Molecular Cell. 2009 Oct;36(1):131-40.

[372] Petroski MD, Deshaies RJ. Function and Regulation of Cullin-RING Ubiquitin Ligases. Nat Rev Mol Cell Biol. 2005 Jan;6(1):9-20.

[373] Bosu DR, Kipreos ET. Cullin-RING Ubiquitin Ligases: Global Regulation and Activation Cycles. Cell Division. 2008 Feb;3:7.

[374] Cavalcante IP, Vaczlavik A, Drougat L, Lotfi CFP, Hecale-Perlemoine K, Ribes C, et al. Cullin 3 Targets the Tumor Suppressor Gene ARMC5 for Ubiquitination and Degradation. Endocrine-Related Cancer. 2020 Feb.

[375] Nalepa G, Rolfe M, Harper JW. Drug Discovery in the Ubiquitin-Proteasome System. Nature Reviews Drug Discovery. 2006 Jul;5(7):596-613.

[376] Rothenberg EV. Transcriptional Control of Early t and b Cell Developmental Choices. Annual Review of Immunology. 2014;32(1):283-321.

[377] Gala HP, Saha D, Venugopal N, Aloysius A, Dhawan J. RNA Polymerase II Pausing Regulates a Quiescence-Dependent Transcriptional Program, Priming Cells for Cell Cycle Reentry. bioRxiv. 2018 Jan:250910.

[378] Selby CP, Sancar A. Cockayne Syndrome Group B Protein Enhances Elongation by RNA Polymerase II. Proceedings of the National Academy of Sciences of the United States of America. 1997 Oct;94(21):11205-9.

[379] Xu J, Lahiri I, Wang W, Wier A, Cianfrocco MA, Chong J, et al. Structural Basis for the Initiation of Eukaryotic Transcription-Coupled DNA Repair. Nature. 2017 Nov;551(7682):653-7.

[380] Lommel L, Bucheli ME, Sweder KS. Transcription-Coupled Repair in Yeast Is Independent from Ubiquitylation of RNA Pol II: Implications for Cockayne's Syndrome. Proceedings of the National Academy of Sciences. 2000 Aug;97(16):9088-92.

[381] Woudstra EC, Gilbert C, Fellows J, Jansen L, Brouwer J, Erdjument-Bromage H, et al. A Rad26-Def1 Complex Coordinates Repair and RNA Pol II Proteolysis in Response to DNA Damage. Nature. 2002;415(6874):929-33.

[382] Ngoc LV, Wang YL, Kassavetis GA, Kadonaga JT. The Punctilious RNA Polymerase II Core Promoter. Genes & Development. 2017 Jul;31(13):1289-301.

[383] Ostrom Q, Chen Y, Devine K, Wolinsky Y, Davitkov P, Robbins S, et al. Family History of Cancer in Benign Brain Tumor Subtypes Versus Gliomas. Frontiers in Oncology. 2012;2.

[384] Yu JS, Koujak S, Nagase S, Li CM, Su T, Wang X, et al. PCDH8 , the Human Homolog of PAPC , Is a Candidate Tumor Suppressor of Breast Cancer. Oncogene. 2008 Aug;27(34):4657-65.

[385] Sun H, You Y, Guo M, Wang X, Zhang Y, Ye S. Tfcp2l1 Safeguards the Maintenance of Human Embryonic Stem Cell Self-Renewal. Journal of Cellular Physiology. 2018;233(9):6944-51.

[386] Smallridge RC, Marlow LA, Copland JA. Anaplastic Thyroid Cancer: Molecular Pathogenesis and Emerging Therapies. Endocrine-related cancer. 2009 Mar;16(1):17-44.

[387] Nishizawa M, Kataoka K, Vogt PK. MafA Has Strong Cell Transforming Ability but Is a Weak Transactivator. Oncogene. 2003 Sep;22(39):7882-90.

[388] Pouponnot C, Sii-Felice K, Hmitou I, Rocques N, Lecoin L, Druillennec S, et al. Cell Context Reveals a Dual Role for Maf in Oncogenesis. Oncogene. 2006 Mar;25(9):1299-310.

[389] Dikstein R, Zhou S, Tjian R. Human TAFII105 Is a Cell Type–Specific TFIID Subunit Related to hTAFII130. Cell. 1996 Oct;87(1):137-46.

