

**ZCCHC8 IS REQUIRED FOR THE DEGRADATION OF PERVASIVE TRANSCRIPTS  
ORIGINATING FROM MULTIPLE GENOMIC REGULATORY REGIONS**

by  
Joshua Wade Collins

A dissertation submitted to Johns Hopkins University in conformity with the  
requirements for the degree of Doctor of Philosophy

Baltimore, Maryland  
April 2021

© 2021 Joshua Collins  
All rights reserved

# Abstract

Antibodies are among the most frequently used tools in biomedical research, but they are often subject to many pitfalls including cross-reactivity, production lot-to-lot variability, and loss of activity. Despite the multiple laboratory applications for which antibodies are used, there are few standardized scientific guidelines for validation of antibody usage. Much of the blame for the recent reproducibility crisis in biomedical research has been placed on the failure of antibodies to perform consistently over time. In chapter one, we report a case study of an anti-BTBD7 antibody that displays cross-reactivity, extreme lot-to-lot variability, and loss of activity across multiple applications. We determined the major cross-reacting protein of interest to be ZCCHC8 and provide a general framework for determining the identity of cross-reacting proteins. We also surveyed a panel of anti-ZCCHC8 antibodies that show an array of non-specificity and in some instances apparent cross-reactivity with BTBD7.

A mechanistic study of ZCCHC8 and its role in regulating non-coding RNA is detailed in chapter two. The vast majority of mammalian genomes are transcribed as non-coding RNA in what is referred to as “pervasive transcription.” Recent studies have uncovered various families of non-coding RNA transcribed upstream of transcription start sites. In particular, highly unstable promoter upstream transcripts known as PROMPTs have been shown to be targeted for exosomal degradation by the nuclear exosome targeting complex (NEXT) consisting of the RNA helicase MTR4, the zinc-knuckle scaffold ZCCHC8, and the RNA binding protein RBM7. Here, we report that in addition to its known RNA substrates, ZCCHC8 and/or the NEXT complex are responsible for the targeted degradation of pervasive transcripts produced at CTCF

binding sites, open chromatin regions, promoters, promoter flanking regions, and transcription factor binding sites. Additionally, we report that a significant number of RIKEN cDNAs and predicted genes display the hallmarks of PROMPTs and are also substrates for ZCCHC8 and/or NEXT complex regulation suggesting these are unlikely to be functional genes. Our results suggest that ZCCHC8 and/or the NEXT complex may play a larger role in the global regulation of pervasive transcription than previously reported.

Primary Reader and Advisor: Kenneth M. Yamada, PhD

Secondary Reader: Trina Schroer, PhD

# Acknowledgements

I have to admit that this section was, in some ways, the most difficult for me to write—chiefly because of the raw emotion that it conjures. The process of obtaining this PhD did not, at all, go according to plan. Not at all, not at all. While the research was at times difficult and even daunting, and the setbacks discouraging and maddening, it was the simultaneous pummeling that life delivered that made finishing all but impossible were it not for my dearest Lee continually picking me up from the rocky bottom. Lee, my greatest love, I owe you my life, my sanity (though I am sure you would argue there is little left), and my eternal devotion. I love you impossibly.

I would like to thank my parents, Terry and Joyce, and my sisters (Kathy, Dottie, Sara, and Joey), for always loving and supporting me, and for providing comfort in dark times. Shout out to my nieces and nephews (Lilly—here is another book for you to read; now go write your own).

I have generally held the term mentor in reserve as I had not previously encountered anyone I thought particularly deserving of the honorific. But I am proud to call Ken Yamada mentor. And so, I offer my eternal gratitude to my mentor, Ken Yamada, for giving me a second start in science. I came to Ken's lab scarred and mentally exhausted from years of turmoil suffered under the regime of a toxic boss. Ken showed me what it is to be a good scientist and a caring, supportive mentor. Always positive, always attentive, always available, Ken continually helped me find ways to solve the unsolvable.

I would like to thank my committee members, Trina Schroer and Tom Misteli, for all their valuable input during committee meetings that helped shape a project from

some very messy data. Thank you to Tatjana Trcek for serving as a reader on my dissertation defense committee.

Thank you to all my colleagues in the Yamada lab for all the support and all the great conversations about life, culture, and science.

Thank you to Sharon Milgram and Phil Wang for love and support and for helping me get out of a bad lab and into a great one.

Thank you to Joan, a legend that, like Prince or Beyonce, is known to all at Hopkins by a single name only, for continual love and support and for solving so many problems.

Thank you to my dear friends Diane and Joanna for being there from start to finish in more ways, and labs, than one.

To my dear little Henry, I hope that you are there, somewhere, watching. I hope that I make you proud. I hope that we see each other again one day.

# Preface

This preface is intended to serve as an orientation, of sorts, for the general reader that could not have possibly been privy to the peculiar series of events that led to the work contained herein. In brief, I began my project studying the role of the extracellular matrix in branching morphogenesis and ended it studying non-coding RNA degradation in the nucleus. This wild transition was the result of a cross-reacting commercial antibody designed to detect BTBD7, a regulator of branching morphogenesis, that I was ultimately able to determine actually detects ZCCHC8, a key member of a complex that helps degrade non-coding RNA. I dare say that this was not an “ordinary” case of antibody non-specificity. The first chapter, then, is dedicated to the many trials and tribulations that occurred, and to which I devoted a significant amount of time, while studying “BTBD7”. This chapter has been styled as a case study of antibody cross-reactivity and the methods I used to uncover the identity of the cross-reacting target protein, ZCCHC8.

To be frank, the first chapter does not contain novel methods or findings—it is intended to serve as a cautionary tale of the impacts of the current reproducibility crisis in biomedical research and provide a blueprint for overcoming issues encountered when working with antibodies. Hopefully, it also demonstrates the many different areas of expertise I developed over this time.

The second chapter is dedicated to the signature research for which this dissertation is named. This chapter details novel research regarding the role of ZCCHC8 and the nuclear exosome targeting complex in degrading pervasive transcription of non-coding RNA.

## Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>Acknowledgments</b>	<b>iv</b>
<b>Preface</b>	<b>vi</b>
<b>Table of contents</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>x</b>
<b>Chapter 1 A Case Study of Antibody Cross-Reactivity</b>	<b>1</b>
<b>Introduction</b>	<b>2</b>
Background	6
Immunofluorescence Validation	8
Determination of the Anti-Btbd7 Cross-Reacting Protein	24
Epitope mapping of the immunogen amino acid sequence	24
Mass spectrometry	29
Immunoprecipitation for protein X	29
Two-dimensional electrophoresis	30
Antibody cross-reactive candidate validation	33
ZCCHC8 expression in salivary glands and cell nuclei	37
Non-Specificity in a Panel of Antibodies	40
Direct western blot and immunofluorescence comparison	40
Human Protein Atlas comparison	44
Sequence Alignment Between ZCCHC8 and BTBD7	47
<b>Discussion</b>	<b>50</b>
Antibody validation	50
Identification of cross-reacting proteins	52
ZCCHC8 and BTBD7 shared epitope	55
Variation in staining patterns of ZCCHC8	56
<b>Materials and Methods</b>	<b>58</b>
Cell culture and maintenance	58
CRISPR/Cas9 knockout of <i>Zcchc8</i> in SIMS cells	58
CRISPR/Cas9 knockout of <i>Btbd7</i> in SIMS cells	59
Antibodies	61
Western blots	62
Immunofluorescence	63
Salivary gland isolation and culture	64
Immunostaining of salivary glands	65
Cell spreading assay	66
Microscopy	67
Epitope mapping and peptide competition assays	68
RNA extraction, cDNA preparation, and qRT-PCR	68

Mass spectrometry	70
2D electrophoresis	71
Karyotyping	72
<b>Bibliography</b>	<b>73</b>
<b>Chapter 2 ZCCHC8 is Required for the Degradation of Pervasive Transcripts Originating from Multiple Genomic Regulatory Features</b>	<b>76</b>
<b>Introduction</b>	<b>77</b>
<b>Results</b>	<b>83</b>
<i>Zcchc8</i> knockout in mouse salivary gland cells	83
RNA sequencing of <i>Zcchc8</i> knockout cells	87
Effects of <i>Zcchc8</i> knockout on genomic regulatory features	94
Differential gene expression in <i>Zcchc8</i> knockout cells	107
Effects of <i>Zcchc8</i> knockout in mice	113
<b>Discussion</b>	<b>118</b>
<b>Materials and Methods</b>	<b>121</b>
Cell Culture and Maintenance	121
Biological Assays	121
<i>MTT proliferation</i>	121
<i>Colony formation assays</i>	122
<i>Soft agar colonization assays</i>	122
<i>Scratch assays</i>	122
CRISPR/Cas9 Knockout of <i>Zcchc8</i> in SIMS Cells	123
Western Blots	124
Immunofluorescence and Confocal Microscopy	126
RNA Extraction and Sequencing Library Preparation	127
Read Mapping and Differential Expression Analysis	128
Bioinformatics and Data Analysis	129
<b>Bibliography</b>	<b>130</b>
<b>Curriculum vitae</b>	<b>134</b>



## List of Tables

<b>Table 1.1 Anti-BTBD7 (NBP2-14364) product information</b>	<b>9</b>
<b>Table 2.1 Top 100 positively correlated loadings for the first principal component</b>	<b>93</b>
<b>Table 2.2 Differentially expressed genomic regulatory features</b>	<b>98</b>
<b>Table 2.3 Gene expression of RNA degradation complex subunits in <i>Zcchc8</i> KO cells</b>	<b>101</b>

## List of Figures

Figure 1.1 Schematic of the structural domains of BTBD7	6
Figure 1.2 Ex vivo cultures of E13.5 salivary glands from <i>Btbd7</i> KO mice	7
Figure 1.3 Ex vivo cultures of E13.5 salivary glands from <i>Btbd7</i> KO mice	10
Figure 1.4 Immunofluorescence images of BTBD7 expression in salivary glands	11
Figure 1.5 Fibronectin induction of BTBD7 in MDCK.2 cells	13
Figure 1.6 Fibronectin induction of BTBD7 in MDCK.2 cells	14
Figure 1.7 Immunofluorescence comparison of antibody production lots	16
Figure 1.8 Antibody validation through peptide competition assays	17
Figure 1.9 Antibody validation through target over-expression	17
Figure 1.10 Antibody validation through RNA interference	19
Figure 1.11 CRISPR/Cas9 knockout of <i>Btbd7</i> in SIMS and NIH/3T3 cells	20
Figure 1.12 Immunofluorescence staining for BTBD7 in SIMS knockout cells	22
Figure 1.13 Immunofluorescence staining for BTBD7 in NIH/3T3 knockout cells	23
Figure 1.14 Protein X immunofluorescence in wild-type salivary glands	25
Figure 1.15 Super-resolution imaging of protein X in cell nuclei	26
Figure 1.16 Epitope mapping of the immunogen used to generate the anti-BTBD7 antibody	28
Figure 1.17 Immunoprecipitation for protein X	31
Figure 1.18 2D western blot for resolving protein X	32
Figure 1.19 CRISPR/Cas9 knockout of <i>Zcchc8</i> in SIMS cells	34
Figure 1.20 Confirmation of the identity of protein X as ZCCHC8	36
Figure 1.21 Immunofluorescence staining for ZCCHC8 in salivary glands	38
Figure 1.22 Super-resolution imaging of ZCCHC8 in cell nuclei	39
Figure 1.23 Western blot comparison of anti-BTBD7 antibodies	41
Figure 1.24 Immunofluorescence comparison of anti-BTBD7 antibodies	42
Figure 1.25 Western blot comparison of anti-ZCCHC8 antibodies	44
Figure 1.26 Immunofluorescence comparison of anti-BTBD7 antibodies	45
Figure 1.27 Comparison of anti-BTBD7 antibody production lots against Human Protein Atlas Results	48
Figure 1.28 Amino acid sequence alignment of ZCCHC8 and BTBD7	49
Figure 2.1 CRISPR/Cas9 knockout of <i>Zcchc8</i> in SIMS cells	84
Figure 2.2 Biological assays of <i>Zcchc8</i> knockout cells	86
Figure 2.3 Principal component analysis of <i>Zcchc8</i> differential gene expression	89
Figure 2.4 Euclidean distance clustering of <i>Btbd7</i> and <i>Zcchc8</i> knockout clones	91
Figure 2.5 Loadings bar chart of the top 20 genes correlating with the variance of the first principal component	92

<b>Figure 2.6</b> Read coverage plots for the 3-kb region upstream of the TSS for all expressed genes in <i>Zcchc8</i> KO cells	<b>95</b>
<b>Figure 2.7</b> Venn diagrams and heat maps of differentially expressed PROMPTs	<b>96</b>
<b>Figure 2.8</b> Read coverage plots for differentially expressed PROMPTs	<b>97</b>
<b>Figure 2.9</b> Probability density functions for chromosomal distribution of PROMPTs	<b>102</b>
<b>Figure 2.10</b> Differential expression of genomic regulatory features in <i>Zcchc8</i> KO cells	<b>104</b>
<b>Figure 2.11</b> Chromosomal distribution of differentially expressed genomic regulatory features in <i>Zcchc8</i> KO cells	<b>106</b>
<b>Figure 2.12</b> Venn diagrams and heat maps of differentially expressed genes	<b>108</b>
<b>Figure 2.13</b> Scatter plots of gene expression data in <i>Zcchc8</i> KO cells	<b>109</b>
<b>Figure 2.14</b> Read coverage plots for differentially expressed RIKEN cDNAs and predicted genes	<b>111</b>
<b>Figure 2.15</b> Differential expression of genomic regulatory features in <i>Zcchc8</i> KO mice	<b>114</b>
<b>Figure 2.16</b> Scatter plots of gene expression data in <i>Zcchc8</i> KO mice	<b>116</b>

# **Chapter 1**

## **A Case Study of Antibody Cross-Reactivity**

# Introduction

Perhaps the most used tool in the biologist's toolbox is that of the antibody. From protein quantitation to structural imaging, disease diagnostics to disease therapy, the multipurpose nature of this tool has made it invaluable to both basic and medical research. According to Antibodypedia, there are currently more than 4.3 million antibodies targeting over 19,000 gene products available from 92 different providers (1). Grand View Research estimated the global research antibody market size to be USD 3.9 billion in 2019 with a predicted 6.2% growth rate through 2027 (2). Clearly, the importance of commercially produced antibodies in laboratory research is indisputable, yet there are currently few scientific standardized guidelines for validation of antibody-to-target specificity (3,4). This may not be entirely surprising considering the myriad antibody-based research techniques. However, the recently acknowledged reproducibility crisis in biomedical research has shone a bright spotlight on results produced with improperly validated antibodies (5-17), and the subsequent calls for validation standards have grown in number and volume (3,4,16,18-25).

Antibodies are a product of life's extraordinary immunological defense system, and the naive researcher can be forgiven for the misconception (no doubt perpetuated somewhat by imprecise terminology disseminated on commercial antibody product datasheets) that a specific antibody is capable of recognizing and binding to only a specific epitope. The more remarkable case, in fact, may be the antibody that is *only* capable of recognizing and binding a single epitope.

Most antibodies are Y-shaped hetero-oligomers comprised of two identical light chains and two identical heavy chains bound together by disulfide bonds. Each light

chain consists of a constant ( $C_L$ ) and a variable region ( $V_L$ ) whereas each heavy chain consists of three constant regions ( $C_{H1-3}$ ) and a single variable region ( $V_H$ ) (26-29). The stem of the “Y” is a heavy chain dimer of constant regions ( $C_{H2-3}$ ) referred to as the Fc fragment and binds to specific receptors on various immune cells as part of the immune response (26-29). The upper “Y” portions are dimers of light and heavy chains ( $C_L V_L$  and  $C_{H1} V_H$ ) referred to as Fab or F(ab) fragments and bind to antigens (26-29). The Fab and Fc fragments are connected by a protease-cleavable hinge region within the heavy chains (26-29).

Antibodies recognize antigens through six complementarity determining regions (CDRs) that associate non-covalently and form the antigen binding site located at the N-terminal tip of each Fab fragment (26-29). These CDRs have enormous sequence variability and are often referred to as hypervariable loops. Within the antigen binding site is a cleft of 50-70 amino acids with several overlapping 15-20 amino acid sites called paratopes (26-29). Paratopes provide structural and chemical complementarity to surface regions of antigens known as epitopes (26-29).

Epitopes can consist of virtually any biomolecule and are not restricted to peptides alone (30-35). Often, epitopes consist of modified proteins with phosphorylated or glycosylated residues, for example. Protein epitopes are usually classified as continuous or discontinuous depending on the continuity of the amino acids within the protein that forms the epitope (28,36,37). Continuous epitopes are linear amino acid sequences, and recent studies suggest the sequence that serves as the epitope ranges from four to twelve residues (36). The majority of protein epitopes

are discontinuous and consist of two to five short stretches of amino acids that are brought together at the surface due to protein folding (28,37).

As an antibody binding site encompasses several individual paratopes, each able to bind differing (or the same) epitopes present in one or more antigens, a single antibody is never monospecific for a single binding partner (28,29,36,37). To be so would require that the remaining majority of amino acids in the antibody binding cleft that do not bind the “specific” binding partner are also unable to bind any other antigenic structure.

While steric hindrance may prevent two antigen molecules from binding simultaneously within the same antibody binding cleft, there is little to prevent two separate antigens from binding the same antibody binding site individually and independently, provided there is sufficient chemical and structural recognition. Within the context of an organism’s robust immunological response, such non-specificity may be of little impact overall. It is of great import to the bench researcher.

Antibody cross-reactivity is not the only pitfall. Antibodies are also subject to lot-to-lot variation and loss of function/activity that may or may not be linked to storage conditions or variable shelf life. Though companies generally provide some validation data, ultimately, it is up to the researcher to validate a particular antibody and subsequently determine its capability to perform well in the required application. The ideal antibody would be target-specific and selective with reproducible results and an absence of variability between production lots. Practically, researchers often must settle for relative specificity, selectivity, and variability while maintaining absolute reproducibility. Validation of an antibody for a particular application, therefore, must be

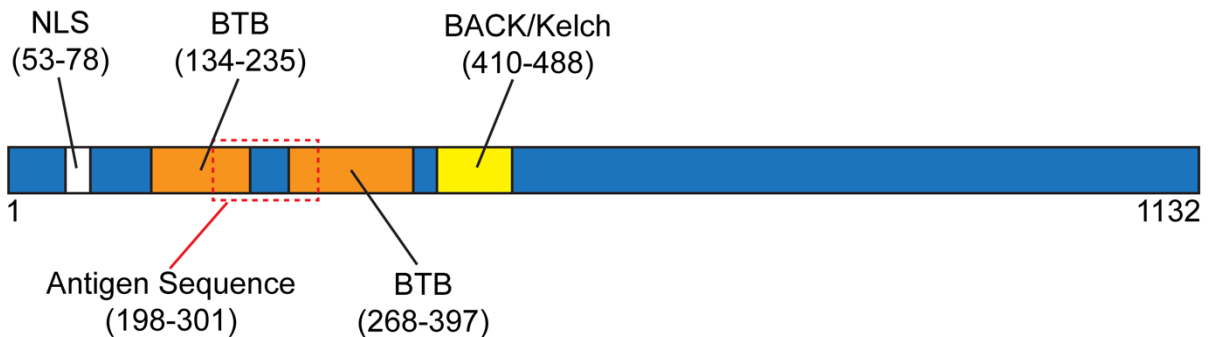
carried out with same care and rigor as the experiments devoted to hypothesis testing. Critical analysis of validation results is imperative as haphazard acceptance can lead to unfavorable research outcomes at best, disastrous at worst.

Here we present a case study of an antibody (anti-BTBD7, Novus NBP2-14364) that, despite careful validation, was ultimately determined to be highly cross-reactive with an unrelated protein. Further, this anti-BTBD7 antibody was found to suffer from broad lot-to-lot variation and application inconsistency. Through extensive efforts, we identified the major cross-reacting protein of interest to be that of the unrelated protein, ZCCHC8. We also examined a panel of anti-ZCCHC8 antibodies and found extensive variability in western blot and immunofluorescence staining patterns. Despite having no amino acid sequence homology between the BTBD7 and ZCCHC8 antigen peptides, our results suggest these two proteins share a common structural epitope.



## Background

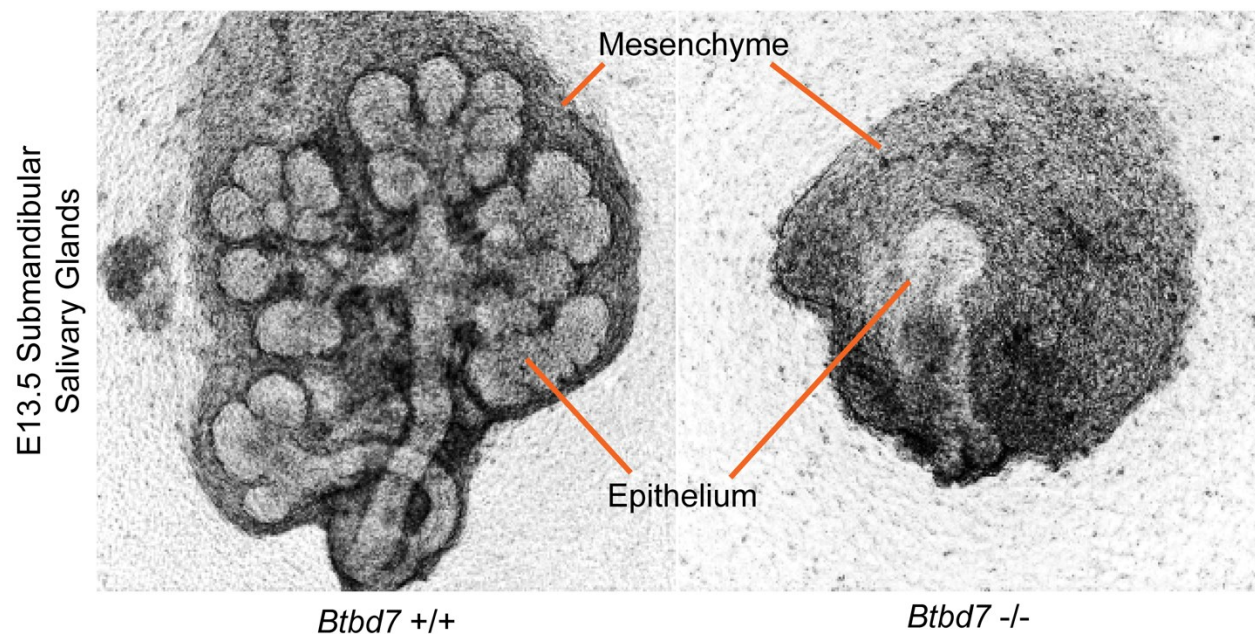
BTBD7 is the seventh member of the BTB/POZ-domain family of proteins and derives its name from the *Drosophila* Bric á brac, Tramtrack, and Broad-complex proteins that contain a highly conserved structural domain (38). BTBD7 is a 126 kD protein that is highly conserved across vertebrate species and is somewhat unique in that it contains two BTB domains. These two BTB domains are in the N-terminal half of the protein and are followed closely by a BACK/Kelch domain (BTB and C-terminal Kelch; a domain found in most BTB-containing proteins) (39). A putative, classical bipartite nuclear localization sequence is located near the N-terminus. A schematic of BTBD7 is shown in Figure 1.1.