[390] Louder RK, He Y, López-Blanco JR, Fang J, Chacón P, Nogales E. Structure of Promoter-Bound TFIID and Model of Human Pre-Initiation Complex Assembly. Nature. 2016 Mar;531(7596):604-9.

[391] Yamit-Hezi A, Dikstein R. TAFII105 Mediates Activation of Anti-Apoptotic Genes by NF-kappaB. Embo j. 1998 Sep;17(17):5161-9.

[392] Gagliardi L, Ling KH, Kok CH, Carolan J, Brautigan P, Kenyon R, et al. Genome-Wide Gene Expression Profiling Identifies Overlap with Malignant Adrenocortical Tumours and Novel Mechanisms of Inefficient Steroidogenesis in Familial ACTH-independent Macronodular Adrenal Hyperplasia. Endocrine-Related Cancer. 2012 Jun;19(3):L19-23.

[393] Kallen CB, Billheimer JT, Summers SA, Stayrook SE, Lewis M, Strauss JF. Steroidogenic Acute Regulatory Protein (StAR) Is A Sterol Transfer Protein*. Journal of Biological Chemistry. 1998 Oct;273(41):26285-8.

[394] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera–a Visualization System for Exploratory Research and Analysis. Journal of Computational Chemistry. 2004 Oct;25(13):1605-12.

[395] Bolger AM, Lohse M, Usadel B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. Bioinformatics (Oxford, England). 2014 Aug;30(15):2114-20.

[396] Bourgey M, Dali R, Eveleigh R, Chen KC, Letourneau L, Fillon J, et al. GenPipes: An Open-Source Framework for Distributed and Scalable Genomic Analyses. GigaScience. 2019 Jun;8(6).

[397] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast Universal RNA-seq Aligner. Bioinformatics. 2013 Jan;29(1):15-21.

[398] Allen M, Poggiali D, Whitaker K, Marshall TR, Kievit RA. Raincloud Plots: A Multi-Platform Tool for Robust Data Visualization. Wellcome Open Res. 2019;4:63.

[399] Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize Implements and Enhances Circular Visualization in R. Bioinformatics. 2014 Oct;30(19):2811-2.

[400] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003 Nov;13(11):2498-504.

[401] Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: A Cytoscape Plug-in to Decipher Functionally Grouped Gene Ontology and Pathway Annotation Networks. Bioinformatics (Oxford, England). 2009 Apr;25(8):1091-3.

[402] Cotney JL, Noonan JP. Chromatin Immunoprecipitation with Fixed Animal Tissues and Preparation for High-Throughput Sequencing. Cold Spring Harbor Protocols. 2015 Feb;2015(2):pdb.prot084848.

[403] Li H, Durbin R. Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. Bioinformatics. 2009;25(14):1754-60.

[404] Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-Based Analysis of ChIP-Seq (MACS). Genome Biology. 2008 Sep;9(9):R137.

[405] Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities. Molecular Cell. 2010 May;38(4):576-89.

[406] Robinson MD, Oshlack A. A Scaling Normalization Method for Differential Expression Analysis of RNA-seq Data. Genome Biology. 2010 Mar;11(3):R25.

[407] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma Powers Differential Expression Analyses for RNA-sequencing and Microarray Studies. Nucleic Acids Research. 2015 Apr;43(7):e47.

[408] Shen L, Shao N, Liu X, Nestler E. Ngs.Plot: Quick Mining and Visualization of next-Generation Sequencing Data by Integrating Genomic Databases. BMC Genomics. 2014 Apr;15(1):284.

[409] Love MI, Huber W, Anders S. Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2. Genome Biology. 2014 Dec;15(12):550.

[410] Timeline Human Development - Embryology; 2016. `https://embryology.med.unsw.edu.au/embryology/index.php/Timeline_human_development`. Website.

[411] Greene ND, Copp AJ. Development of the Vertebrate Central Nervous System: Formation of the Neural Tube. Prenat Diagn. 2009 Apr;29(4):303-11.

[412] Mitchell LE, Adzick NS, Melchionne J, Pasquariello PS, Sutton LN, Whitehead AS. Spina Bifida. The Lancet. 2004 Nov;364(9448):1885-95.

[413] Blencowe H, Kancherla V, Moorthie S, Darlison MW, Modell B. Estimates of Global and Regional Prevalence of Neural Tube Defects for 2015: A Systematic Analysis. Ann N Y Acad Sci. 2018 Feb;1414(1):31-46.