**Figure 1.1** Schematic depicting the locations of the major structural domains within the human BTBD7 protein. The red, dashed box represents the location of the peptide sequence used to generate the anti-BTBD7 antibody used in this case study. Amino acid numbers are in parenthesis.

Previous research in our lab identified BTBD7 as a regulator of branching morphogenesis (40). At the time of this research, *Btbd7* was relatively unstudied, and commercially produced anti-BTBD7 antibodies were not readily available. Our lab contracted with multiple companies to produce an antibody but ultimately these

products were deemed unsuitable for our desired applications. Thus, BTBD7 expression studies were relegated to mRNA and GFP-fusion cDNA constructs. Consequently, our lab generated a *Btbd7* knockout mouse in order to more thoroughly explore epithelial dynamics in branching tissues. These knockout mice displayed a high rate of embryonic mortality with most deaths occurring between days E13 and E17 (though occasional pups were born but succumbed after a few days). The *Btbd7* knockout mice displayed diminished branching of certain epithelial tissues including salivary glands, an important model for developmental epithelial dynamics used extensively in our lab (Figure 1.2).



**Figure 1.2** *Ex vivo* cultures of E13.5 submandibular salivary glands taken from wild-type (left) and *Btbd7* knockout (right) mice. Glands were cultured for two hours before imaging. Experiment and images courtesy of Will Daley.

## Immunofluorescence Validation

In the time between the publication of *Btbd7* as a regulator of branching morphogenesis and the generation of our lab's knockout mouse, commercial anti-BTBD7 antibodies for various applications became available. We purchased an anti-BTBD7 antibody from Novus Biologicals (NBP2-14364) for use in immunocytochemistry/immunofluorescence (ICC/IF) experiments. For simplicity, we will refer to the antibody production lot numbers as  $k_0$  for the initial production lot used in our experiments with increasing subscripts for each successive production lot. The product details from the accompanying data sheet are listed in Table 1.1. Of note, this antibody was developed against a 104-amino acid sequence that partially spans the two BTB domains (Figure 1.1) and is stated to work for immunohistochemistry (IHC) and ICC/IF in human, mouse, and rat tissues. The 104-amino acid region is highly conserved with only two amino acid substitutions between humans and mice (amino acids 35 and 83 of the antigen) and only one substitution between both humans and rats, and humans and canines (amino acid 83 of the antigen). Thus, the antibody might reasonably be predicted to work in multiple species of cell lines.

Immunofluorescence staining of knockout tissues using anti-BTBD7 antibody lot  $k_0$  revealed a unique staining pattern in wild-type mice and an absence of staining in knockout tissues (Figure 1.3). BTBD7 staining appeared uniform across mesenchymal cell nuclei, and large cytoplasmic granules were visible throughout the epithelium and mesenchyme (Figure 1.4A-B). Close examination of BTBD7 immunofluorescence revealed an enriched, nuclear staining pattern in peripheral epithelial cells (Figure 1.4A-

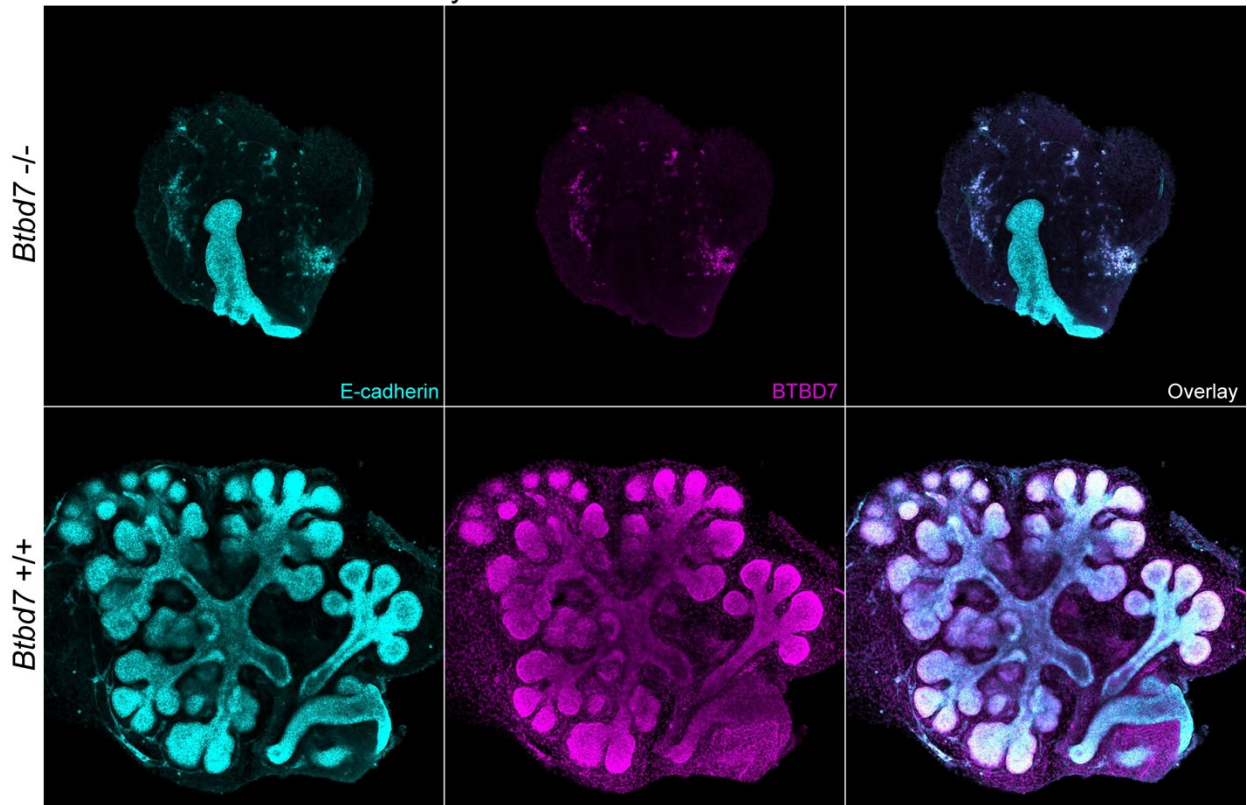
B). Higher resolution imaging of peripheral epithelial nuclei revealed a uniquely punctate BTBD7 staining pattern (Figure 1.4C).

**Table 1.1** Anti-BTBD7 (NBP2-14364) product information

<b>Summary</b>	
<b>Reactivity</b>	Human, Mouse, Rat
<b>Applications</b>	ICC/IF, IHC, IHC-Paraffin
<b>Clonality</b>	Polyclonal
<b>Host</b>	Rabbit
<b>Conjugate</b>	Unconjugated
<b>BTBD7 Antibody Summary</b>	
<b>Immunogen</b>	This antibody was developed against a recombinant protein corresponding to amino acids: LLHYLYTGEFGMEDSRFQNVLDILVQLSEEFGTPNSLDVD MRGLFDYMCYYDVVLSFSSDSELVEAFGGNQNCLDEEL KAHKAVISARSPFFRNLLQRRIRTGEE
<b>Specificity</b>	Specificity of human BTBD7 antibody verified on a Protein Array containing target protein plus 383 other non-specific proteins.
<b>Predicted Species</b>	Mouse (98%), Rat (99%). Backed by 100% Guarantee.
<b>Isotype</b>	IgG
<b>Clonality</b>	Polyclonal
<b>Host</b>	Rabbit
<b>Gene Symbol/Gene ID</b>	Btbd7/55727
<b>Purity</b>	Immunogen affinity purified
<b>Unit Size</b>	0.1 ml
<b>Concentration</b>	Concentrations vary lot to lot. See vial label for concentration. If unlisted, please contact technical services.
<b>Storage</b>	Store at 4C short term. Aliquot and store at -20C long term. Avoid freeze-thaw cycles.
<b>Applications/Dilutions</b>	
<b>Dilutions</b>	<ul style="list-style-type: none"> <li>Immunocytochemistry/Immunofluorescence 0.25-2 ug/ml</li> <li>Immunohistochemistry 1:200 - 1:500</li> <li>Immunohistochemistry-Paraffin 1:200 - 1:500</li> </ul>
<b>Application Notes</b>	For IHC-Paraffin, HIER pH 6 retrieval is recommended. ICC/IF Fixation Permeabilization: Use PFA/Triton X-100.

[https://www.novusbio.com/products/btbd7-antibody\\_nbp2-14364](https://www.novusbio.com/products/btbd7-antibody_nbp2-14364)

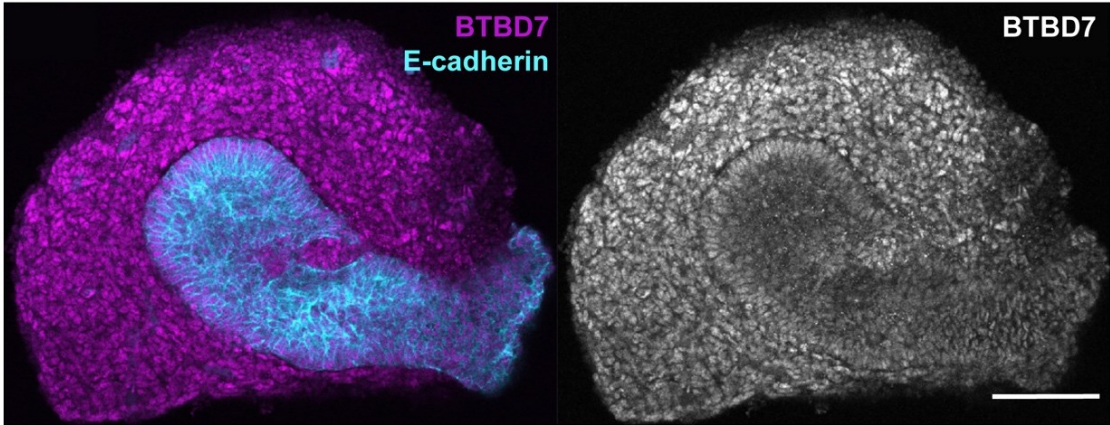
### E13.5 Submandibular Salivary Glands



**Figure 1.3** *Ex vivo* cultures of E13.5 submandibular salivary glands taken from *Btbd7* knockout (top) and wild-type (bottom) mice. BTBD7 is absent from the epithelial bud in knockout tissue (middle). Glands were cultured for two hours before imaging. Note the presence of multi-channel, auto-fluorescent cells that are abundant in the mesenchyme. These cells are frequently present in varying number during *ex vivo* salivary gland culture. Experiment and images courtesy of Will Daley.

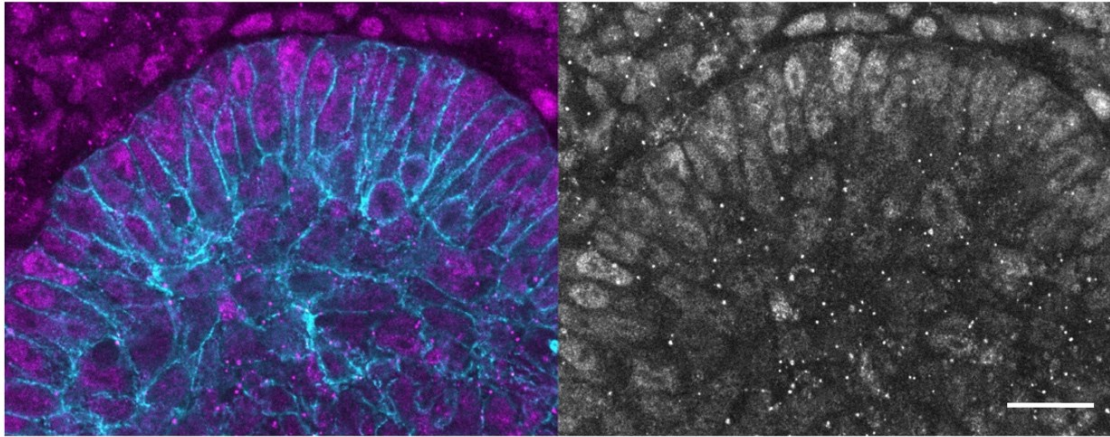
**A**

E11.5 Submandibular  
Salivary Gland (20X)



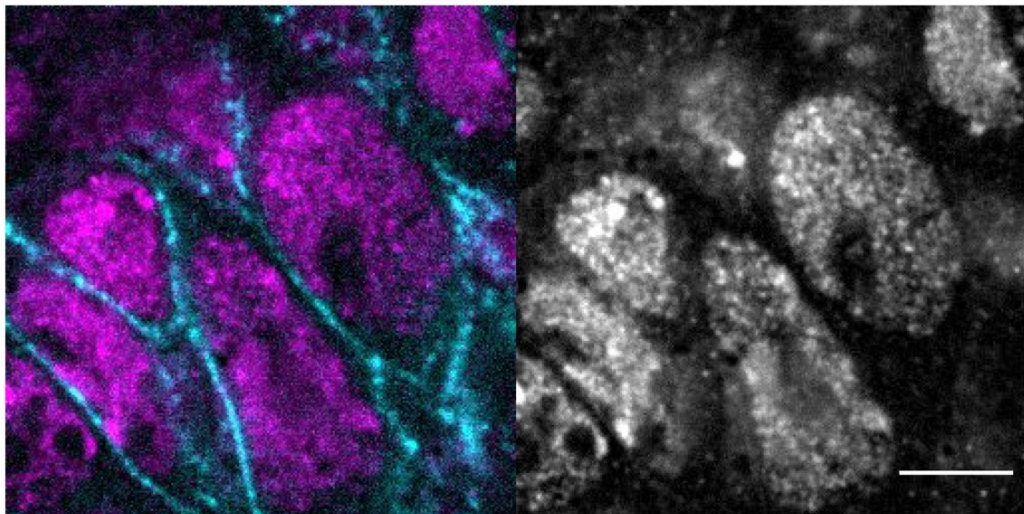
**B**

E11.5 Submandibular  
Salivary Gland (100X)



**C**

E13.5 Submandibular  
Salivary Gland (150X)

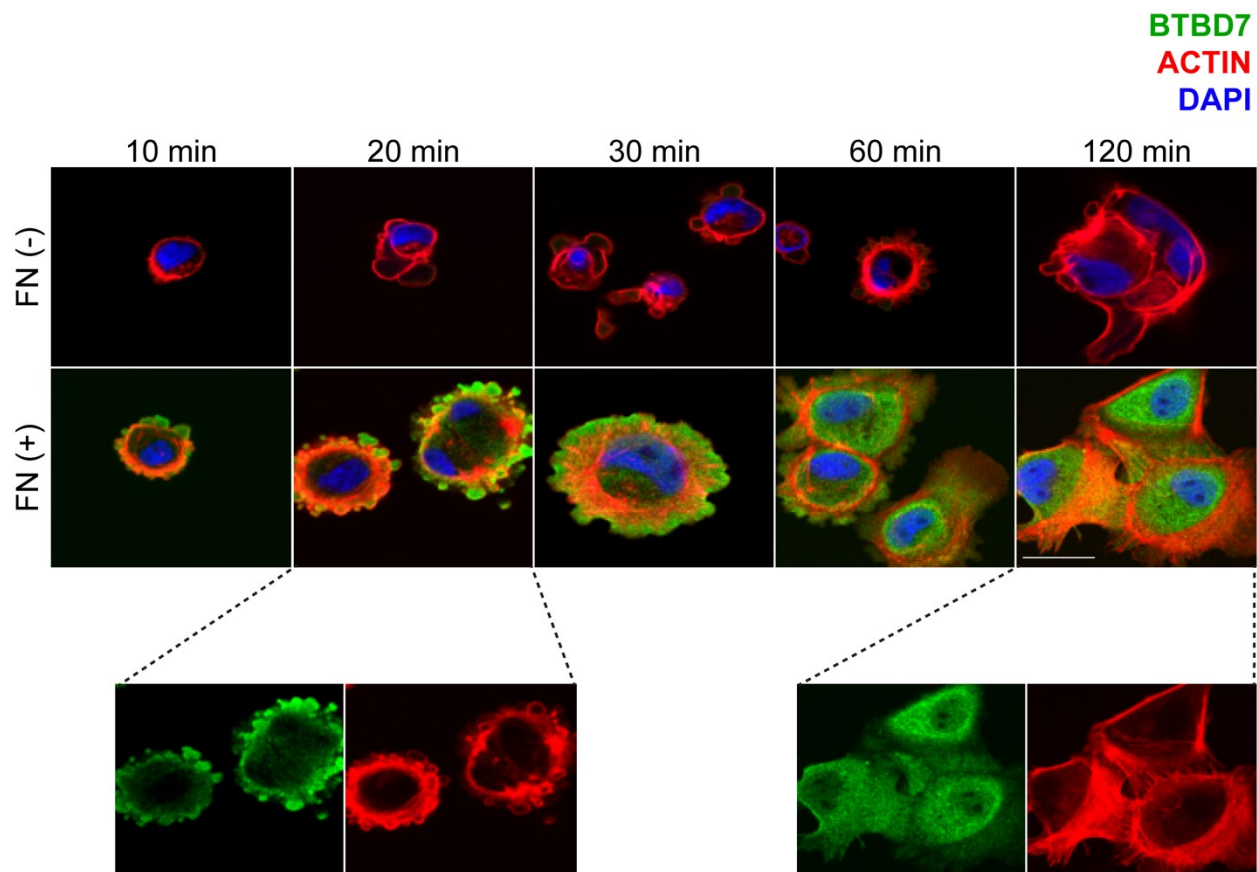


**Figure 1.4** (A) 20X IF images of E11.5 submandibular salivary glands from wild-type mice. Epithelial tissues are delineated by E-cadherin in cyan. BTBD7 staining is in magenta and grayscale. Note the enhanced staining of BTBD7 in peripheral epithelial cells (grayscale). Scale bar = 100  $\mu\text{m}$ . (B) 100X IF images of E11.5 submandibular salivary glands from wild-type mice. Epithelial cells and BTBD7 are stained as in (A). Note the large cytoplasmic granules in both the epithelium and mesenchyme. Scale bar = 20  $\mu\text{m}$ . (C) 150X IF images of E11.5 submandibular salivary glands from wild-type mice. Epithelial cells and BTBD7 are stained as in (A). Note the punctate pattern of BTBD7 staining in the epithelial cell nuclei. Images were processed using a rolling-ball background subtraction method. Scale bar = 5  $\mu\text{m}$ . Images in (A-C) are from single axial planes taken mid-gland. Images in (A) and (B) courtesy of Shaohe Wang.

In the initial characterization of BTBD7 as a regulator of branching morphogenesis, our lab showed that fibronectin could induce *Btbd7* mRNA expression (40). Later, preliminary immunofluorescence experiments seemed to confirm these results. Madin-Darby canine kidney cells (MDCK.2) plated on fibronectin coated MatTek dishes displayed increased BTBD7 staining compared to cells plated on uncoated dishes. BTBD7 staining was enhanced in peripheral blebs of cells undergoing cell spreading. Over time, BTBD7 staining appeared uniformly diffuse and enriched in the cytoplasm of spread cells with lesser diffuse staining evident in nuclei. These results are displayed in Figure 1.5. It should be noted that this staining pattern differs from the staining pattern seen *in vivo* (Figure 1.4C).