[414] Davidson LA, Keller RE. Neural Tube Closure in Xenopus Laevis Involves Medial Migration, Directed Protrusive Activity, Cell Intercalation and Convergent Extension. Development. 1999 Oct;126(20):4547-56.

[415] Kobus-Bianchini K, Bourckhardt GF, Ammar D, Nazari EM, Müller YMR. Homocysteine-Induced Changes in Cell Proliferation and Differentiation in the Chick Embryo Spinal Cord: Implications for Mechanisms of Neural Tube Defects (NTD). Reprod Toxicol. 2017 Apr;69:167-73.

[416] Suzuki M, Morita H, Ueno N. Molecular Mechanisms of Cell Shape Changes That Contribute to Vertebrate Neural Tube Closure. Dev Growth Differ. 2012 Apr;54(3):266-76.

[417] Yang SL, Yang M, Herrlinger S, Liang C, Lai F, Chen JF. MiR-302/367 Regulate Neural Progenitor Proliferation, Differentiation Timing, and Survival in Neurulation. Dev Biol. 2015 Dec;408(1):140-50.

[418] Agopian AJ, Tinker SC, Lupo PJ, Canfield MA, Mitchell LE, Study NBDP. Proportion of Neural Tube Defects Attributable to Known Risk Factors. Birth Defects Res A Clin Mol Teratol. 2013 Jan;97(1):42-6.

[419] Imbard A, Benoist JF, Blom HJ. Neural Tube Defects, Folic Acid and Methylation. Int J Environ Res Public Health. 2013 Sep;10(9):4352-89.

[420] Greene ND, Stanier P, Copp AJ. Genetics of Human Neural Tube Defects. Hum Mol Genet. 2009 Oct;18(R2):R113-29.

[421] Yadav U, Kumar P, Yadav SK, Mishra OP, Rai V. "Polymorphisms in Folate Metabolism Genes as Maternal Risk Factor for Neural Tube Defects: An Updated Meta-Analysis". Metab Brain Dis. 2015 Feb;30(1):7-24.

[422] Zhu Q, Li L, Wang T, Jiang W, Ding J, Liu M, et al. Genetic Study of 12 SNPs Involved in 11 Folate Metabolism Genes and Neural Tube Defects in Suzhou Children. Molecular and Genetic Medicine. 2016;10(2):1-7.

[423] Sims RJ, Mandal SS, Reinberg D. Recent Highlights of RNA-polymerase-II-mediated Transcription. Current Opinion in Cell Biology. 2004 Jun;16(3):263-71.

[424] Statello L, Guo CJ, Chen LL, Huarte M. Gene Regulation by Long Non-Coding RNAs and Its Biological Functions. Nature Reviews Molecular Cell Biology. 2021 Feb;22(2):96-118.

[425] Woychik NA. Fractions to Functions: RNA Polymerase II Thirty Years Later. Cold Spring Harb Symp Quant Biol. 1998;63:311-7.

[426] Kurihara LJ, Semenova E, Levorse JM, Tilghman SM. Expression and Functional Analysis of Uch-L3 during Mouse Development. Mol Cell Biol. 2000 Apr;20(7):2498-504.

[427] Li Y, Liu Y, Xu H, Jiang G, der Jeught KV, Fang Y, et al. Heterozygous Deletion of Chromosome 17p Renders Prostate Cancer Vulnerable to Inhibition of RNA Polymerase II. Nature Communications. 2018 Oct;9(1):4394.

[428] Gao Q, Gao YM. Hyperglycemic Condition Disturbs the Proliferation and Cell Death of Neural Progenitors in Mouse Embryonic Spinal Cord. Int J Dev Neurosci. 2007 Oct;25(6):349-57.

[429] Zhu X, Li B, Ai Z, Xiang Z, Zhang K, Qiu X, et al. A Robust Single Primate Neuroepithelial Cell Clonal Expansion System for Neural Tube Development and Disease Studies. Stem Cell Reports. 2016 Feb;6(2):228-42.

[430] Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the Deleteriousness of Variants throughout the Human Genome. Nucleic Acids Res. 2019 Jan;47(D1):D886-d894.

[431] Chen RH. Cullin 3 and Its Role in Tumorigenesis. Adv Exp Med Biol. 2020;1217:187-210.

[432] Galbraith MD, Bender H, Espinosa JM. Therapeutic Targeting of Transcriptional Cyclin-Dependent Kinases. Transcription. 2019 Apr;10(2):118-36.

[433] Korolchuk VI, Menzies FM, Rubinsztein DC. Mechanisms of Cross-Talk between the Ubiquitin-Proteasome and Autophagy-Lysosome Systems. FEBS Lett. 2010 Apr;584(7):1393-8.

[434] Yau R, Rape M. The Increasing Complexity of the Ubiquitin Code. Nature Cell Biology. 2016 Jun;18(6):579-86.

[435] Marz M, Donath A, Verstraete N, Nguyen VT, Stadler PF, Bensaude O. Evolution of 7SK RNA and Its Protein Partners in Metazoa. Mol Biol Evol. 2009 Dec;26(12):2821-30.

[436] Castelo-Branco G, Amaral PP, Engström PG, Robson SC, Marques SC, Bertone P, et al. The Non-Coding snRNA 7SK Controls Transcriptional Termination, Poising, and Bidirectionality in Embryonic Stem Cells. Genome Biol. 2013;14(9):R98.

[437] Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Research. 2020 Apr;9:304.

[438] Sun XM, Bowman A, Priestman M, Bertaux F, Martinez-Segura A, Tang W, et al. Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size. Curr Biol. 2020 Apr;30(7):1217-30.e7.

[439] Kimura H, Tao Y, Roeder RG, Cook PR. Quantitation of RNA Polymerase II and Its Transcription Factors in an HeLa Cell: Little Soluble Holoenzyme but Significant Amounts of Polymerases Attached to the Nuclear Substructure. Mol Cell Biol. 1999;19(8):5383-92.

[440] Devlin AM, Ling EH, Peerson JM, Fernando S, Clarke R, Smith AD, et al. Glutamate Carboxypeptidase II: A Polymorphism Associated with Lower Levels of Serum Folate and Hyperhomocysteinemia.

Hum Mol Genet. 2000 Nov;9(19):2837-44.

[441] Database GHG. FOLH1 Gene - GeneCards; 2021. `https://www.genecards.org/cgi-bin/carddisp.pl?gene=FOLH1`. Website.

[442] Ohta T, Michel JJ, Schottelius AJ, Xiong Y. ROC1, a Homolog of APC11, Represents a Family of Cullin Partners with an Associated Ubiquitin Ligase Activity. Molecular Cell. 1999/04/01/;3(4):535-41.

[443] de Bie P, Ciechanover A. Ubiquitination of E3 Ligases: Self-Regulation of the Ubiquitin System via Proteolytic and Non-Proteolytic Mechanisms. Cell Death Differ. 2011 Sep;18(9):1393-402.

[444] Nguyen AT, Prado MA, Schmidt PJ, Sendamarai AK, Wilson-Grady JT, Min M, et al. UBE2O Remodels the Proteome during Terminal Erythroid Differentiation. Science. 2017 Aug;357(6350):eaan0218.

[445] Lao L, Bourdeau I, Gagliardi L, He X, Shi W, Hao B, et al. ARMC5 Is Part of an RPB1-specific Ubiquitin Ligase Implicated in Adrenal Gland Hyperplasia. Nucleic Acids Research. 2022;(Under Revision).

[446] Steurer B, Marteijn JA. Traveling Rocky Roads: The Consequences of Transcription-Blocking DNA Lesions on RNA Polymerase II. Journal of Molecular Biology. 2017 Oct;429(21):3146-55.

[447] Wilson MD, Harreman M, Svejstrup JQ. Ubiquitylation and Degradation of Elongating RNA Polymerase II: The Last Resort. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2013 Jan;1829(1):151-7.

[448] Liptak GS, Dosa NP. Myelomeningocele. Pediatr Rev. 2010 Nov;31(11):443-50.

[449] Eubanks JD, Cheruvu VK. Prevalence of Sacral Spina Bifida Occulta and Its Relationship to Age, Sex, Race, and the Sacral Table Angle: An Anatomic, Osteologic Study of Three Thousand One Hundred Specimens. Spine (Phila Pa 1976). 2009 Jul;34(15):1539-43.

[450] Karikkineth AC, Scheibye-Knudsen M, Fivenson E, Croteau DL, Bohr VA. Cockayne Syndrome: Clinical Features, Model Systems and Pathways. Ageing research reviews. 2017 Jan;33:3-17.

[451] Gartel AL, Radhakrishnan SK. Lost in Transcription: P21 Repression, Mechanisms, and Consequences. Cancer Res. 2005 May;65(10):3980-5.