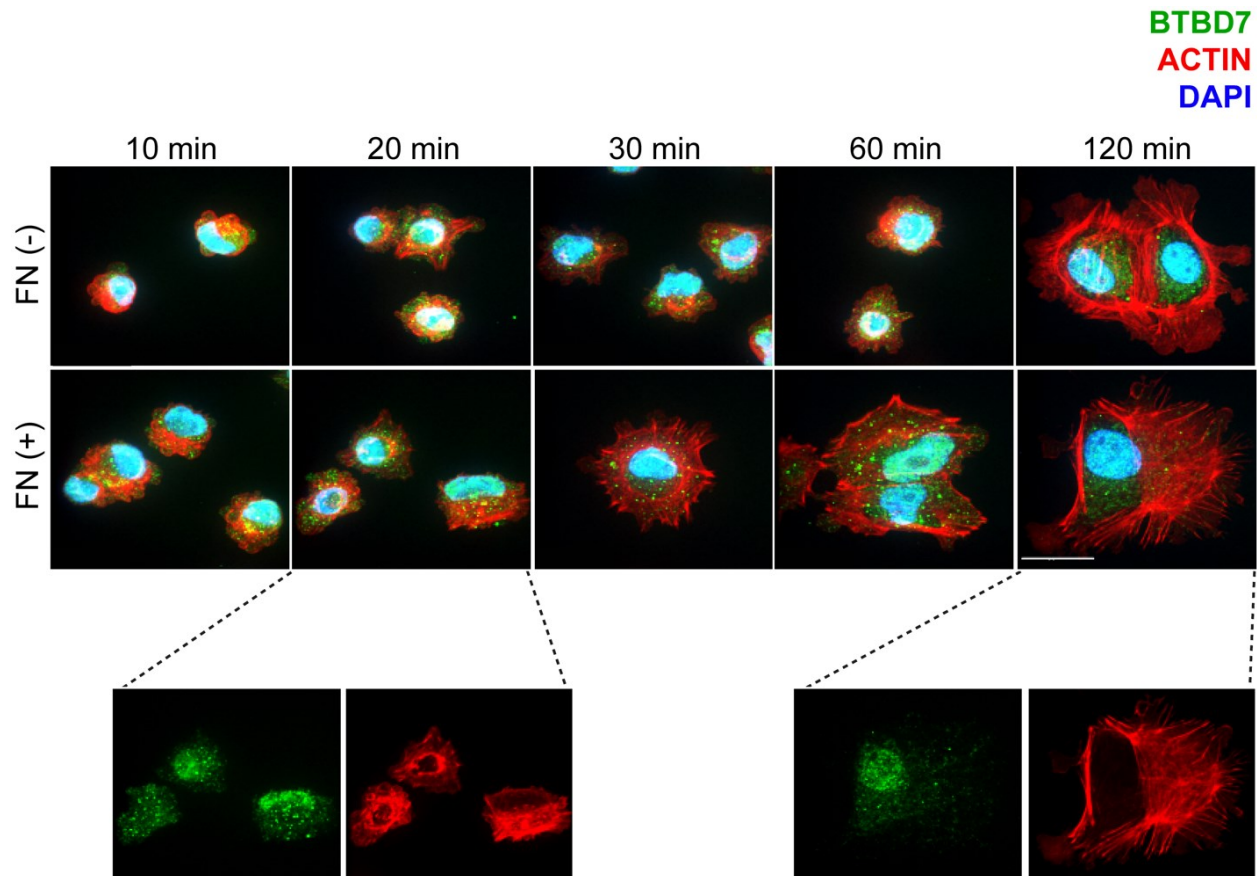
These preliminary findings went unexplored for some time. Eventually, our lab's focus returned to mechanistic studies of *Btbd7* regulation. Attempts to reproduce the findings depicted in Figure 1.5 were confounding. Under identical experimental conditions, BTBD7 staining of MDCK.2 cells plated on fibronectin-coated MatTek dishes were indistinguishable from cells on uncoated dishes. Further, rather than the uniformly diffuse cytoplasmic enrichment of the initial experiments, BTBD7 staining was highly

enriched in the nucleus with a unique punctate pattern. The now scarce cytoplasmic staining displayed large, non-uniform granules. These results are displayed in Figure 1.6 and more closely resemble the *in vivo* staining shown in Figure 1.4C.



**Figure 1.5** Madin-Darby Canine Kidney (MDCK.2) cells were plated on uncoated (FN -) or fibronectin-coated (FN +; 100  $\mu\text{g}/\text{ml}$ ) MatTek dishes. At the indicated intervals of post-plating time, cells were fixed and stained for BTBD7, actin (rhodamine phalloidin), and nuclei (DAPI). The anti-BTBD7 antibody was from production lot k<sub>0</sub>. Note the insets at 20 and 120 minutes show individual BTBD7 and ACTIN staining channels. BTBD7 staining changes from highly concentrated at peripheral blebs to uniformly diffuse throughout the cytoplasm over time. BTBD7 staining is enriched in the cytoplasm versus the nucleus at 120 minutes. Images are maximum intensity projections. Scale bar = 10  $\mu\text{m}$ . Experiment and images courtesy of Will Daley.





**Figure 1.6** Madin-Darby Canine Kidney (MDCK.2) cells were plated on uncoated (FN -) or fibronectin coated (FN +; 100  $\mu\text{g}/\text{ml}$ ) MatTek dishes. At the indicated intervals of post-plating time, cells were fixed and stained for BTBD7, actin (rhodamine phalloidin), and nuclei (DAPI). The anti-BTBD7 antibody was from production lot  $k_1$ . Note the insets at 20 and 120 minutes show individual BTBD7 and ACTIN staining channels. BTBD7 now appears consistent over time with obvious enrichment in the nucleus versus the cytoplasm. Cytoplasmic staining is scarce with large granules present. Images are maximum intensity projections. Scale bar = 10  $\mu\text{m}$ .

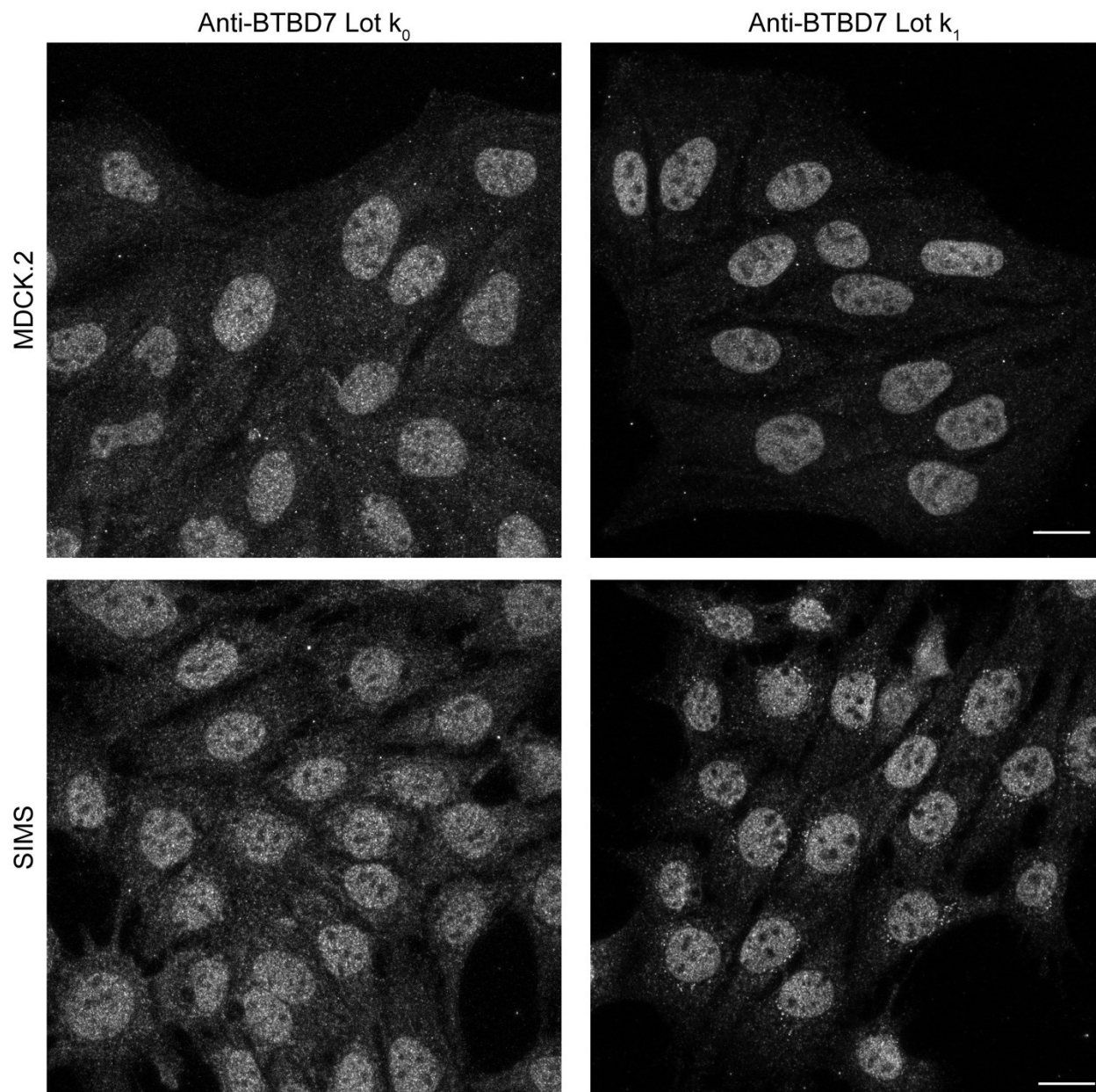
Initially, this discrepancy was attributed to polyclonal antibody lot-to-lot variation as the experiment in Figure 1.5 used production lot  $k_0$  and the experiment in Figure 1.6 used production lot  $k_1$ . However, side-by-side comparison immunofluorescence revealed highly similar results across antibody production lots in both MDCK.2 cells and SIMS mouse submandibular salivary gland cells (Figure 1.7). After identical immunofluorescent staining parameters, microscopy settings, and image processing,

the nuclear staining pattern appeared highly similar between anti-BTBD7 lots  $k_0$  and  $k_1$ . Cytoplasmic staining using lot  $k_1$  appeared slightly diminished relative to lot  $k_0$  giving the impression of reduced, non-specific background staining.

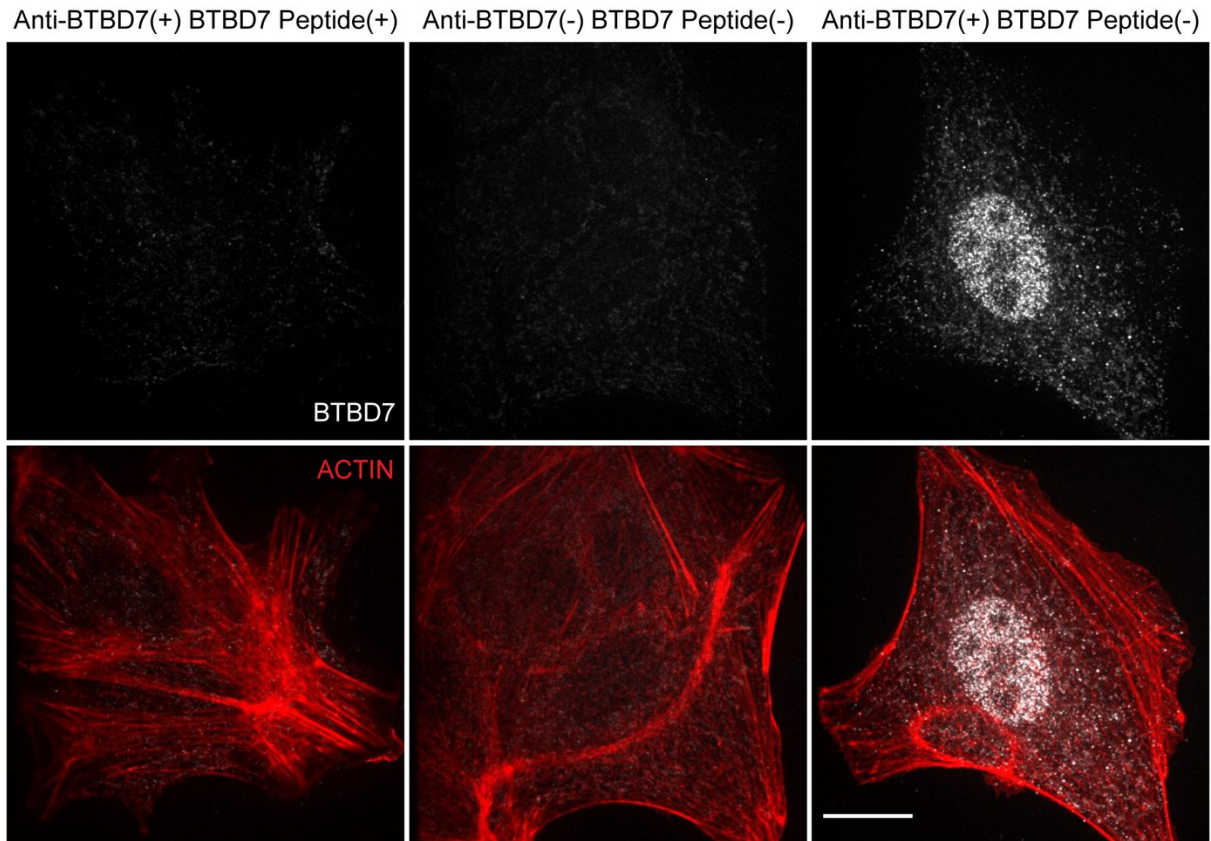
The extreme differences in results between the experiment depicted in Figure 1.5, and those repeated experiments for which the results are depicted in Figure 1.6, were ultimately attributed to variances in experimenter handling. Nevertheless, we sought to re-validate the anti-BTBD7 antibody for use in cell-culture based immunofluorescence. As our supply of anti-BTBD7 lot  $k_0$  was nearing depletion, we focused our validation efforts on lot  $k_1$ . We first performed peptide competition assays using the 104-amino acid peptide that served as the antigen for producing the antibody. The anti-BTBD7 antibody was incubated with a 10X excess amount of BTBD7 peptide prior to immunostaining MDCK.2 cells. A slight residual staining was still evident after immunostaining, though this was likely due to secondary antibody background staining as a comparable weak staining pattern was evident in the control condition lacking the primary anti-BTBD7 antibody (Figure 1.8). The results show that anti-BTBD7 lot  $k_1$  antibodies preferentially bind to the BTBD7 antigen.

In previous research, our lab created an MDCK.2 cell line that stably expresses a tetracycline-regulated, GFP-BTBD7 fusion protein (40). We used this cell line to test if lot  $k_1$  anti-BTBD7 antibodies would immunostain this fusion protein. The results of this experiment are shown in Figure 1.9 and confirm that anti-BTBD7 lot  $k_1$  successfully binds BTBD7 when over-expressed.

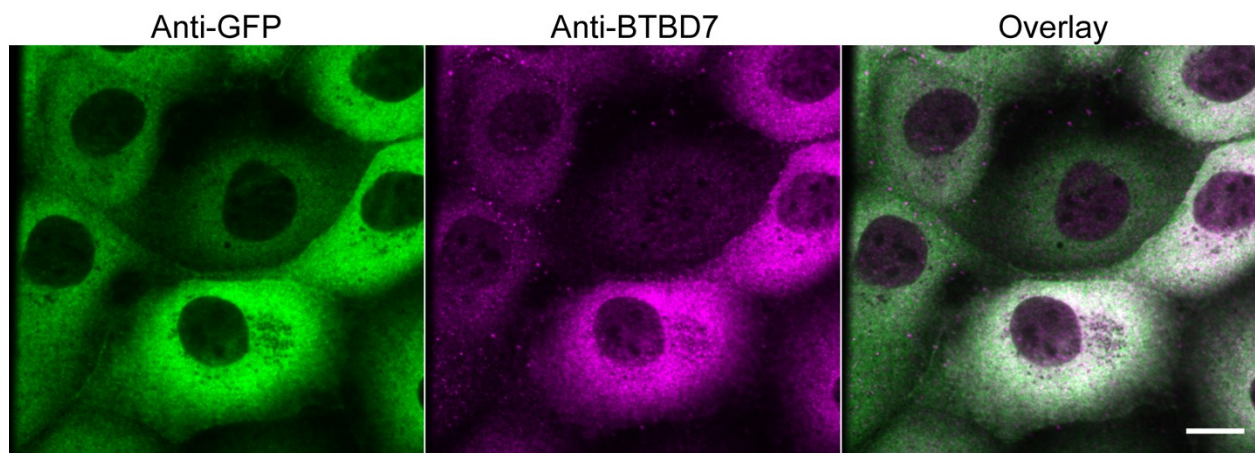
With limited information, resources, and tools readily available for working with BTBD7 in mouse and canine cells, we looked to expand our capabilities by using



**Figure 1.7** MDCK.2 and SIMS cells were stained for BTBD7 using two different production lots ( $k_0$  and  $k_1$ ) of the same antibody. The nuclear staining pattern appears consistent, while the cytoplasmic staining appears diminished in lot  $k_1$ . Immunofluorescence, microscopy settings, and image processing are identical across antibody lots. Images are maximum intensity projections. Scale bar = 15  $\mu\text{m}$ .



**Figure 1.8** MDCK.2 cells were used for peptide competition assays. Anti-BTBD7 lot  $k_1$  was incubated with 10X excess BTBD7 peptide for 30 minutes at room temperature prior to staining. Cells were then immunostained with either anti-BTBD7 plus peptide (left), neither (center), or anti-BTBD7 alone (right). Immunofluorescence, microscopy settings, and image processing were identical across conditions. Images are maximum intensity projections. Scale bar = 15  $\mu\text{m}$ .

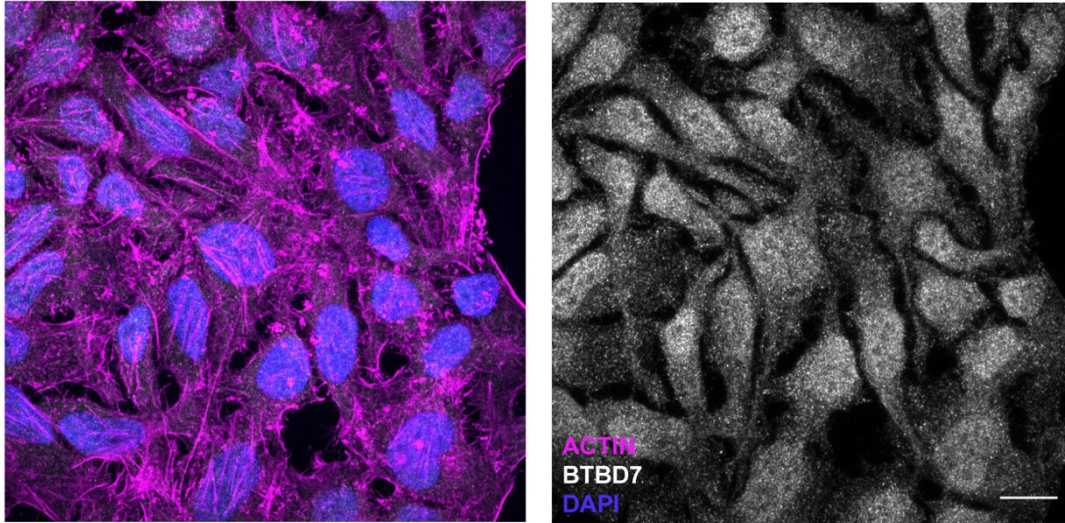
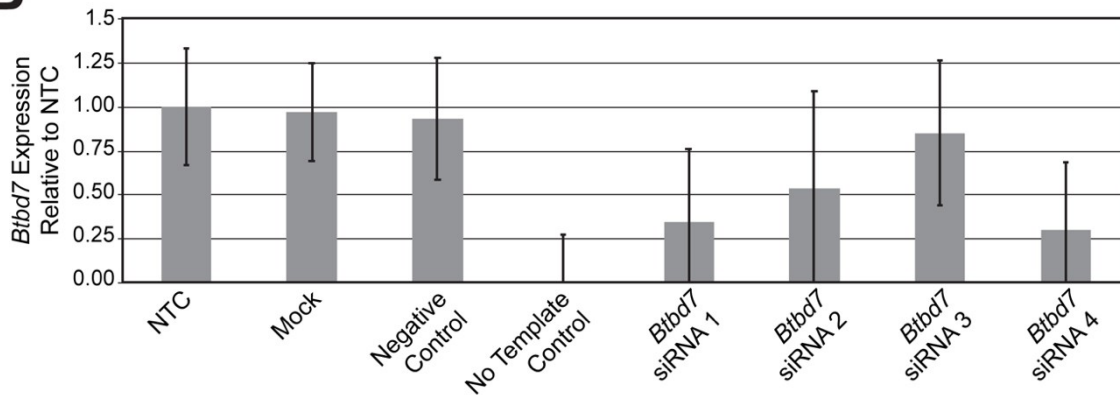
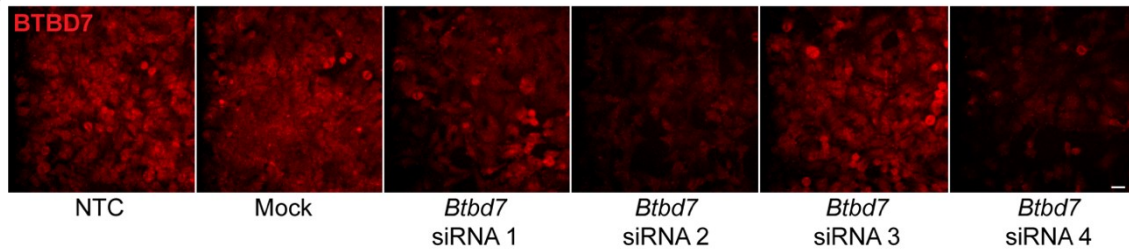


**Figure 1.9** MDCK.2 cells stably expressing a GFP-BTBD7 fusion protein were immunostained for anti-GFP (left) and anti-BTBD7 lot  $k_1$  (center). Note the enriched, diffuse cytoplasmic staining. Images are maximum intensity projections. Scale bar = 15  $\mu\text{m}$ .