[452] Podust VN, Podust LM, Goubin F, Ducommun B, Hübscher U. Mechanism of Inhibition of Proliferating Cell Nuclear Antigen-Dependent DNA Synthesis by the Cyclin-Dependent Kinase Inhibitor P21. Biochemistry. 1995 Jul;34(27):8869-75.

[453] Spencer SL, Cappell SD, Tsai FC, Overton KW, Wang CL, Meyer T. The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit. Cell. 2013 Oct;155(2):369-83.

[454] Vairapandi M, Azam N, Balliet AG, Hoffman B, Liebermann DA. Characterization of MyD118, Gadd45, and Proliferating Cell Nuclear Antigen (PCNA) Interacting Domains. PCNA Impedes

MyD118 AND Gadd45-mediated Negative Growth Control. J Biol Chem. 2000 Jun;275(22):16810-9.

[455] Vairapandi M, Balliet AG, Hoffman B, Liebermann DA. GADD45b and GADD45g Are Cdc2/cyclinB1 Kinase Inhibitors with a Role in S and G2/M Cell Cycle Checkpoints Induced by Genotoxic Stress. J Cell Physiol. 2002 Sep;192(3):327-38.

[456] Qiu W, David D, Zhou B, Chu PG, Zhang B, Wu M, et al. Down-Regulation of Growth Arrest DNA Damage-Inducible Gene 45beta Expression Is Associated with Human Hepatocellular Carcinoma. The American Journal of Pathology. 2003/06//;162(6):1961-74.

[457] Chandler CJ, Wang TT, Halsted CH. Pteroylpolyglutamate Hydrolase from Human Jejunal Brush Borders. Purification and Characterization. J Biol Chem. 1986 Jan;261(2):928-33.

[458] Han L, Picker JD, Schaevitz LR, Tsai G, Feng J, Jiang Z, et al. Phenotypic Characterization of Mice Heterozygous for a Null Mutation of Glutamate Carboxypeptidase II. Synapse. 2009 Aug;63(8):625-35.

[459] Tsai G, Dunham KS, Drager U, Grier A, Anderson C, Collura J, et al. Early Embryonic Death of Glutamate Carboxypeptidase II (NAALADase) Homozygous Mutants. Synapse. 2003 Dec;50(4):285-92.

[460] Guo J, Xie H, Wang J, Zhao H, Wang F, Liu C, et al. The Maternal Folate Hydrolase Gene Polymorphism Is Associated with Neural Tube Defects in a High-Risk Chinese Population. Genes Nutr. 2013 Mar;8(2):191-7.

[461] Singh B, Gupta R, Rai S, Rai S, Singh R. In-Silico Study of Neural Tube Defect in Relation to FOHL1 Gene. International Journal of Scientific Research. 2012 Jun;3:366-7.

[462] Thiffault I, Wolf NI, Forget D, Guerrero K, Tran LT, Choquet K, et al. Recessive Mutations in POLR1C Cause a Leukodystrophy by Impairing Biogenesis of RNA Polymerase III. Nat Commun. 2015 Jul;6:7623.

[463] Liu G, Zhang J, Larsen B, Stark C, Breitkreutz A, Lin ZY, et al. ProHits: Integrated Software for Mass Spectrometry-Based Interaction Proteomics. Nature Biotechnology. 2010 Oct;28:1015-7.

[464] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. Nat Methods. 2016 Sep;13(9):731-40.

[465] Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289-300.

[466] Roberts TC, Hart JR, Kaikkonen MU, Weinberg MS, Vogt PK, Morris KV. Quantification of Nascent Transcription by Bromouridine Immunocapture Nuclear Run-on RT-qPCR. Nature Protocols. 2015 Aug;10(8):1198-211.

[467] Hillman P, Baker C, Hebert L, Brown M, Hixson J, Ashley-Koch A, et al. Identification of Novel Candidate Risk Genes for Myelomeningocele within the Glucose Homeostasis/Oxidative Stress and Folate/One-Carbon Metabolism Networks. Molecular genetics & genomic medicine. 2020;8(11):e1495-5.

[468] Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. Hum Mutat. 2016 Mar;37(3):235-41.

[469] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. Nature. 2020 May;581(7809):434-43.

[470] Myers TA, Chanock SJ, Machiela MJ. LDlinkR: An R Package for Rapidly Calculating Linkage Disequilibrium Statistics in Diverse Populations. Front Genet. 2020;11:157.

[471] Zhang S, Übelmesser N, Josipovic N, Forte G, Slotman JA, Chiang M, et al. RNA Polymerase II Is Required for Spatial Chromatin Reorganization Following Exit from Mitosis. bioRxiv. 2020 Oct:2020.10.27.356915.

[472] Akhrymuk I, Kulemzin SV, Frolova EI. Evasion of the Innate Immune Response: The Old World Alphavirus nsP2 Protein Induces Rapid Degradation of Rpb1, a Catalytic Subunit of RNA Polymerase II. Journal of Virology. 2012;86(13):12.

[473] Wang J, Zhao S, He W, Wei Y, Zhang Y, Pegg H, et al. A Transcription Factor IIA-binding Site Differentially Regulates RNA Polymerase II-mediated Transcription in a Promoter Context-Dependent Manner. Journal of Biological Chemistry. 2017 Jul;292(28):11873-85.

[474] London E, Bloyd M, Stratakis CA. PKA Functions in Metabolism and Resistance to Obesity: Lessons from Mouse and Human Studies. Journal of Endocrinology. 2020 Sep;246(3):R51-64.

[475] Mao CG, Jiang SS, Shen C, Long T, Jin H, Tan QY, et al. BCAR1 Promotes Proliferation and Cell Growth in Lung Adenocarcinoma via Upregulation of POLR2A. Thoracic Cancer. 2020 Nov;11(11):3326-36.

[476] Dickson JR, Yoon H, Frosch MP, Hyman BT. Cytoplasmic Mislocalization of RNA Polymerase II Subunit RPB1 in Alzheimer Disease Is Linked to Pathologic Tau. Journal of Neuropathology & Experimental Neurology. 2021 May;80(nlab040).

[477] Santi L, Maggioli C, Mastroroberto M, Tufoni M, Napoli L, Caraceni P. Acute Liver Failure Caused by Amanita Phalloides Poisoning. International Journal of Hepatology. 2012 Jul;2012:e487480.

[478] Hwang CS, Shemorry A, Varshavsky A. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. Science. 2010 Feb;327(5968):973-7.

[479] Park SE, Kim JM, Seok OH, Cho H, Wadas B, Kim SY, et al. Control of Mammalian G Protein Signaling by N-terminal Acetylation and the N-end Rule Pathway. Science. 2015 Mar;347(6227):1249-52.

[480] Chen SJ, Wu X, Wadas B, Oh JH, Varshavsky A. An N-end Rule Pathway That Recognizes Proline and Destroys Gluconeogenic Enzymes. Science. 2017 Jan;355(6323).

[481] Timms RT, Zhang Z, Rhee DY, Harper JW, Koren I, Elledge SJ. A Glycine-Specific N-degron Pathway Mediates the Quality Control of Protein N -Myristoylation. Science. 2019 Jul;365(6448):eaaw4912.

[482] Rusnac DV, Lin HC, Canzani D, Tien KX, Hinds TR, Tsue AF, et al. Recognition of the Diglycine C-End Degron by CRL2KLHDC2 Ubiquitin Ligase. Molecular Cell. 2018 Dec;72(5):813-22.e4.

[483] Lin HC, Yeh CW, Chen YF, Lee TT, Hsieh PY, Rusnac DV, et al. C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases. Molecular Cell. 2018 May;70(4):602-13.e3.

[484] Soond SM, Terry JL, Colbert JD, Riches DWH. TRUSS, a Novel Tumor Necrosis Factor Receptor 1 Scaffolding Protein That Mediates Activation of the Transcription Factor NF-κB. Molecular and Cellular Biology. 2003 Nov;23(22):8334-44.

[485] Sakamoto KM, Kim KB, Kumagai A, Mercurio F, Crews CM, Deshaies RJ. Protacs: Chimeric Molecules That Target Proteins to the Skp1-Cullin-F Box Complex for Ubiquitination and Degradation. Proceedings of the National Academy of Sciences. 2001 Jul;98(15):8554-9.

[486] Mullard A. Targeted Protein Degraders Crowd into the Clinic. Nature Reviews Drug Discovery. 2021 Mar;20(4):247-50.