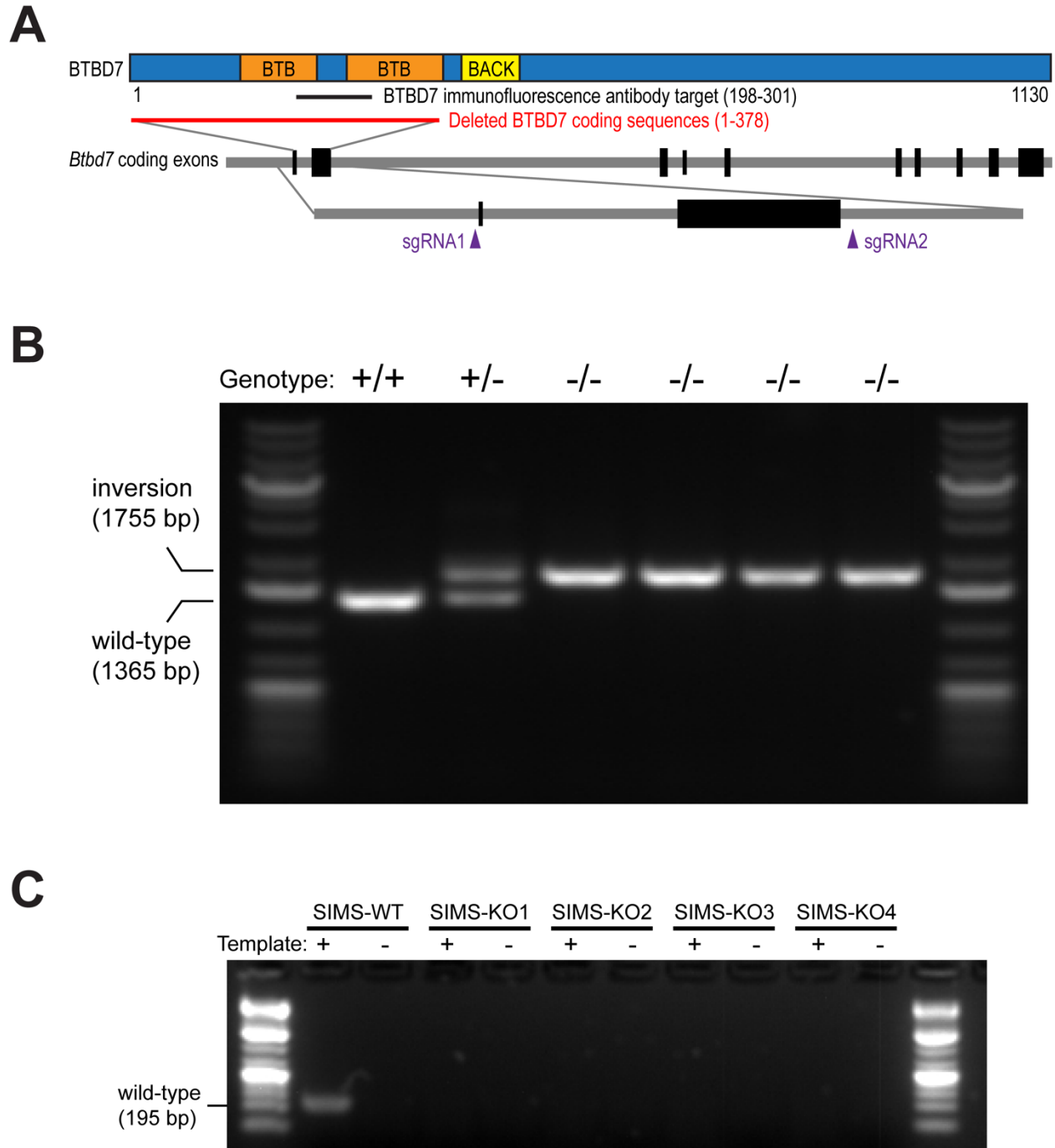
human cells. Immunostaining HeLa cells with anti-BTBD7 lot k<sub>1</sub> displayed only a slight nuclear enrichment in contrast to the prominent nuclear enrichment in MDCK.2 or SIMS cells (Figure 1.10A). We performed BTBD7 siRNA knockdown experiments in HeLa cells and confirmed the results with immunofluorescence. HeLa cells were transfected with four separate *Btbd7* siRNA constructs and mRNA expression was measured 48 hours post-transfection using qRT-PCR. Three of the four siRNAs resulted in decreased mRNA expression when compared to controls (Figure 1.10B). In complementary experiments, siRNA transfected cells were immunostained for BTBD7 and the results closely mirrored the qRT-PCR results (Figure 1.10C). These results confirmed the relative specificity of anti-BTBD7 lot k<sub>1</sub> in HeLa cells and also served as a validation of anti-BTBD7 lot k<sub>1</sub> for immunostaining human cells.

Based on the results of multiple validation strategies, we concluded that the anti-BTBD7 antibody from Novus Biologicals (NBP2-14364) was adequately target specific and sensitive for use in immunofluorescence experiments.

As mechanistic research regarding BTBD7 continued in our lab, we eventually turned our attention to CRISPR/Cas9 gene deletion experiments. In order to precisely mimic the conditions used in producing the *Btbd7* knockout mouse (40), a CRISPR/Cas9 strategy was developed to delete the first two coding exons in *Btbd7* (Figure 1.11A). Multiple *Btbd7* knockout clones in SIMS cells were generated and genomic analysis revealed an inversion of the CRISPR/Cas9 targeted region (Figure 1.11B) resulting in the complete ablation of functional *Btbd7* mRNA expression (Figure 1.11C).

**A****B****C**

**Figure 1.10** (A) Immunostaining for BTBD7 in HeLa cells. Images are maximum intensity projections. Scale bar = 15  $\mu$ m. (B) qRT-PCR results for mRNA expression of *Btbd7* 48 hours post siRNA transfection in HeLa cells. Mean expression  $\pm$  s.d. ( $n = 3$ ) is relative to non-transfected control (NTC). (C) Immunofluorescence images of HeLa cells transfected with siRNAs as in (B). Images are maximum intensity projections of representative fields of view taken with a 20X objective. Scale bar = 25  $\mu$ m.

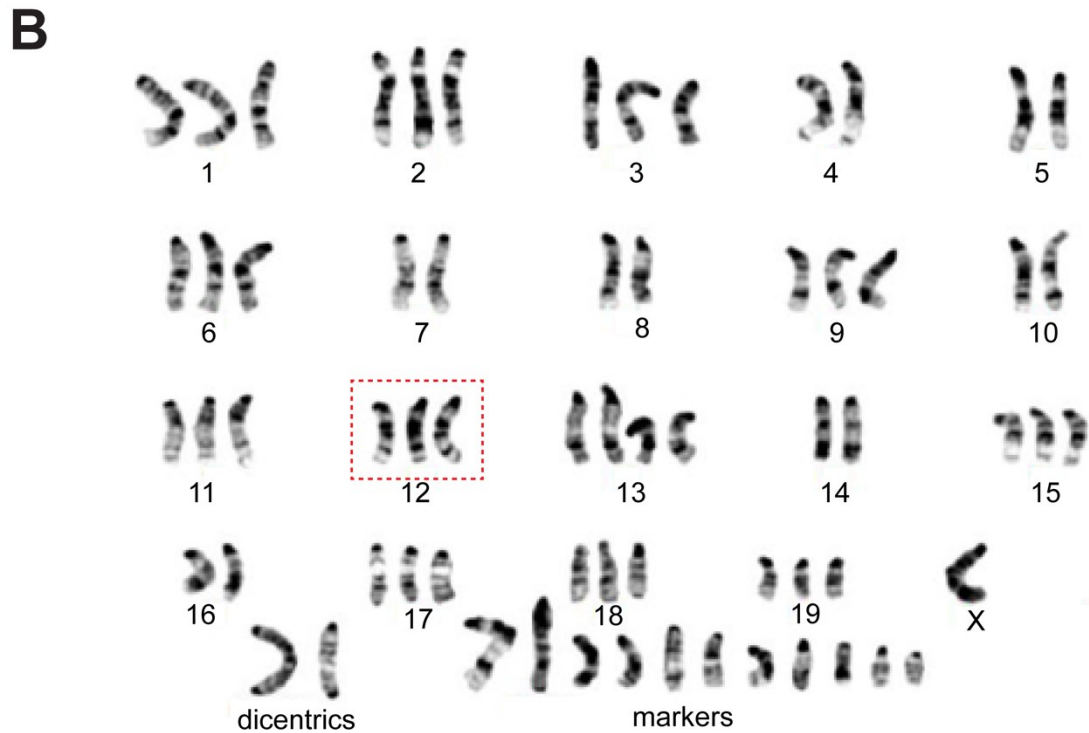
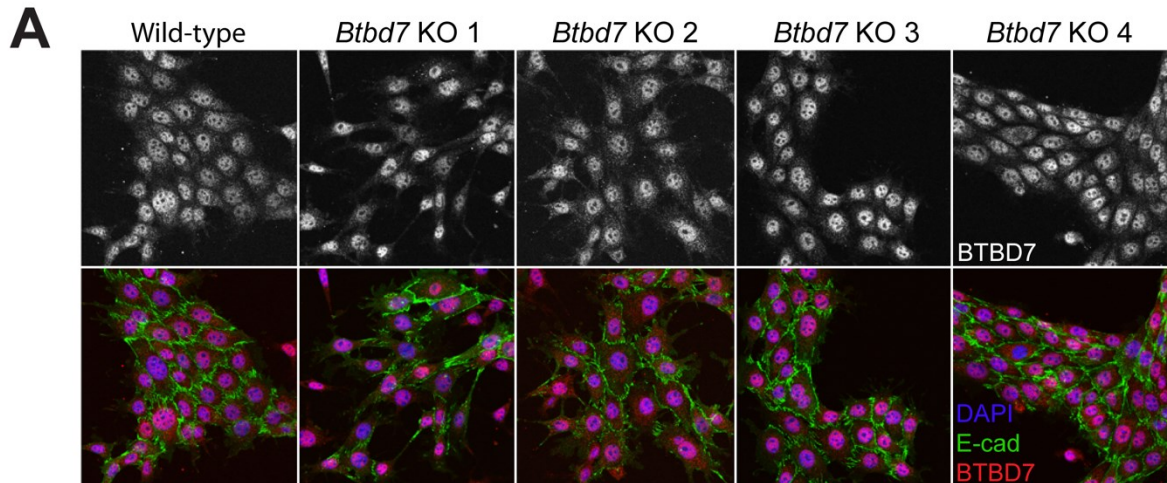


**Figure 1.11** (A) CRISPR/Cas9 strategy for disrupting *Btbd7* in SIMS and NIH/3T3 cells. (B) Genomic PCR results confirming the successful disruption of *Btbd7* in multiple SIMS clones. The upper band is the product of inverting the targeted gene region in (A) during the DNA repair process. Note the presence of two bands in lane 3 signifying at least one wild-type *Btbd7* copy exists and at least one inversion event occurred. (C) RT-PCR results confirming the absence of functional *Btbd7* mRNA for the three homozygous knockout clones in (B). Experiments courtesy of Shaohe Wang.

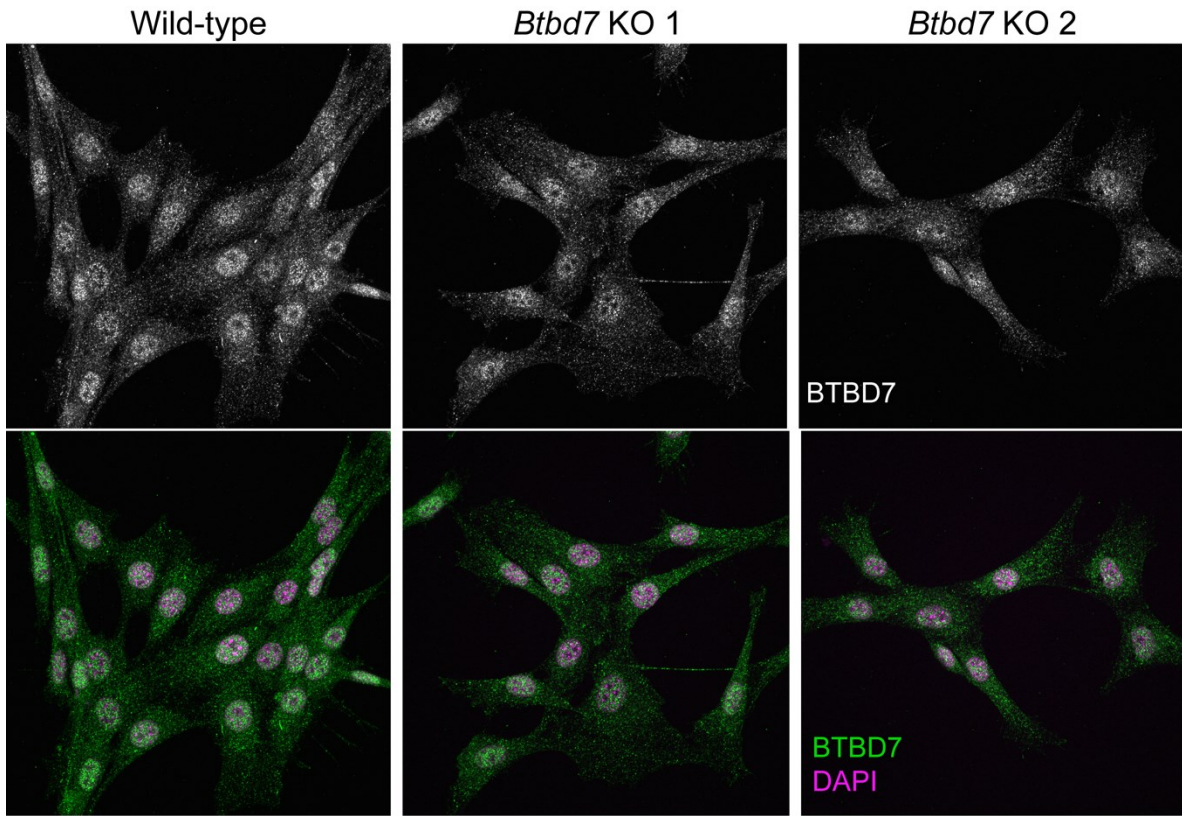
We then attempted to verify *Btbd7* gene disruption through immunofluorescence for BTBD7 protein. Unexpectedly, each *Btbd7* knockout clone tested was indistinguishable from the wild-type (Figure 1.12A). Considering our extensive antibody validation results, we assumed that our CRISPR/Cas9 deletion attempts were incomplete and that perhaps the SIMS cells contained multiple copies of the *Btbd7* gene. Indeed, karyotyping of the SIMS cells revealed 60-65 chromosomes with at least 3 copies of chromosome 12 which contains *Btbd7*. Additionally, there were numerous dicentric and marker chromosomes present (Figure 1.12B).

Considering the potential difficulty of disrupting an unknown number of *Btbd7* gene copies in SIMS cells, we opted to use our CRISPR/Cas9 deletion strategy in NIH/3T3 cells. Shockingly, after multiple strategies confirmed successful *Btbd7* gene disruption in these cells, immunofluorescence for BTBD7 protein was indistinguishable between control and knockout cells (Figure 1.13).





**Figure 1.12** (A) Immunofluorescence for BTBD7 in SIMS wild-type and knockout clones suggesting either anti-BTBD7 antibody cross-reactivity or incomplete knockout of *Btbd7*. Images are maximum intensity projections. Experiment courtesy of Shaohe Wang. (B) Karyotype of SIMS wild-type cells showing at least three copies of *Btbd7* on chromosome 12 are present. Note the numerous dicentric and marker chromosomes of unknown origin.



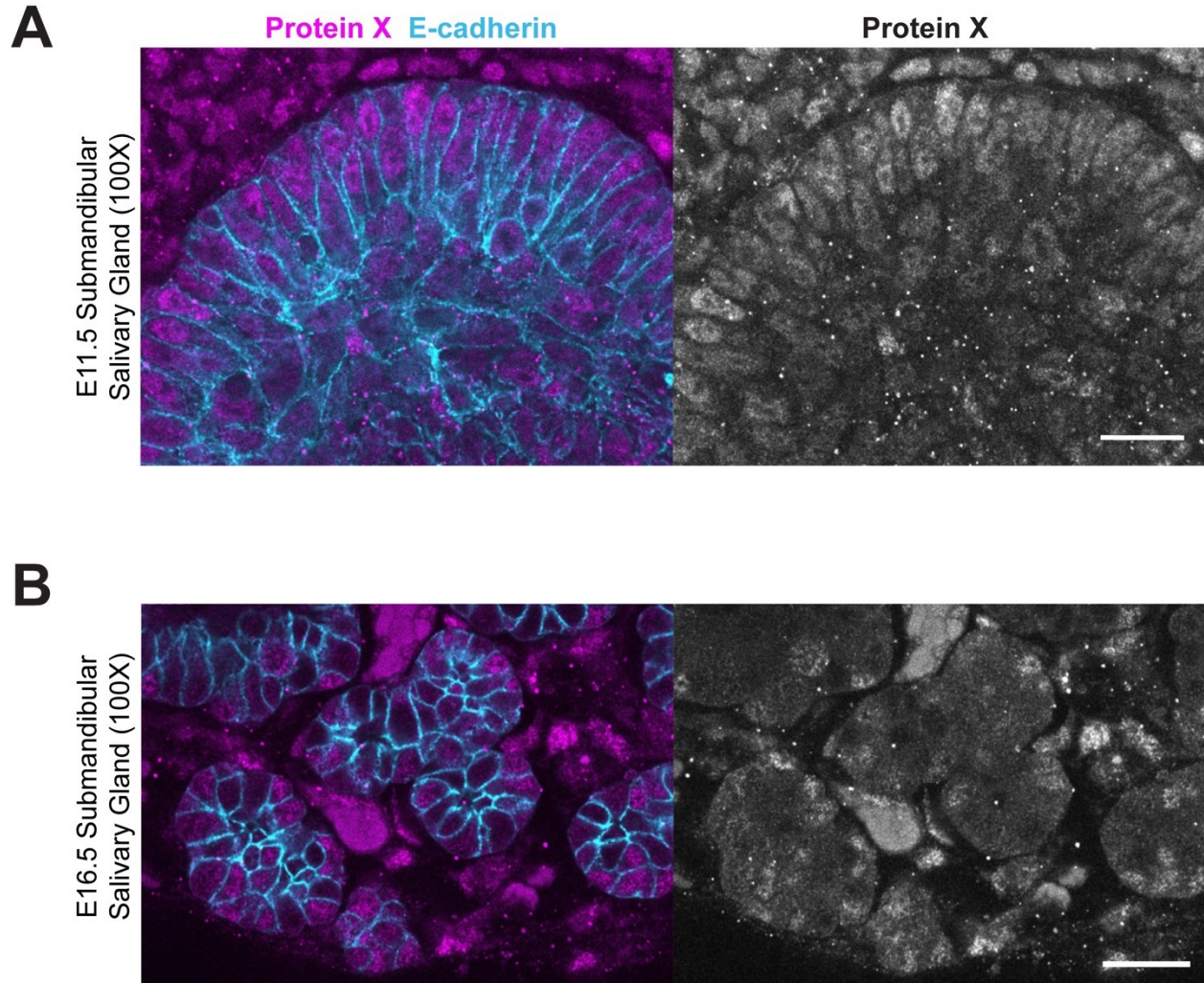
**Figure 1.13** (A) Immunofluorescence for BTBD7 in NIH/3T3 wild-type and knockout clones confirming cross-reactivity of the anti-BTBD7 antibody. Images are maximum intensity projections. Experiment courtesy of Shaohe Wang.

## **Determination of the Anti-BTBD7 Cross-Reacting Protein**

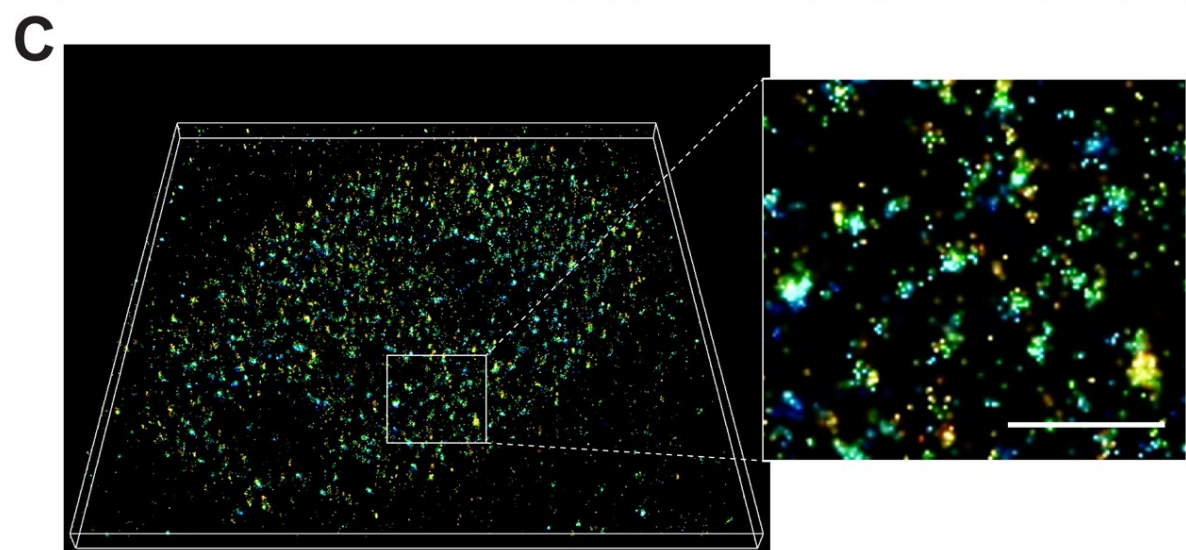
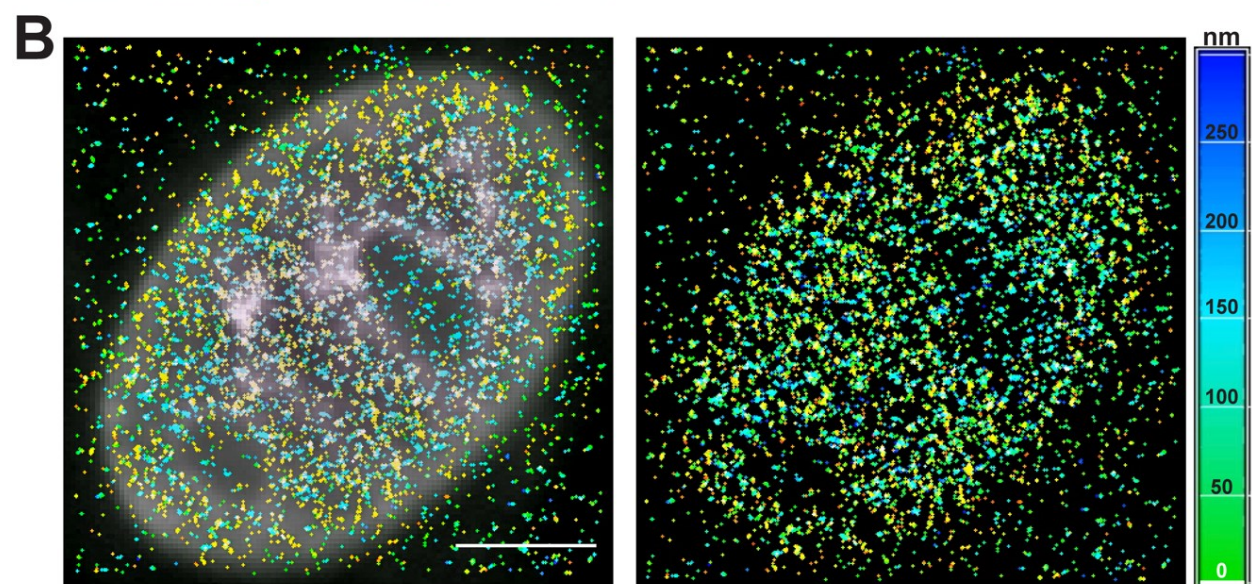
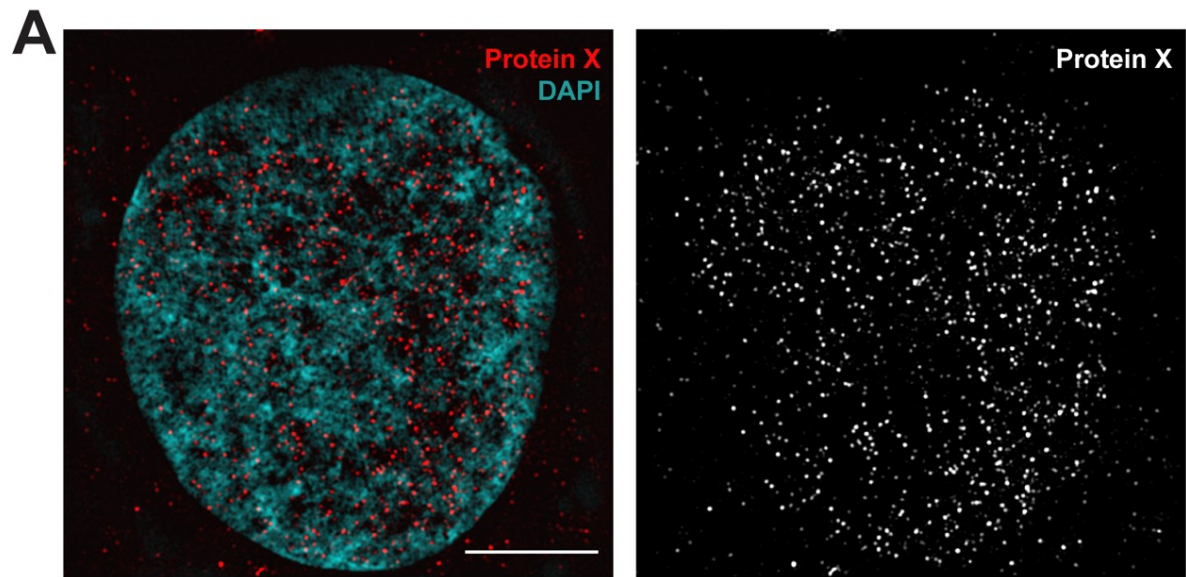
We were now confronted with the unpleasant reality that the results of our experiments that relied heavily on immunofluorescence with this anti-BTBD7 antibody were invalid. Nevertheless, we were intrigued by several properties revealed by the staining patterns of the unknown, cross-reacting protein (hereafter referred to as protein X). First, protein X expression in developing salivary glands displayed a remarkable enrichment in peripheral epithelial cells as early as day E11.5 (Figure 1.14A). This peripheral enrichment is largely absent by day E16.5, suggesting a potential early developmental role for protein X (Figure 1.14B). Second, protein X staining in the cell nucleus displayed a unique pattern of ~1,000 puncta up to ~200 nm in diameter consisting of multiple individual proteins in DAPI sparse regions, as revealed by super-resolution imaging (Figure 1.15A-C). We sought to uncover the true identity of protein X in order to study these intriguing properties.