[487] Lu G, Middleton RE, Sun H, Naniong M, Ott CJ, Mitsiades CS, et al. The Myeloma Drug Lenalidomide Promotes the Cereblon-Dependent Destruction of Ikaros Proteins. Science. 2014 Jan;343(6168):305-9.

[488] Krönke J, Udeshi ND, Narla A, Grauman P, Hurst SN, McConkey M, et al. Lenalidomide Causes Selective Degradation of IKZF1 and IKZF3 in Multiple Myeloma Cells. Science. 2014 Jan;343(6168):301-5.

[489] Brand M, Jiang B, Bauer S, Donovan KA, Liang Y, Wang ES, et al. Homolog-Selective Degradation as a Strategy to Probe the Function of CDK6 in AML. Cell Chemical Biology. 2019 Feb;26(2):300-6.e9.

[490] He Y, Khan S, Huo Z, Lv D, Zhang X, Liu X, et al. Proteolysis Targeting Chimeras (PROTACs) Are Emerging Therapeutics for Hematologic Malignancies. Journal of Hematology & Oncology. 2020 Jul;13(1):103.

[491] Olson CM, Jiang B, Erb MA, Liang Y, Doctor ZM, Zhang Z, et al. Pharmacological Perturbation of CDK9 Using Selective CDK9 Inhibition or Degradation. Nature Chemical Biology. 2018 Feb;14(2):163-70.

[492] Nalawansha DA, Crews CM. PROTACs: An Emerging Therapeutic Modality in Precision Medicine. Cell Chemical Biology. 2020 Aug;27(8):998-1014.

[493] Bondeson DP, Mares A, Smith IED, Ko E, Campos S, Miah AH, et al. Catalytic in Vivo Protein Knockdown by Small-Molecule PROTACs. Nature Chemical Biology. 2015 Aug;11(8):611-7.

[494] Crew AP, Raina K, Dong H, Qian Y, Wang J, Vigil D, et al. Identification and Characterization of Von Hippel-Lindau-Recruiting Proteolysis Targeting Chimeras (PROTACs) of TANK-Binding Kinase 1. Journal of Medicinal Chemistry. 2018 Jan;61(2):583-98.

[495] Cromm PM, Samarasinghe KTG, Hines J, Crews CM. Addressing Kinase-Independent Functions of Fak via PROTAC-Mediated Degradation. Journal of the American Chemical Society. 2018 Dec;140(49):17019-26.

[496] Gechijian LN, Buckley DL, Lawlor MA, Reyes JM, Paulk J, Ott CJ, et al. Functional TRIM24 Degrader via Conjugation of Ineffectual Bromodomain and VHL Ligands. Nature Chemical Biology. 2018 Apr;14(4):405-12.

[497] Farnaby W, Koegl M, Roy MJ, Whitworth C, Diers E, Trainor N, et al. BAF Complex Vulnerabilities in Cancer Demonstrated via Structure-Based PROTAC Design. Nature Chemical Biology. 2019 Jul;15(7):672-80.

[498] Schneekloth JS, Fonseca FN, Koldobskiy M, Mandal A, Deshaies R, Sakamoto K, et al. Chemical Genetic Control of Protein Levels: Selective in Vivo Targeted Degradation. Journal of the American Chemical Society. 2004 Mar;126(12):3748-54.

[499] Hines J, Lartigue S, Dong H, Qian Y, Crews CM. MDM2-Recruiting PROTAC Offers Superior, Synergistic Antiproliferative Activity via Simultaneous Degradation of BRD4 and Stabilization of P53. Cancer Research. 2019 Jan;79(1):251-62.

[500] Lu M, Liu T, Jiao Q, Ji J, Tao M, Liu Y, et al. Discovery of a Keap1-dependent Peptide PROTAC to Knockdown Tau by Ubiquitination-Proteasome Degradation Pathway. European Journal of Medicinal Chemistry. 2018 Feb;146:251-9.

[501] Jouanne M, Rault S, Voisin-Chiret AS. Tau Protein Aggregation in Alzheimer's Disease: An Attractive Target for the Development of Novel Therapeutic Agents. European Journal of Medicinal Chemistry. 2017 Oct;139:153-67.

[502] Cui HK, Zhao B, Li Y, Guo Y, Hu H, Liu L, et al. Design of Stapled α-Helical Peptides to Specifically Activate Wnt/β-Catenin Signaling. Cell Research. 2013 Apr;23(4):581-4.