### **Epitope mapping of the immunogen amino acid sequence**

We first performed a BLAST search using the 104-amino acid sequence of the immunogen used to generate the anti-protein X (anti-BTBD7) antibody. These results were specific to BTBD7 only. Further, relaxed BLAST alignment parameters resulted in too many candidates to be of practical benefit. Thus, we conducted epitope mapping assays of the immunogen sequence to narrow the target region for BLAST alignment. We synthesized four overlapping peptides of 31 amino acids that spanned the entire immunogen sequence and then performed immunofluorescence-based, peptide competition assays (Figure 1.16A-B). Based on these results, we found the epitope



**Figure 1.14** (A) 100X immunofluorescence images of E11.5 submandibular salivary glands from wild-type mice. Epithelial tissues are delineated by E-cadherin in cyan. Protein X staining is in magenta and grayscale. Note the enhanced staining of protein X in peripheral epithelial cells. (B) 100X immunofluorescence images of E16.5 submandibular salivary glands from wild-type mice. Epithelial cells and protein X are stained as in (A). Note the diminished staining of protein X in peripheral epithelial cells. Scale bar = 20  $\mu\text{m}$  for both. Images in (A-B) are from single axial planes taken mid-gland. Experiment courtesy of Shaohe Wang.

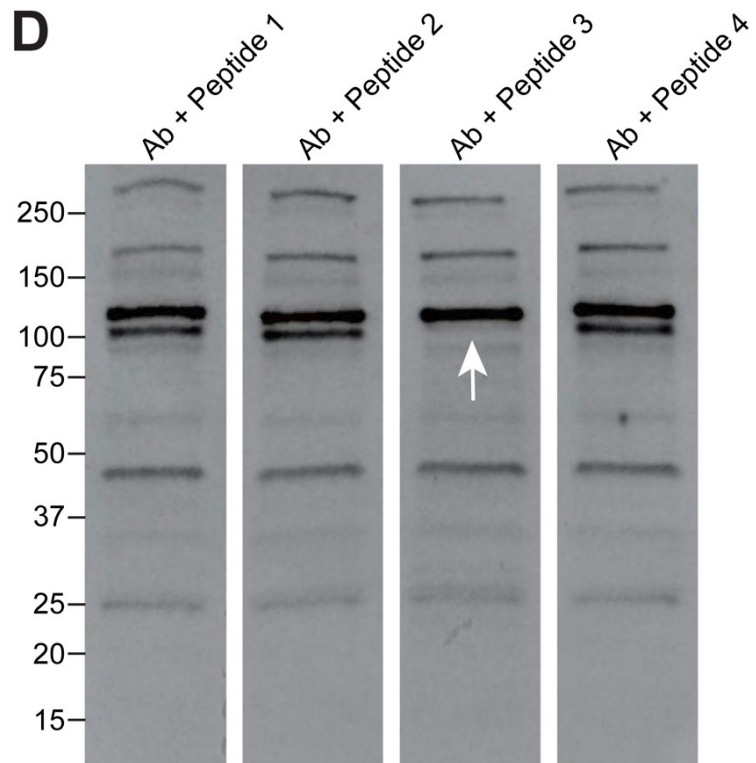
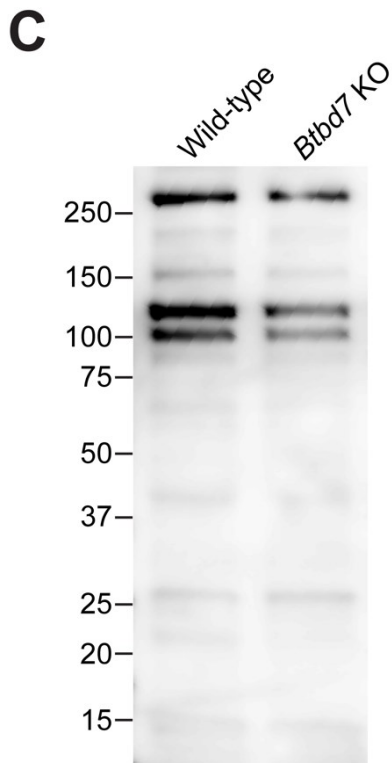
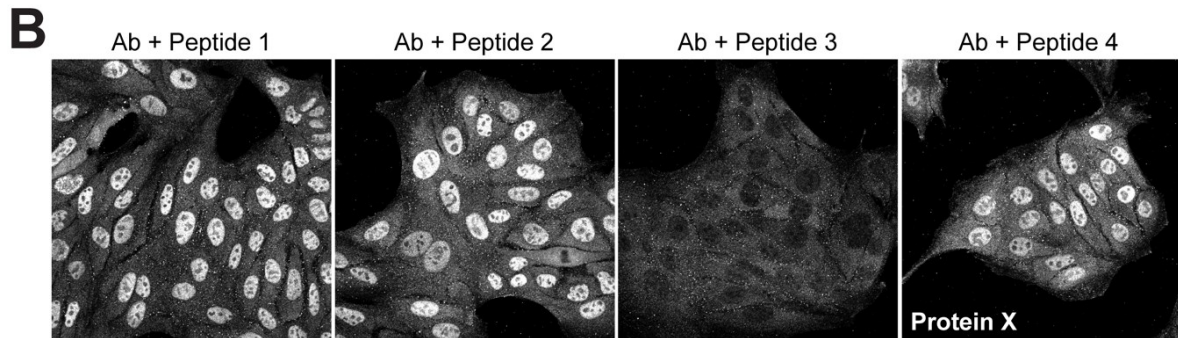
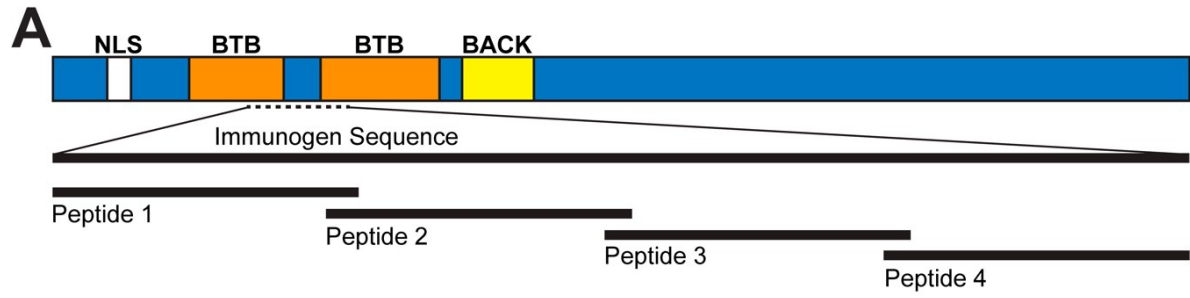


**Figure 1.15** (A) DeltaVision OMX structured illumination super-resolution imaging of protein X in an MDCK.2 nucleus. Image is of a single mid-nuclear axial plane. Scale bar = 5  $\mu\text{m}$ . (B) Nikon N-STORM super-resolution 2D mapping of individual protein X molecules from the middle 600 nm of the nucleus. Colors represent axial depth relative to the focal plane (0 nm). AF647-labeled anti-BTBD7 lot k<sub>1</sub> was used for STORM imaging. Scale bar = 5  $\mu\text{m}$ . (C) Gaussian 3D rendering of the individual protein X molecules from (B). Inset scale bar = 1  $\mu\text{m}$ . Note the large clusters of protein X molecules.

region responsible for the nuclear staining of protein X resides in the 31-amino acid region of peptide 3 depicted in Figure 1.16A (see Materials and Methods for peptide sequence).

Additionally, though not stated as a validated application on the company's antibody product datasheet, we performed western blots using lysates from SIMS wild-type and *Btbd7* knockout cells. We found that probing with the anti-protein X antibody resulted in multiple, prominent bands of varying molecular weight (Figure 1.16C). Notably, there was no difference in the banding pattern between SIMS wild-type and *Btbd7* knockout cells. We then asked if western-blot-based peptide competition assays using the same peptides above would reveal the apparent molecular size of protein X. Indeed, we found that the 31 amino acid peptide that resulted in the loss of nuclear immunofluorescence staining also resulted in the loss of a 105 kD band in western blots (Figure 1.16D).

We then performed extensive BLAST analyses and searched for nuclear-localized candidate proteins. We found 27 candidates ranging in molecular weight from 359 to 48 kD. To account for molecular weight shifts due to post-translational modifications, we limited candidates to those between 85 and 125 kD. These criteria resulted in four primary candidates: KDM4A, PHF8, KDM7A, and TACC1.



**Figure 1.16** (A) Schematic depicting the overlapping peptides used for epitope mapping of the immunogen sequence used to generate the anti-BTBD7 antibody used in this case study. (B) Peptide competition assay results confirming the epitope responsible for the nuclear immunofluorescence staining pattern resides on the 31-amino acid region of peptide 3. (C) Western blot of cell lysates from SIMS wild-type and *Btbd7* knockout cells. Westerns were probed using the same anti-BTBD7 antibody as in (B). Note that the banding pattern is identical in both samples. (D) Peptide competition assay results confirming the epitope for a 105 kD protein resides on the 31-amino acid region of peptide 3.

## Mass spectrometry

In order to determine if the candidate proteins were expressed in our SIMS cells, and if these candidates were post-translationally modified in such a way as to result in an apparent 105 kD molecular weight, we contracted with MS Bioworks to perform mass spectrometry analysis on the 105 kD region excised from a Coomassie-stained gel. We found a total of 187 proteins in our sample. Filtering these results for proteins that localize to the nucleus resulted in 79 candidates. The proteins with the greatest number of peptides detected were ACTN4, TIF1B, ZCCHC8, SNUT1, and HS105. Interestingly, though perhaps not surprising, cross-referencing the 79 mass spectrometry candidates with the 27 from our BLAST analysis resulted in zero matches.

## Immunoprecipitation for protein X

With too many candidates to be of practical benefit, we sought strategies to refine our candidate list. Based on the results of the peptide competition assays, we reasoned that the protein X antibody would work in IP assays with the goal of identifying the IP product with mass spectrometry analysis.

After initial refinement of the IP conditions, we were able to successfully IP protein X (Figure 1.17A). However, our success was short-lived, and repeated attempts

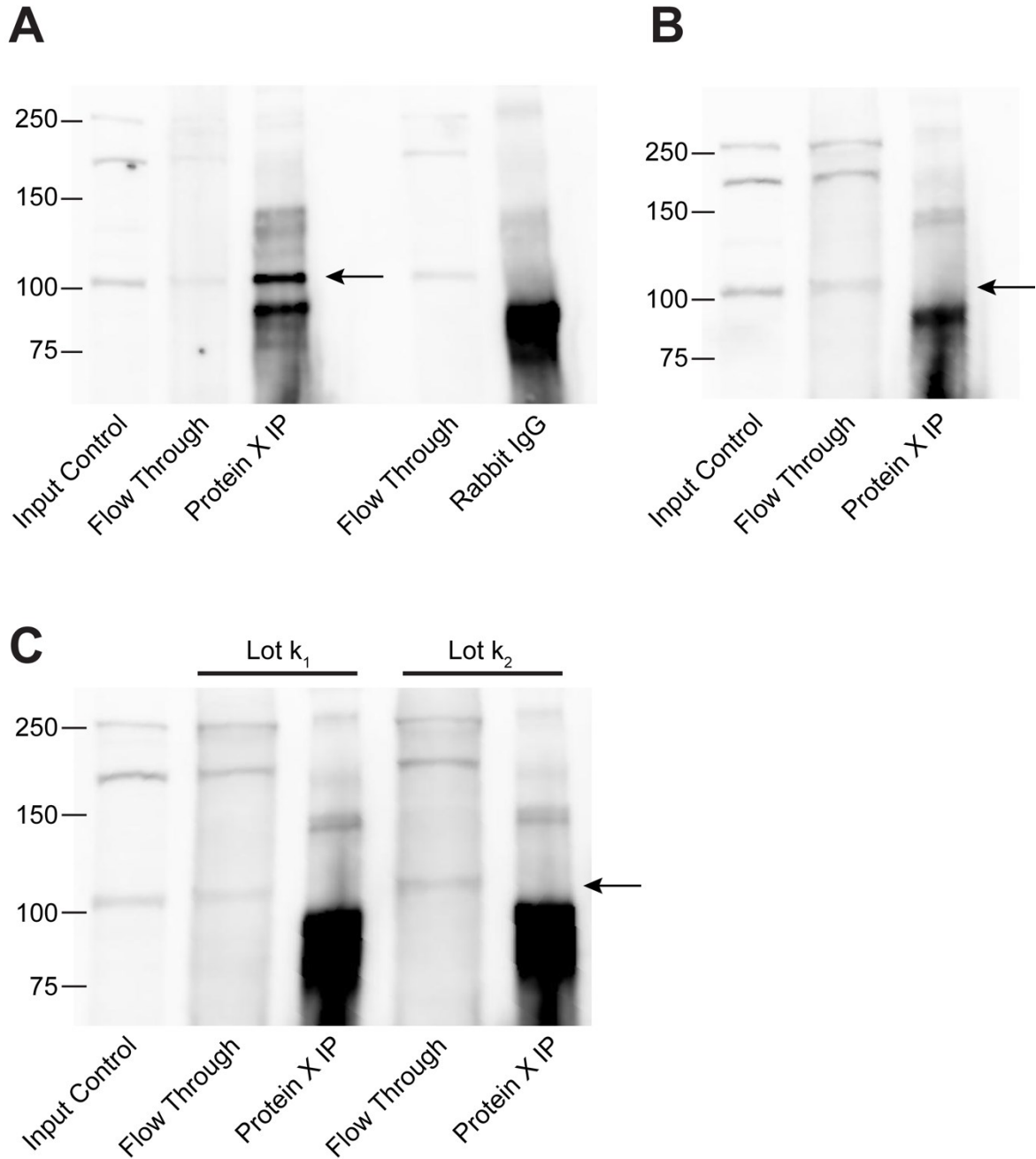


failed before we could perform mass spectrometry on the IP product (Figure 1.17B). After multiple attempts, our supply of protein X antibody lot  $k_1$  was nearing depletion, and we opted to reserve a small supply for future validation experiments. Attempts to IP protein X with new antibody production lot  $k_2$  were also unsuccessful (Figure 1.17C).

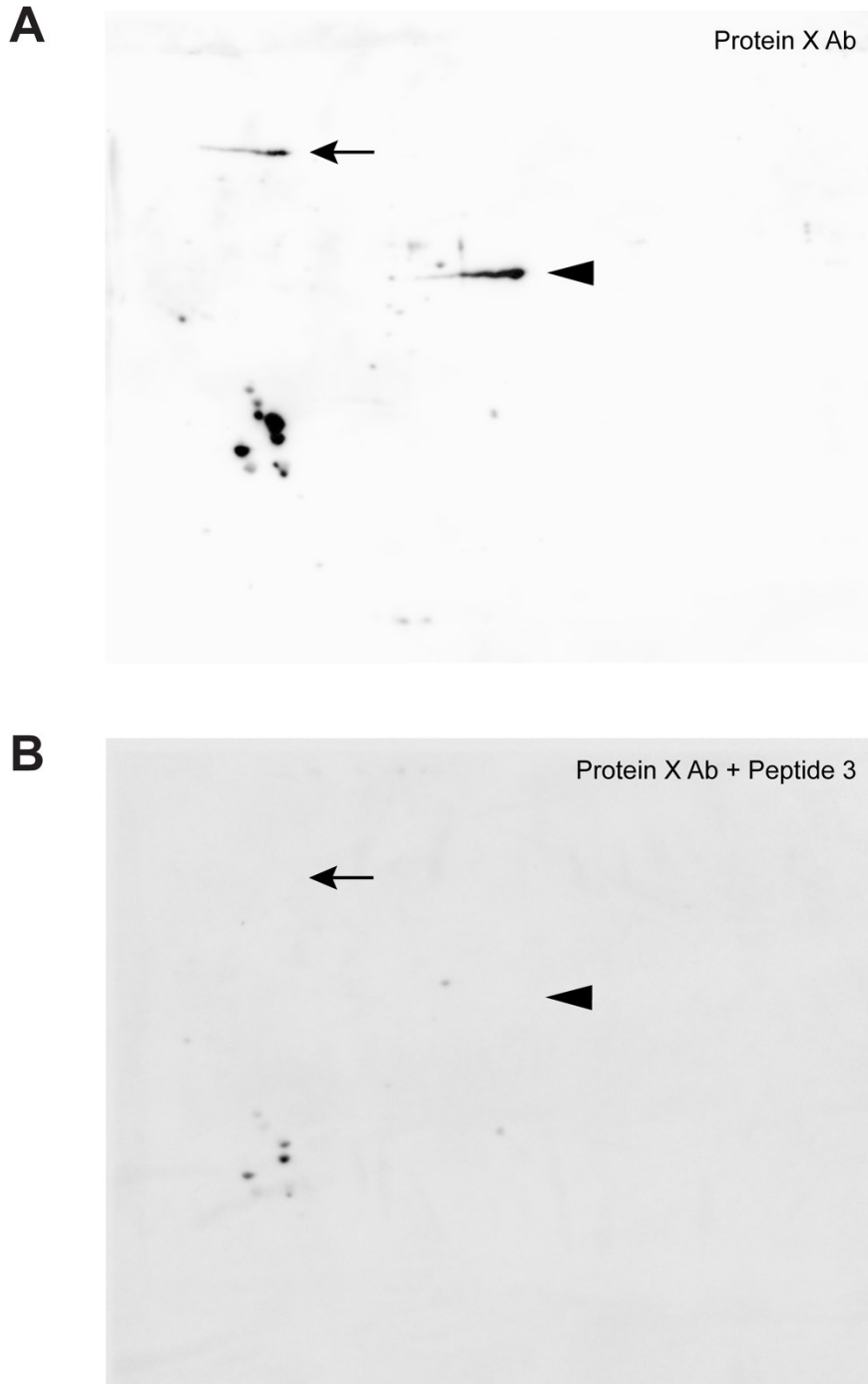
### **Two-dimensional electrophoresis**

We next sought to refine our candidate list using 2D electrophoresis followed by mass spectrometry. We contracted with Kendrick Laboratories to perform the 2D electrophoresis and gel transfer of our SIMS cell lysate. We then performed our peptide competition assays to identify any protein X candidate spots on the 2D membrane. Two highly phosphorylated proteins in regions corresponding to molecular weights of ~105 kD and ~65 kD were chosen for mass spectrometry (Figure 1.18A-B). A single protein corresponding to each region was identified: ZCCHC8 and PDIA3, respectively.