[503] Clark VE, Harmancı AS, Bai H, Youngblood MW, Lee TI, Baranoski JF, et al. Recurrent Somatic Mutations in POLR2A Define a Distinct Subset of Meningiomas. Nature Genetics. 2016 Oct;48(10):1253-9.

[504] Hansen AW, Arora P, Khayat MM, Smith LJ, Lewis AM, Rossetti LZ, et al. Germline Mutation in POLR2A: A Heterogeneous, Multi-Systemic Developmental Disorder Characterized by Transcriptional Dysregulation. Human Genetics and Genomics Advances. 2021 Jan;2(1):100014.

[505] Haijes HA, Koster MJE, Rehmann H, Li D, Hakonarson H, Cappuccio G, et al. De Novo Heterozygous POLR2A Variants Cause a Neurodevelopmental Syndrome with Profound Infantile-Onset Hypotonia. The American Journal of Human Genetics. 2019 Aug;105(2):283-301.

[506] Povlsen LK, Beli P, Wagner SA, Poulsen SL, Sylvestersen KB, Poulsen JW, et al. Systems-Wide Analysis of Ubiquitylation Dynamics Reveals a Key Role for PAF15 Ubiquitylation in DNA-damage Bypass. Nature Cell Biology. 2012 Oct;14(10):1089-98.

[507] Elia AEH, Boardman AP, Wang DC, Huttlin EL, Everley RA, Dephoure N, et al. Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response. Molecular Cell. 2015 Sep;59(5):867-81.

[508] Boeing S, Williamson L, Encheva V, Gori I, Saunders RE, Instrell R, et al. Multiomic Analysis of the UV-Induced DNA Damage Response. Cell Reports. 2016 May;15(7):1597-610.

[509] Graumann J, Hubner NC, Kim JB, Ko K, Moser M, Kumar C, et al. Stable Isotope Labeling by Amino Acids in Cell Culture (SILAC) and Proteome Quantitation of Mouse Embryonic Stem Cells to a Depth of 5,111 Proteins. Molecular & Cellular Proteomics. 2008 Apr;7(4):672-83.

[510] Shiber A, Ravid T. Chaperoning Proteins for Destruction: Diverse Roles of Hsp70 Chaperones and Their Co-Chaperones in Targeting Misfolded Proteins to the Proteasome. Biomolecules. 2014 Jul;4(3):704-24.

[511] Kundrat L, Regan L. Identification of Residues on Hsp70 and Hsp90 Ubiquitinated by the Cochaperone CHIP. Journal of Molecular Biology. 2010 Jan;395(3):587-94.

[512] Moore DJ, West AB, Dikeman DA, Dawson VL, Dawson TM. Parkin Mediates the Degradation-Independent Ubiquitination of Hsp70. Journal of Neurochemistry. 2008;105(5):1806-19.

[513] Chen D, Kon N, Li M, Zhang W, Qin J, Gu W. ARF-BP1/Mule Is a Critical Mediator of the ARF Tumor Suppressor. Cell. 2005 Jul;121(7):1071-83.

[514] Adhikary S, Marinoni F, Hock A, Hulleman E, Popov N, Beier R, et al. The Ubiquitin Ligase HectH9 Regulates Transcriptional Activation by Myc and Is Essential for Tumor Cell Proliferation. Cell. 2005 Nov;123(3):409-21.

[515] Li M, Xu X, Chang CW, Liu Y. TRIM28 Functions as the SUMO E3 Ligase for PCNA in Prevention of Transcription Induced DNA Breaks. Proceedings of the National Academy of Sciences. 2020 Sep;117(38):23588-96.

[516] Neo SH, Itahana Y, Alagu J, Kitagawa M, Guo AK, Lee SH, et al. TRIM28 Is an E3 Ligase for ARF-Mediated NPM1/B23 SUMOylation That Represses Centrosome Amplification. Molecular and Cellular Biology. 2015 Aug;35(16):2851-63.

[517] Czerwińska P, Mazurek S, Wiznerowicz M. The Complexity of TRIM28 Contribution to Cancer. Journal of Biomedical Science. 2017 Aug;24(1):63.

[518] Dove KK, Kemp HA, Di Bona KR, Reiter KH, Milburn LJ, Camacho D, et al. Two Functionally Distinct E2/E3 Pairs Coordinate Sequential Ubiquitination of a Common Substrate in Caenorhabditis Elegans Development. Proceedings of the National Academy of Sciences. 2017 Aug;114(32):E6576-84.