Interestingly, ZCCHC8, a 76 kD protein, is highly phosphorylated and results in a 105 kD band on western blots. In addition, it was among the highest scoring candidates in our initial mass spectrometry results. PDIA3 was not present in our initial mass spectrometry results. Therefore, we chose ZCCHC8 as the primary candidate for validation studies.



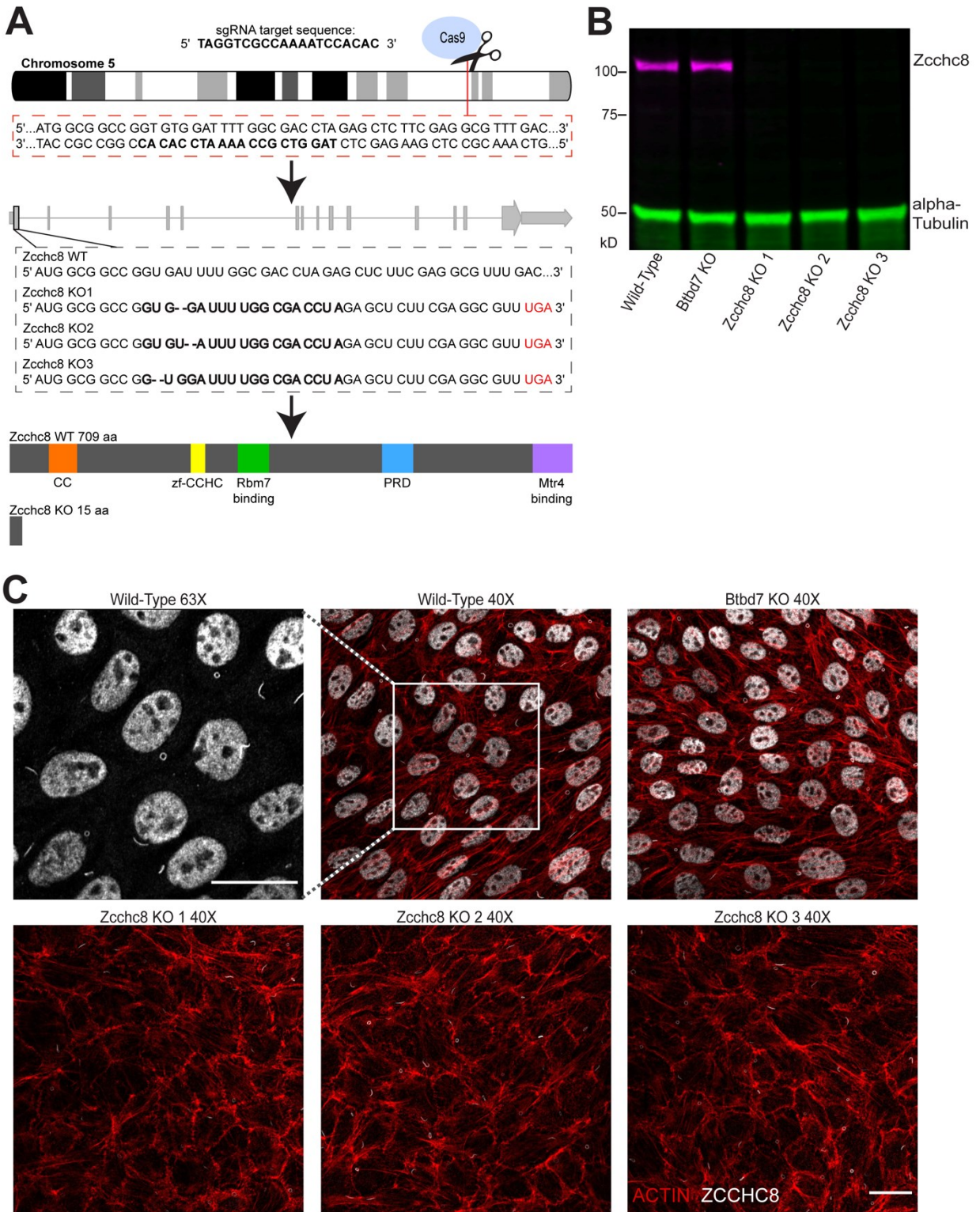
**Figure 1.17** (A) Immunoprecipitation for protein X using the cross-reacting anti-BTBD7 antibody. Arrow marks the 105 kD protein X band. Note the enhanced signal in the IP lane with little residual signal in the flow-through. (B) Failed protein X IP results due to unknown circumstances. The conditions in (A) and (B) are identical. Arrow marks the location of the expected 105 kD protein X band. Note the increased residual signal in the flow-through. (C) Failed protein X IP results using two different production lots of the anti-BTBD7 antibody ( $k_1$  and  $k_2$ ). The conditions in (A) and (C) are identical. Arrow marks the location of the expected 105 kD protein X band. Note the increased residual signal in the flow-through lanes.



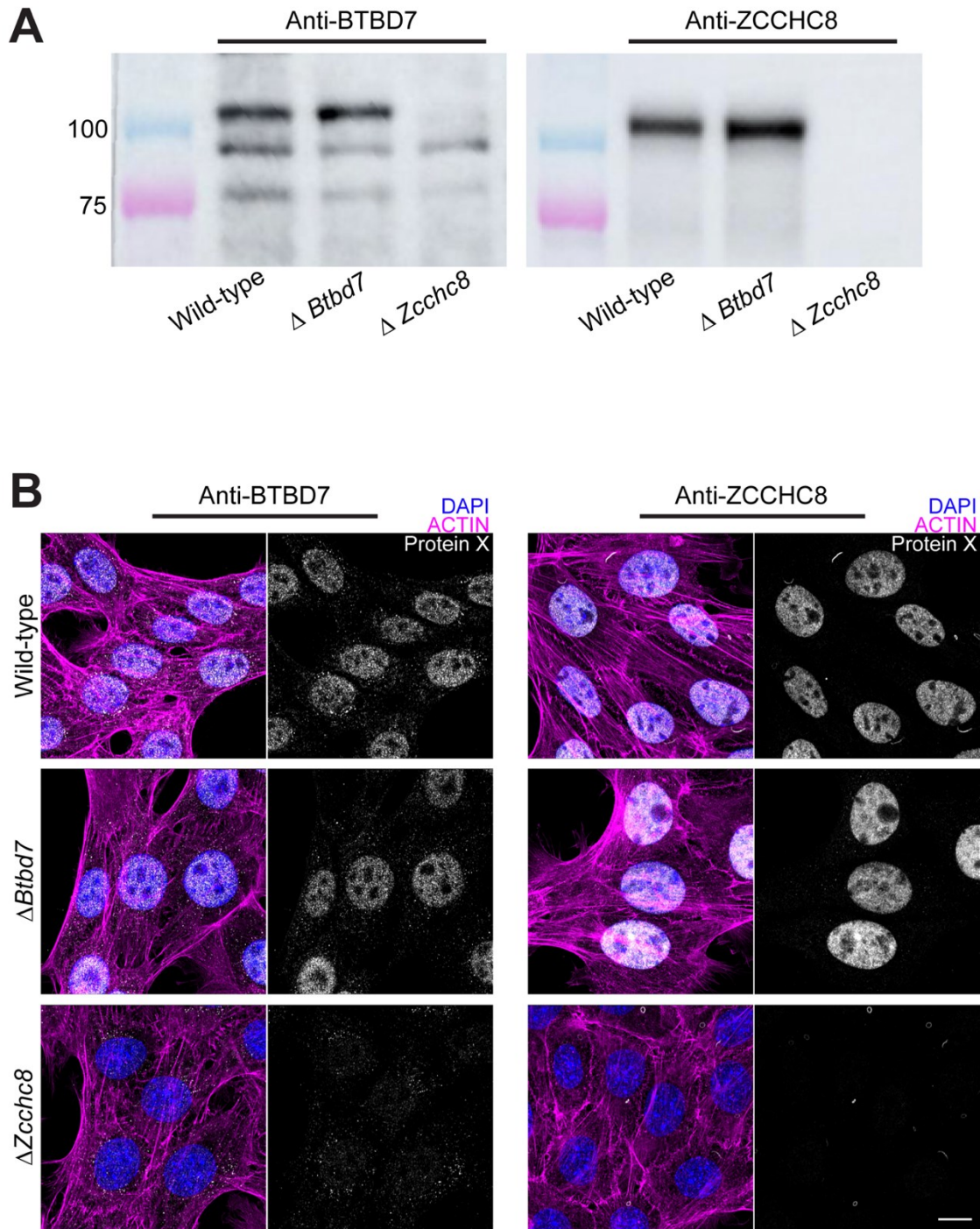
**Figure 1.18** (A) 2D western blot for protein X using the cross-reacting anti-BTBD7 antibody (lot  $k_1$ ). Arrow marks a highly phosphorylated protein of ~105 kD. Arrowhead marks a highly phosphorylated protein of ~65 kD. (B) Peptide competition assay results on a 2D western blot reveal two major candidates for protein X. Note the absence of the ~105 kD spot (arrow) and the ~65 kD spot (arrowhead). The antibody was incubated with 10X excess peptide 3, on which the protein X epitope resides, prior to western blot probing.

## Antibody cross-reactive candidate validation

Considering our experience with previous antibody validation techniques, we opted to use the CRISPR/Cas9 system to knockout *Zcchc8* in the SIMS cell line (Figure 1.19A). We then isolated >20 clonal populations using standard single-cell cloning techniques. After western blotting for ZCCHC8, we randomly selected three individual cell lines from those with depleted ZCCHC8 expression. Genetic sequencing confirmed differing two-base-pair deletions near the CRISPR/Cas9 target sequence that resulted in newly generated stop codons shortly downstream from the deletion site in each clone (Figure 1.19A). Follow-up western blotting and immunofluorescence analyses confirmed the absence of ZCCHC8 protein expression in the selected clones (Figure 1.19B-C). After successful knockout of *Zcchc8* in SIMS cells, we then confirmed ZCCHC8 as the identity of protein X through comparison western blotting and immunofluorescence in SIMS *Btbd7* and *Zcchc8* knockout cells (Figure 1.20).



**Figure 1.19** (A) Schematic depicting the CRISPR/Cas9 strategy for disrupting *Zcchc8* in SIMS mouse salivary gland cells. Single-cell cloning produced three separate knockout clones with two-base-pair deletions ~12-15 bp downstream of the TSS. The resulting frameshift produced a translational stop codon at amino acid position 16. (B) Western blot confirmation of ZCCHC8 ablation in three separate clones. Western blots were probed with a rabbit, polyclonal anti-ZCCHC8 antibody from Proteintech and a mouse, monoclonal anti- $\alpha$ -Tubulin antibody from MilliporeSigma. (C) Immunofluorescence confirmation of ZCCHC8 ablation in three separate clones. The same anti-ZCCHC8 antibody as in (B) was used. Note the presence of antibody cross-reactivity to primary cilia before and after *Zcchc8* knockout. Actin fibers were stained with rhodamine-labeled phalloidin from Thermo Fisher. Scale bar = 20 $\mu$ m.



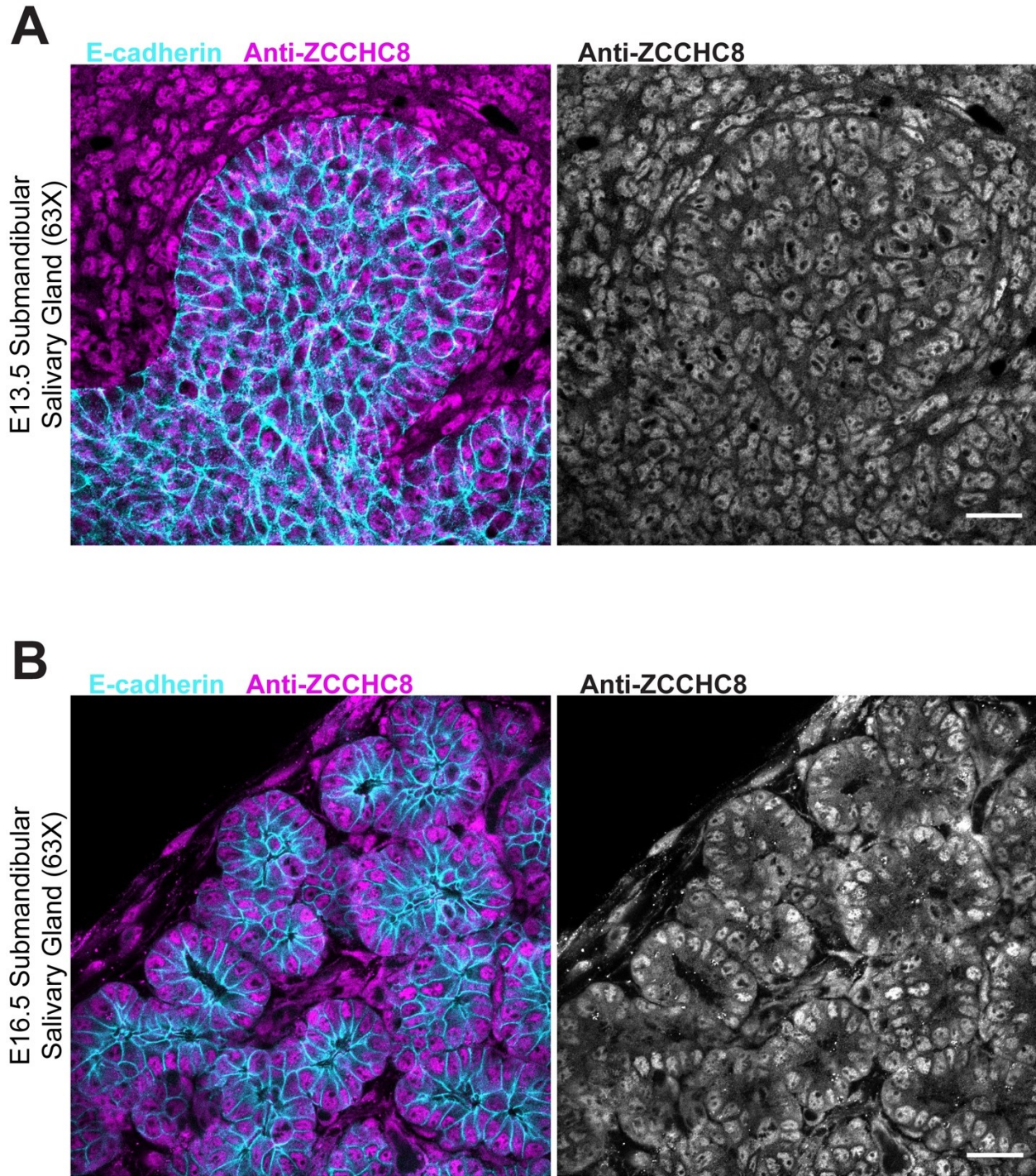
**Figure 1.20** (A) Western blots confirming the identity of protein X as ZCCHC8. SIMS wild-type, *Btd7*, and *Zcchc8* knockout cell lysates were transferred to nitrocellulose and probed with either the anti-BTBD7 lot k<sub>1</sub> antibody (left) or an anti-ZCCHC8 (Proteintech, 23374-1-AP) antibody (right). Note the absence of the 105 kD band in the *Zcchc8* knockout lane of both blots. (B) Immunofluorescence confirming the identity of protein X as ZCCHC8. SIMS wild-type, *Btd7*, and *Zcchc8* knockout cells were immunostained with same antibodies as in (A). Note the absence of nuclear staining in the *Zcchc8* knockout cells probed with both antibodies. Images are maximum intensity projections. Scale bar = 5μm.

## **ZCCHC8 expression in salivary glands and cell nuclei**

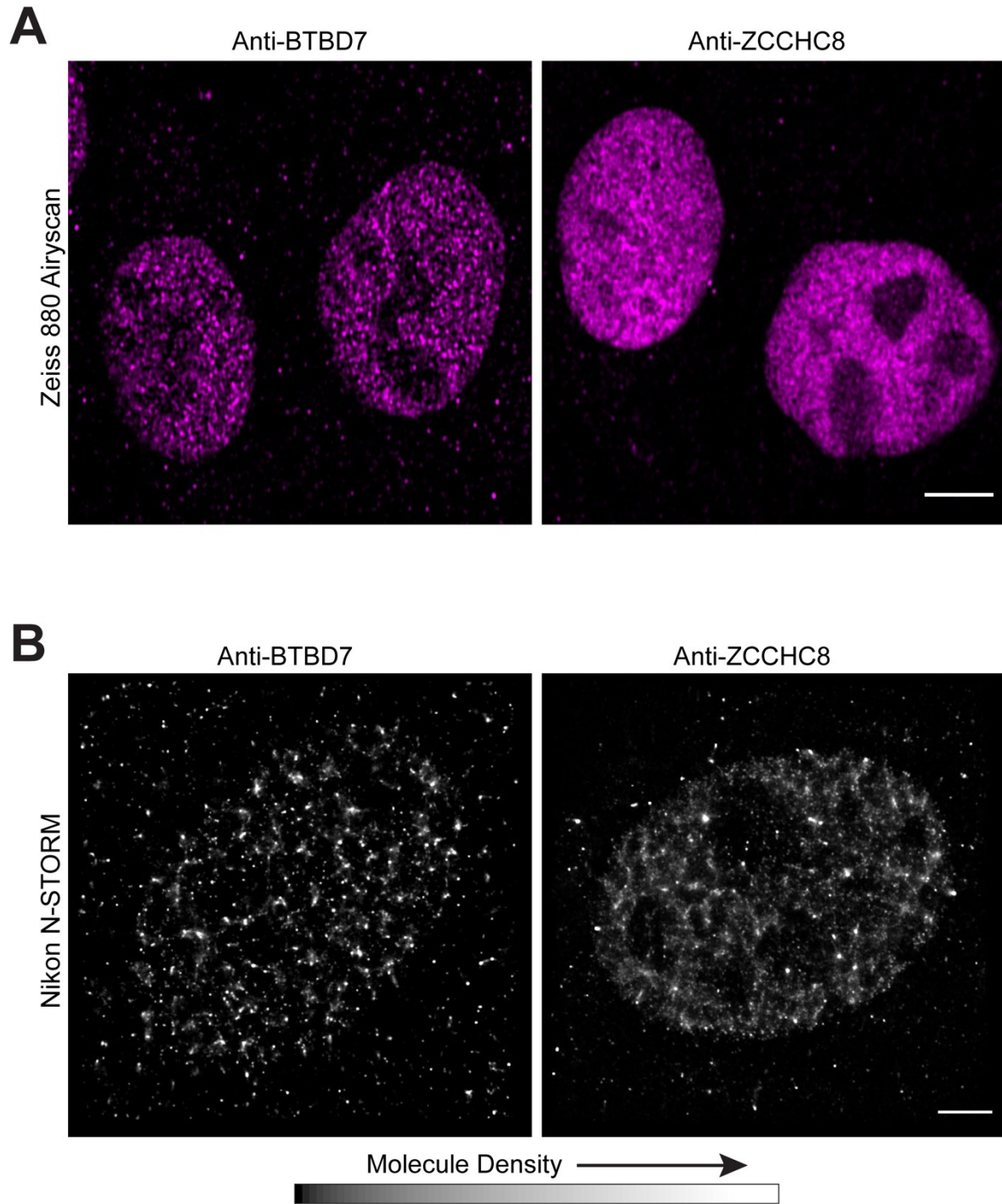
Having discovered the identity of protein X as that of ZCCHC8, we asked if the intriguing expression pattern we had originally observed would be recapitulated with antibodies intended explicitly for ZCCHC8. We first examined immunofluorescence staining patterns in developing salivary glands using an anti-ZCCHC8 antibody purchased from Proteintech (23374-1-AP). ZCCHC8 staining of E13.5 submandibular salivary glands displayed primarily nuclear and uniform staining throughout the epithelium and mesenchyme (Figure 1.21A) and remained consistent in E16.5 glands (Figure 1.21B). In contrast, original staining with the cross-reacting, anti-BTBD7 lot k<sub>0</sub> antibody displayed a remarkable enrichment in peripheral epithelial cells with large cytoplasmic granules present throughout (Figure 1.4A-B, 1.14A). Anti-BTBD7 staining of epithelial cells in E16.5 glands was largely absent (Figure 1.4B).

We also noticed that ZCCHC8 immunofluorescence staining of cell nuclei appeared more diffuse and less punctate. Airyscan imaging of MDCK.2 nuclei stained with the cross-reactive anti-BTBD7 lot k<sub>1</sub> antibody revealed distinct, large puncta located primarily throughout the nucleoplasm and, to a lesser extent, within the nucleolus (Figure 1.22A). Conversely, staining with an anti-ZCCHC8 antibody resulted in a dense and diffuse pattern with fewer apparent puncta (Figure 1.22A). N-STORM super-resolution imaging confirmed these results (Figure 1.22B). Particle analysis of the images in Figure 1.22B revealed 116 puncta within the nucleus stained with anti-BTBD7 lot k<sub>1</sub> and only 20 puncta within the nucleus stained with anti-ZCCHC8.





**Figure 1.21** (A) 63X immunofluorescence images of E13.5 submandibular salivary glands from wild-type mice. Epithelial tissues are delineated by E-cadherin in cyan. ZCCHC8 staining is in magenta and grayscale. (B) 63X immunofluorescence images of E16.5 submandibular salivary glands from wild-type mice. Epithelial cells and ZCCHC8 are stained as in (A). Note the diminished staining of protein X in peripheral epithelial cells at this later stage. Scale bar = 20  $\mu$ m for both. Images in (A-B) are from single axial planes taken mid-gland.



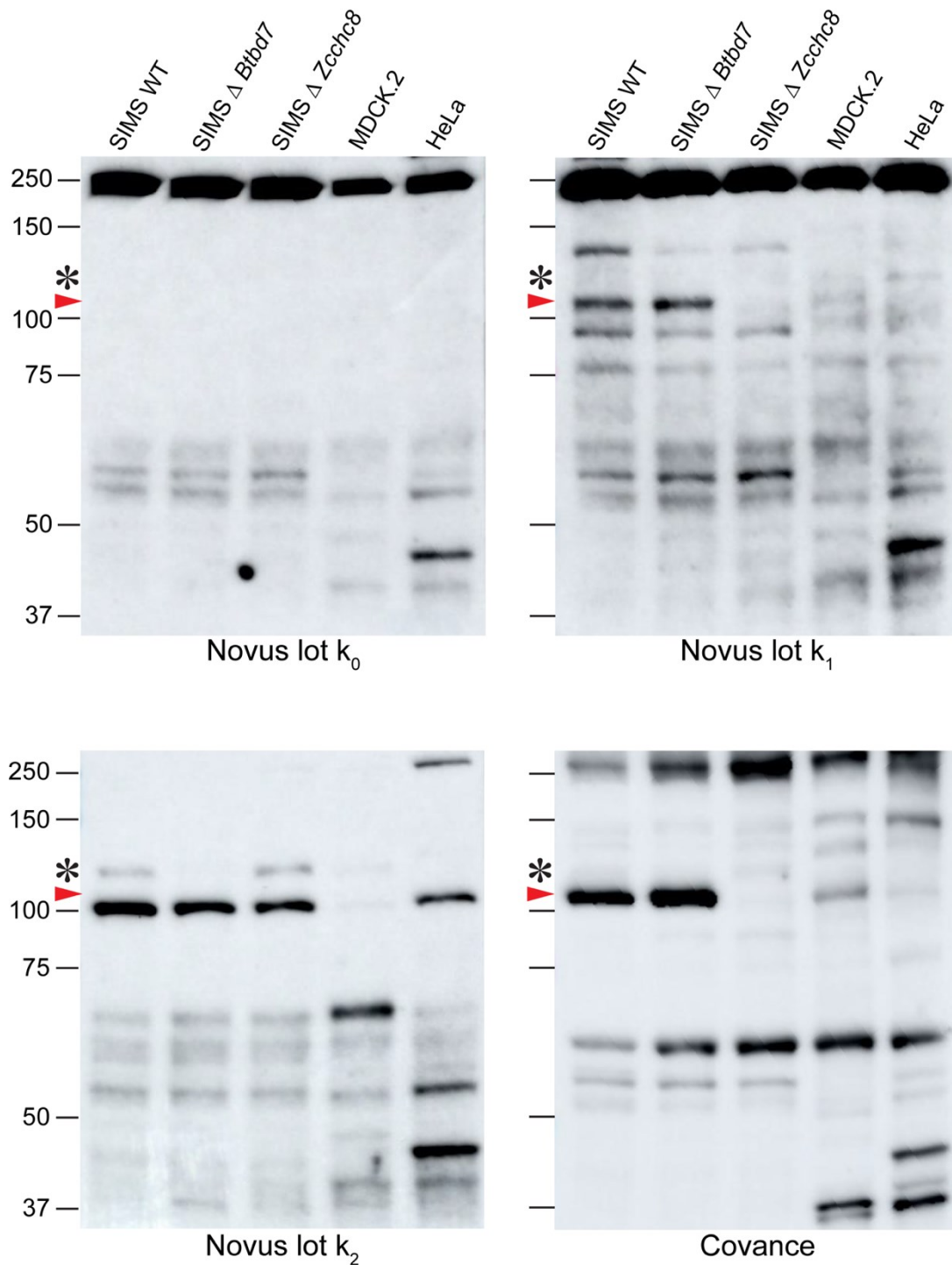
**Figure 1.22** (A) Zeiss 880 immunofluorescence Airyscan images of MDCK.2 cell nuclei immunostained with anti-BTBD7 (lot  $k_1$ ) or anti-ZCCHC8. Anti-BTBD7 results in large, distinct puncta throughout the nucleus. Anti-ZCCHC8 results in a more diffuse pattern with fewer puncta. Images are of a single mid-nuclear axial plane. Scale bar = 5  $\mu\text{m}$ . (B) Nikon N-STORM super-resolution molecular density mappings of MDCK.2 nuclei immunostained with AF647-labeled anti-BTBD7 (lot  $k_1$ ) or anti-ZCCHC8. Note the more diffuse staining with fewer puncta in the nucleus immunostained with anti-ZCCHC8. Scale bar = 2.5  $\mu\text{m}$ .

## Non-Specificity in a Panel of Antibodies

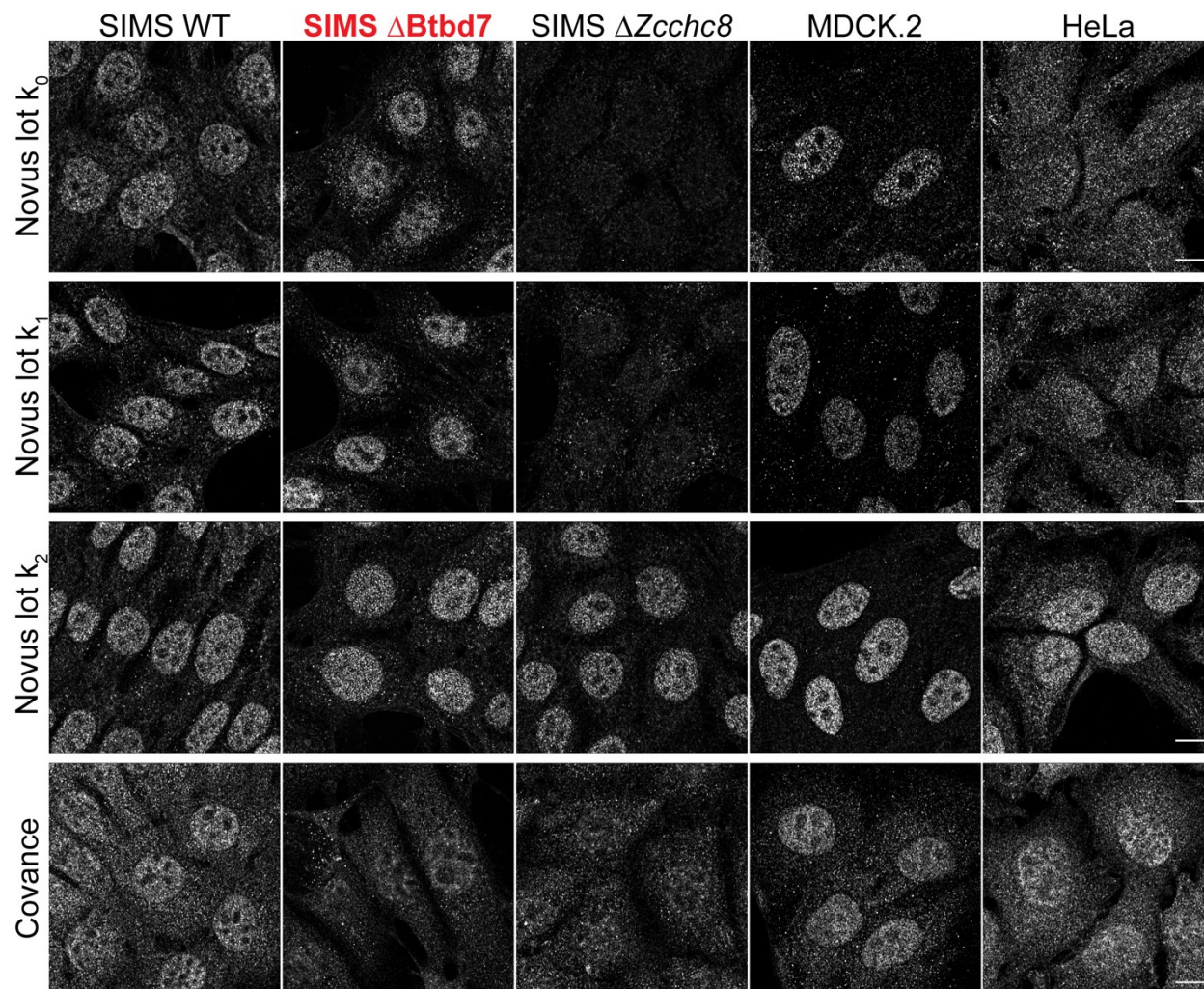
### Direct western blot and immunofluorescence comparison

During our validation efforts, we nearly exhausted our remaining supply of anti-BTBD7 lot  $k_1$ . We acquired a new production lot of anti-BTBD7 (lot  $k_2$ ) and performed comparison immunofluorescence and western blot experiments against lot  $k_1$ . Our initial immunofluorescence staining appeared indistinguishable. However, western blots showed a molecular weight shift in the expected bands. Alarmed at the implications of such lot-lot-lot variability, we then contracted with Covance to produce a rabbit, polyclonal antibody using the BTBD7 31-amino acid peptide sequence we previously identified as carrying the epitope for protein X/ZCCHC8. We then performed comparison immunofluorescence and western blot analysis using the three lots of anti-BTBD7 and our Covance antibody in mouse, canine, and human cell lines. We found that anti-BTBD7 lot  $k_0$  and  $k_1$  were indistinguishable in immunofluorescence but differed greatly in western blots. In western blots (Figure 1.23), both lots of antibody detected multiple nonspecific proteins, but lot  $k_0$  failed to detect either a 105 kD (ZCCHC8) or 126 kD (BTBD7) band. Lot  $k_1$  detected a 105 kD band that corresponded to ZCCHC8 but not a 126 kD band. Conversely, lot  $k_2$  produced a faint band at 126 kD that corresponded to BTBD7 and another distinct band at ~101 kD that did not correspond to ZCCHC8. Further, our purified Covance antibody also produced multiple non-specific bands and a prominent 105 kD band that correlated with ZCCHC8.

Immunofluorescence staining (Figure 1.24) with anti-BTBD7 lots  $k_0$ ,  $k_1$ , and  $k_2$  appear nearly indistinguishable in SIMS wild-type and *Btbd7* knockout cells. As demonstrated previously, lots  $k_0$  and  $k_1$  do not produce the characteristic nuclear



**Figure 1.23** Western blots using different production lots of the same anti-BTBD7 antibody (Novus NBP2-14364) or a custom antibody (Covance) showing non-specific staining in a panel of cell lines. The custom antibody from Covance was generated using the BTBD7 31-amino acid peptide used in our peptide competition assays as the immunogen. Asterisks mark the position of BTBD7. Red arrowheads mark the position of ZCCHC8. Note the extreme lot-to-lot variability of the Novus antibody.



**Figure 1.24** Immunofluorescence using different production lots of the same anti-BTBD7 antibody (Novus NBP2-14364) or a custom antibody (Covance) showing non-specific staining in a panel of cell lines. The custom antibody from Covance was generated using the BTBD7 31-amino acid peptide used in our peptide competition assays as the immunogen. The *Btbd7* knockout cells are denoted in red. Note all production lots of the Novus anti-BTBD7 antibody produce a strong, cross-reacting nuclear pattern in *Btbd7* knockout cells. This nuclear staining is absent in *Zcchc8* knockout cells stained with anti-BTBD7 lots  $k_0$  and  $k_1$  but not  $k_2$ . The enhanced nuclear pattern produced with the Covance antibody is diminished in both *Btbd7* and *Zcchc8* knockout cells. Images are maximum intensity projections. Scale bar = 5  $\mu$ m.

staining pattern in *Zcchc8* knockout cells. Strangely, lot  $k_2$  stains the nucleus in an identical manner in both SIMS wild-type and *Zcchc8* knockout cells. Further, lot  $k_2$  showed enhanced nuclear staining in MDCK.2 and HeLa cells compared to lots  $k_0$  and  $k_1$ .

Immunofluorescence staining with our Covance antibody produced a characteristic enhanced nuclear staining pattern in SIMS wild-type, MDCK.2, and HeLa cells. This staining was diminished, though not eliminated, in both *Btbd7* and *Zcchc8* knockout SIMS cells (Figure 1.24).

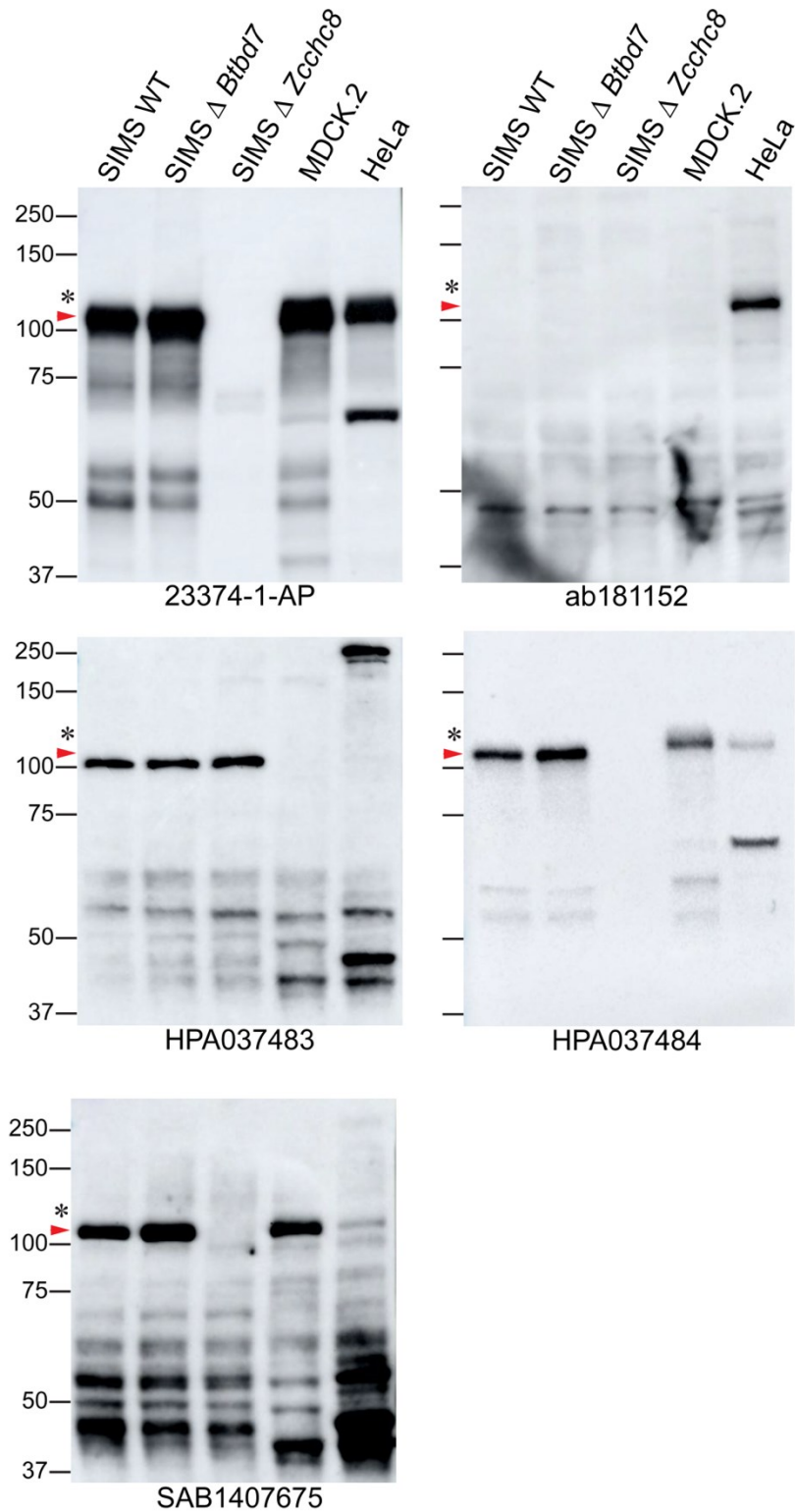
We noticed that the anti-ZCCHC8 antibody that we used also seemed to suffer from non-specific staining artifacts. In particular, this antibody appears to stain primary cilia in both wild-type and *Zcchc8* knockout SIMS cells (Figure 1.19C, Figure 1.20B). While this artifact can be easily differentiated from the expected nuclear staining, we, nevertheless, sought to find an antibody that did not suffer from such cross-reactivity.

After an exhaustive search, we identified a total of five anti-ZCCHC8 antibodies that were produced using unique immunogens. We purchased these antibodies from Proteintech (23374-1-AP, rabbit polyclonal), Abcam (ab181152, rabbit monoclonal), and MilliporeSigma (HPA037483, rabbit polyclonal; HPA037484 rabbit polyclonal; SAB1407675, mouse polyclonal). We then performed comparison immunofluorescence and western blot analysis using these antibodies. In western blots (Figure 1.25), we found three of these antibodies produced distinct bands at 105 kD that corresponded to ZCCHC8 in mouse, canine, and human cells. The Abcam rabbit monoclonal antibody appeared to detect only the human form of ZCCHC8. All of the antibodies displayed non-specific cross-reactivity.

Immunofluorescence staining with these antibodies produced a range of results (Figure 1.26). In general, all of the antibodies displayed some form of cross-reactivity. The Proteintech (23374-1-AP) antibody appeared to cross-react with primary cilia but otherwise displayed little cross-reactivity. Interestingly, the Abcam rabbit monoclonal antibody (ab181152) appeared to produce a nuclear staining pattern due to BTBD7 rather than ZCCHC8. The rabbit polyclonal antibodies (HPA037483 and HPA037484) from MilliporeSigma gave similar results to each other, and both appeared to cross-react with an undetermined cytoplasmic protein. Of interest, and at the time of this writing, these two antibodies are featured in the Human Protein Atlas as confirmation of nuclear localization for ZCCHC8 in a panel of human cell lines. Our results with HPA037484 in HeLa cells differ from the human A-431, U-2 OS, and U-251 cells used in the Human Protein Atlas, suggesting possible lot-to-lot variation with this antibody. The Sigma mouse polyclonal antibody (SAB1407675) displayed diminished staining in *Zcchc8* knockout cells though the overall pattern lacked the nuclear enrichment characteristic of the other antibodies.

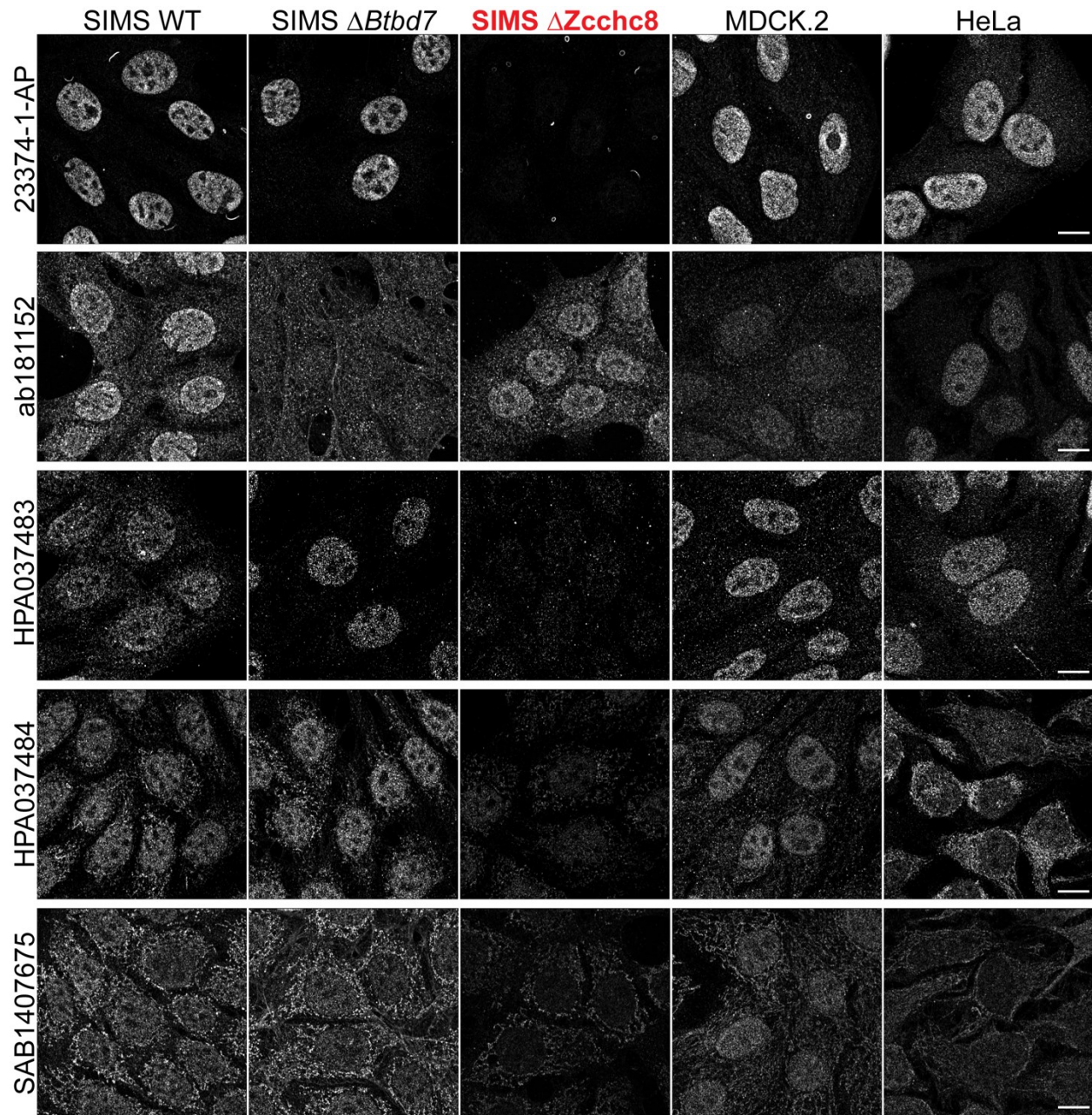
### **Human protein atlas comparison**

Over the course of our investigation for the identity of protein X, the Human Protein Atlas validated and confirmed that BTBD7 was localized to focal adhesions in human RT4 cells using the anti-BTBD7 antibody that we used extensively (Figure 1.27A).



**Figure 1.25** Western blots using different anti-ZCCHC8 antibodies showing non-specific staining in a panel of cell lines. Asterisks mark the position of BTBD7. Red arrowheads mark the position of ZCCHC8. Antibodies were purchased from Proteintech (23374-1-AP), Abcam (ab181152), and MilliporeSigma (HPA037483, HPA037484, SAB1407675).



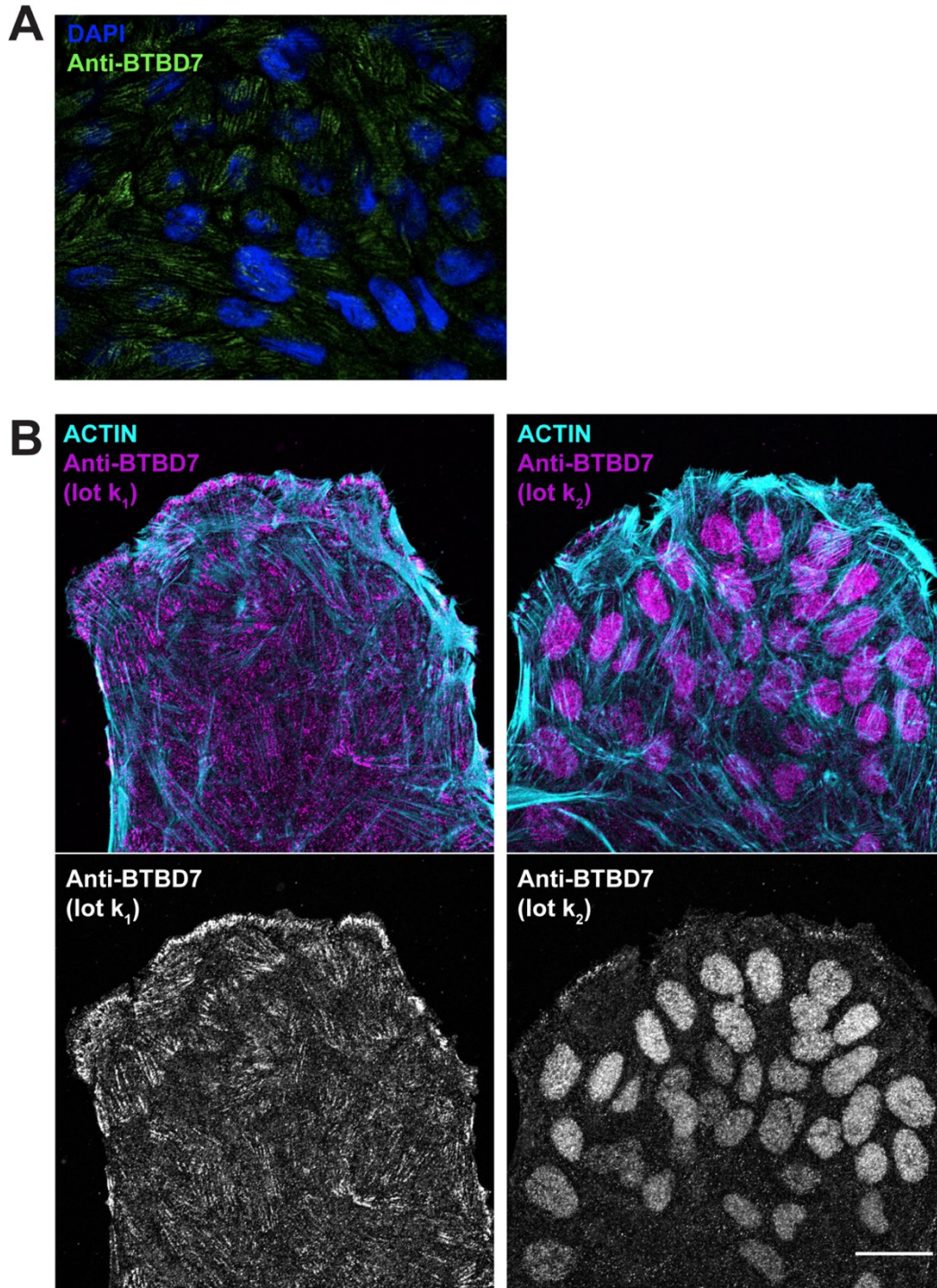


**Figure 1.26** Immunofluorescence using different anti-ZCCHC8 antibodies showing non-specific staining in a panel of cell lines. The *Zcchc8* knockout cells are denoted in red. Note antibody ab181152 apparently stains BTBD7 in the nucleus but not ZCCHC8. Antibodies were purchased from Proteintech (23374-1-AP), Abcam (ab181152), and MilliporeSigma (HPA037483, HPA037484, SAB1407675). Images are maximum intensity projections. Scale bar = 5  $\mu$ m.

We noticed that the staining pattern in human HeLa cells often displayed a faint but perceptible focal adhesion pattern with anti-BTBD7 lots  $k_0$  and  $k_1$ . However, lot  $k_2$  immunofluorescence displayed an enhanced nuclear staining not previously observed. We wondered if this discrepancy would be evident in the RT4 cells featured in the Human Protein Atlas. We purchased RT4 cells from ATCC and performed comparison immunofluorescence using anti-BTBD7 lots  $k_1$  and  $k_2$ . As expected, the results revealed a high, lot-to-lot variability with lot  $k_2$  displaying high nuclear enrichment and little focal adhesion staining (Figure 1.27B).

### **Sequence Alignment Between ZCCHC8 and BTBD7**

Our results from the immunofluorescence comparison of anti-BTBD7 and anti-ZCCHC8 antibodies suggest that ZCCHC8 and BTBD7 may share a common epitope. We asked if an amino acid sequence alignment of the antigen regions would reveal a shared, continuous epitope. Sequence alignment between the BTBD7 104-amino acid immunogen sequence and the ZCCHC8 353-amino acid immunogen sequence (C-terminal region 355-707) found no sequence similarities despite varying the alignment algorithm parameters. Attempts at sequence alignment using the narrowed BTBD7 31-amino acid epitope region and the ZCCHC8 immunogen sequence also produced no similarities. Finally, we aligned the entire amino acid sequences of both proteins and found a partial match between amino acids 7-28 in ZCCHC8 and 745-766 in BTBD7, regions well outside the immunogen sequences used to produce the respective antibodies (Figure 1.28).



**Figure 1.27** (A) Screenshot taken from the Human Protein Atlas of immunofluorescence staining in RT4 cells using the anti-BTBD7 antibody used in this case study (production lot unknown). Note the strong focal adhesion-like pattern and lack of nuclear staining. (B) Comparison immunofluorescence staining in RT4 cells using two production lots ( $k_1$  and  $k_2$ ). Staining with lot  $k_1$  recapitulates the Human Protein Atlas results while staining with lot  $k_2$  shows extreme cross-reactivity with an unknown nuclear protein. Images are maximum intensity projections. Scale bar = 25  $\mu\text{m}$ .

Btbd7 Length: 1130  
Range 745 to 766

Zcchc8 Length: 709  
Range: 7 to 28

Score:25.0 bits(53), Expect:0.020,  
Method:Compositional matrix adjust.,  
Identities:11/22(50%), Positives:12/22(54%), Gaps:0/22(0%)

BTBD7	745	FTDLDSFVAFHPPLPPPPPPYH	766
		F DL+ F AF PP P P H	
ZCCHC8	7	FGDLELFEAFDPPEESTPKPVH	28

**Figure 1.28** Protein sequence alignment of full-length BTBD7 and ZCCHC8 reveals only a small region of similarity. This aligned region falls outside of the immunogen sequences used to produce antibodies used in this case study.

# Discussion

## Antibody Validation

This case study provides a broad examination of the potential pitfalls of antibody usage in laboratory applications. Our results show that reliance on traditional validation techniques is not always sufficient to ascertain an antibody's target specificity and selectivity.

We employed peptide competition assays as part of our validation strategy. But, can peptide competition assays reveal cross-reactivity? Perhaps, if the conditions are just right. Competition assays typically employ antibody pre-incubation with large excess of the peptide/antigen used to generate the antibody. If the antibody has a high binding affinity to this antigen (as one would expect), then most, if not all, of the antibody will bind preferentially to the excessive antigen and cross-reactivity will not be observed. Ultimately, then, these competition assays assess an antibody's preference to bind to the antigen used to generate the antibody.

We also employed a target over-expression system as part of our validation. Once again, this system relies on an excess of the target, and so binding affinity is at the core of the results. However, rather than assessing the antibody's ability to bind to the antigen against which it was raised, this assay determines the antibody's ability to bind to the protein/target containing the antigen. This is an important added aspect, yet it does not fully account for possible cross-reactivity. Assessing cross-reactivity is possible under the ideal conditions but, for many of the same reasons as with peptide competition assays, not guaranteed.

As another validation strategy, we employed siRNA knockdown of target gene expression. Again, this strategy is helpful but insufficient to deduce specificity. Cells expressing various siRNA constructs showed reduced immunofluorescence staining in our experiments. What of the remaining immunofluorescence staining? Is this remaining immunofluorescence due to existing target protein or non-specific staining? Knockdown validation assays can show relative target specificity but may be insufficient to show absolute or even adequate specificity.

Perhaps the gold standard of validation assays is that of gene knockout to wild-type comparison. Indeed, the strength of our conviction in the specificity of the anti-BTBD7 antibody was due, almost entirely, to the presence and absence of immunofluorescence staining in wild-type and *Btbd7* knockout mouse tissues. Here, again, there is an important caveat in this validation: *Btbd7* knockout in mice resulted in embryonic lethality. The tissues used in our immunofluorescence experiments were from embryos with extreme developmental defects. It is likely that these embryos suffered severe disruption to the proteome and that potential antibody cross-reacting proteins were degraded, improperly processed, or not expressed at all and therefore undetectable through immunofluorescence. Though the practicality of generating a knockout mouse strictly for antibody validation purposes is unreasonable, should this approach be available, it may be best reserved for use with fully developed, adult mice. This, of course, limits the targets that can be successfully validated as not all gene knockouts result in viable offspring.

Though the results of experiments we conducted using the anti-BTBD7 antibody were often unpredicted, confused, and difficult to interpret, it was not until the need for

CRISPR/Cas9 knockout cell culture experiments arose that we were forced to acknowledge the off-target reactivity of this antibody. Here again, however, our conviction in the knockout mouse results was so strong that we dismissed the initial CRISPR/Cas9 knockout cell line as an incomplete knockout and only after a second, independent NIH/3T3 knockout cell line was created did we fully realize the extent of cross-reactivity.

It is tempting to think of gene knockout in cell lines as the ultimate antibody validation effort. In large part, this may be true. But, this strategy is not without difficulty. Some genes are essential for cell survival and proliferation, and therefore complete knockout may not be achievable. Many genes are part of extensive pathways with products of one gene or pathway inducing expression of many others. An unlucky researcher may unknowingly have ablated the downstream expression of cross-reacting products after deleting the target gene. As unlikely as these scenarios may be, knowing of the possibilities precludes gene knockout as the end-all, be-all validation strategy. To be successful, validation strategies must work in concert with one another and must take into account careful consideration of the results.

## **Identification of Cross-Reacting Proteins**

Often, understanding that antibody cross-reactivity occurs is sufficient to eliminate, or retain, a particular antibody for a desired application. Occasionally, identifying cross-reacting proteins may be desirable. Our work provides a framework for such identification.

The first requirement for identifying cross-reacting proteins is a candidate list. In the absence of experimental evidence for particular candidates, a mass spectrometry approach is advisable. Thanks to large, online repositories of proteome informatics, such as UniProt or NCBI, candidate lists can be easily filtered based on properties like cell line expression or subcellular localization. One can seek to validate these candidates through “brute force” or “trial and error” methods but, often, with further refinement of the candidate list, much effort can be saved.

Ideally, candidates are generated in a manner that limits the initial possibilities. Immunoprecipitation (if possible) or 2D-gel electrophoresis followed by mass spectrometry are excellent methods for limiting candidates. However, IP may not successfully capture sufficient quantities of protein for detection by mass spectrometry. If the cross-reacting protein of interest displays subcellular compartmentalization, then subcellular fractionation may successfully limit candidates in subsequent mass spectrometry.

Sequence alignment, in some instances, may also identify potential candidates. Interestingly, the first mention of *Zcchc8* in the literature came as the result of one laboratory’s undertaking to uncover conflicting immunofluorescence results of subnuclear localization by c-Myc in HeLa cells (41). Gustafson and coworkers noticed an apparent paradox in subnuclear localization of c-Myc to the nucleolus when performing immunofluorescence with an anti-pT58 c-Myc antibody. By carefully examining the antibody staining patterns in western blots and immunofluorescence assays, Gustafson et al., identified at least two cross-reacting proteins. Using partially purified nucleoli followed by mass spectrometry, they identified the major nucleolar



protein to be Ebp2. Subsequently, they performed IP followed by mass spectrometry and identified Zcchc8 as the other non-nucleolar, cross-reacting protein of interest. In this instance, c-Myc, Ebp2, and Zcchc8 all share a common epitope region including the pT58 residue of c-Myc (41):

c-Myc (56-69)	L P T P P L S P
Ebp2 (1-8)	M D T P P L S D
Zcchc8 (490-497)	K G T P P L T P

Caution must be used with this approach. The majority of antigens are of the discontinuous variety. Sequence alignments may yield continuous, false candidates, and considerable time and effort could be lost in ensuing validation efforts. As we have shown, BLAST alignment of the 31-amino acid peptide recognized by the anti-BTBD7 antibody generated multiple candidates. Mass spectrometry analysis failed to detect these proteins in our samples, and we therefore de-prioritized them as candidates. Eventually, ZCCHC8 was identified as the anti-BTBD7 cross-reacting protein. ZCCHC8 and BTBD7 have no sequence similarity within the regions used to generate the immunogens.

Our efforts demonstrate that multiple approaches may be required to successfully identify cross-reacting proteins. Success is not guaranteed, however, as much depends on the antibody in question. Within our case study, we also encountered considerable lot-to-lot variability and even the complete loss of activity/function in IP assays. The final production lot of anti-BTBD7 (lot k<sub>2</sub>) apparently detects a protein that

is neither the intended target, BTBD7, nor the major cross-reacting target, ZCCHC8, detected by the initial production lots! One antibody, three production lots, three different results—one begins to understand why antibodies receive much of the blame for the reproducibility crisis.

## **ZCCHC8 and BTBD7 Shared Epitope**

Interestingly, despite no sequence similarities detected in their respective immunogen regions, our results suggest that BTBD7 and ZCCHC8 may share a higher-order protein folding domain. First, the anti-BTBD7 antibody, generated with a 104-amino acid peptide that is unique to BTBD7, detects a nuclear protein in immunofluorescence and 105 kD protein in western blots that was determined to be ZCCHC8 (Figure 1.23, Figure 1.24). Second, at least one anti-ZCCHC8 antibody, generated with a 101-amino acid peptide unique to ZCCHC8, appears to bind BTBD7 (Figure 1.26). Lastly, through epitope mapping, we narrowed the antigenic region of the peptide bound by the anti-BTBD7 antibody to a 31-amino acid sequence. We then used this 31-amino acid peptide to generate a polyclonal antibody that produces enhanced nuclear staining in wild-type cells but shows diminished staining in both *Btbd7* and *Zcchc8* knockout cells (Figure 1.24).

The cross-reactivity of anti-BTBD7 antibodies to ZCCHC8, and anti-ZCCHC8 antibodies to BTBD7, demonstrates a shared epitope within these proteins, at a minimum. The complete lack of sequence similarity between the BTBD7 and ZCCHC8 immunogens further demonstrates that the epitope is of the discontinuous variety. Therefore, the epitope must be the product of higher dimensional protein folding.

Alas, diminished supplies of the initial antibody production lots, and the extreme variability in the final production lot, prevented any further elucidation. Further, any investigative efforts beyond those discussed fall well beyond the scope of research interests in our lab.

## **Variation in Staining Patterns of ZCCHC8**

Our initial interest in identifying the anti-BTBD7 cross-reacting protein was three-fold. First, expression in developing salivary glands displayed a remarkable enrichment in peripheral epithelial cells as early as day E11.5 (Figure 1.14A), which suggested the possibility of a developmental niche. Second, this peripheral enrichment, indeed expression altogether, was largely absent by day E16.5, suggesting an early developmental requirement (Figure 1.14B). Lastly, staining in the cell nucleus displayed a unique pattern of ~1,000 puncta up to ~200 nm in diameter consisting of multiple individual proteins that were concentrated in DAPI-sparse regions (Figure 1.15A-C).

Unfortunately, after confirming ZCCHC8 as the cross-reacting target, immunofluorescence staining with anti-ZCCHC8 antibodies did not recapitulate these findings. Rather than the peripheral epithelial enrichment initially seen only at early developmental stages in salivary glands, ZCCHC8 staining is largely uniform throughout (Figure 1.21A) and consistently expressed during development (Figure 1.21B).

This discrepancy is difficult to reconcile. The peripheral enhancement may be due to staining and/or imaging artifacts. Salivary glands are large, three dimensional objects that are rarely positioned such that the imaging plane neatly bisects every, or any one, bud. Consequently, a 2D image may capture regions that are not within the

same axial plane thereby giving the artificial impression of an enriched region. Further, differences in light penetration through dense tissue may enhance this effect. Lastly, insufficient permeabilization of the entire gland may prevent uniform antibody penetration, resulting in an uneven, artificial staining appearance. Indeed, in the time since our original salivary gland images were produced, our lab has refined our salivary gland imaging and staining techniques resulting in deeper penetrance for antibodies and molecular probes.

Cell culture imaging, on the other hand, should not be subject to these pitfalls, and yet we observed a marked difference in nuclear staining. Super-resolution imaging of cell nuclei stained with an anti-ZCCHC8 antibody revealed fewer puncta and a more diffuse overall staining pattern (Figure 1.22). One possibility for this discrepancy may be due to quaternary structural limitations of antibody binding. ZCCHC8 is a scaffold protein for a ternary association known as the nuclear exosome targeting complex. It is possible that the anti-ZCCHC8 antibody binds primarily free ZCCHC8, resulting in a diffuse pattern, while the anti-BTBD7 antibody primarily binds ZCCHC8 in a complex resulting in the more punctate appearance.

In the final analysis, it is difficult to ascribe to a protein, with any certainty, a unique expression pattern when the antibody used to produce the results is so obviously flawed. Contrary to our initial interest, our results suggest that ZCCHC8 is not expressed in a unique development niche. However, ZCCHC8 is a relatively understudied protein and, even though we were unable to recapitulate the results that captured our initial interest, we decided to undertake a mechanistic study in the hope of

uncovering a unique function for ZCCHC8. We will explore this undertaking in the next chapter.

## **Material and Methods**

### **Cell culture and maintenance**

MDCK.2, HeLa, and RT4 cells were purchased from ATCC (CRL-2936, CCL-2, and HTB-2, respectively). MDCK.2, HeLa, and SIMS cells were maintained in phenol red-free DMEM (GE Healthcare/Cytiva, SH30284.01) supplemented with 10% fetal bovine serum (FBS; GE Healthcare/Cytiva, SH30070.03) and incubated at 37°C with 10% CO<sub>2</sub>. RT4 cells were maintained in McCoy's 5a Modified Medium (ATCC, 30-2007) supplemented with 10% fetal bovine serum. Cells were passaged every three to four days using trypsin-EDTA (Thermo Fisher, 25300120) after rinsing with HBSS (Thermo Fisher, 14170161). Cell concentrations were determined using an automated cell counter (Nexcelom Cellometer Auto 2000).

### **CRISPR/Cas9 knockout of *Zcchc8* in SIMS cells**

The *Zcchc8* KN2.0 non-homology mediated mouse gene knockout kit (KN519669) was purchased from OriGene. Either the pCas-Guide CRISPR vector (OriGene, KN519669G1) containing a single guide RNA target sequence 5'-TAGGTGCGCCAAAATCCACAC-3' or the pCas-Guide CRISPR vector (OriGene, KN519669G2) containing a single guide RNA target sequence 5'-CGAGGCGTTTGACCCACCAG-3' or a combination of the two was transfected into the SIMS cell line using Thermo Fisher's Lipofectamine 3000 Reagent kit (L3000008) in the

following manner. Approximately  $1 \times 10^5$  SIMS cells were plated in each well of a 6-well plate the day prior to transfection. On the day of transfection, either 3.75 or 7.5  $\mu$ l of Lipofectamine 3000 reagent was diluted into 125  $\mu$ l of serum-free Opti-MEM in separate 1.5 ml microfuge tubes for each guide RNA or the combination. In separate 1.5 ml microfuge tubes, 1.0  $\mu$ g of either pCas-Guide CRISPR vector or the combination, 1.0  $\mu$ g of linear donor cassette with EF1a promoter followed by eGFP-P2A-Puromycin resistance (OriGene, KN519669D), and 2.0  $\mu$ l of P3000 reagent per  $\mu$ l DNA was diluted into 125  $\mu$ l of serum-free Opti-MEM. The diluted DNA mixtures were then added to the respective Lipofectamine microfuge tubes and the reactions were incubated for 15 minutes at room temperature. After incubation, the DNA-lipid mixtures were added drop-wise to the SIMS cells in the respective individual wells. After 48 hours post-transfection, the cells were split 1:10 into DMEM (Thermo Fisher, 11965118) + 10% FBS every three days for a total of 4 passages. The cells were then grown in the selective medium DMEM + 10% FBS + 2  $\mu$ g/ml puromycin (MilliporeSigma, P8833) for approximately one month. A subset of cells was harvested to check for *Zcchc8* knockout efficiency via western blot, immunofluorescence, and PCR. At this time, it was determined that the greatest knockout efficiency had been achieved using the pCas-Guide CRISPR vector with guide RNA sequence of 5'-TAGGTCGCCAAAATCCACAC-3' (OriGene, KN519669G1). These cells were chosen to produce individual clones using standard single cell cloning techniques in 96-well plates.

## **CRISPR/Cas9 knockout of *Btbd7* in SIMS cells**

Aneuploidy in SIMS cells required multiple rounds of CRISPR/Cas9 transfection to achieve complete *Btbd7* knockout. Separate single guide RNA sequences were cloned into separate plasmids containing a Cas9 and puromycin selection cassette: sgRNA1 5'-CACTCCGGCATTCCATCTGC-3' and sgRNA2 5'-TATCATCAGTAATGAATGCT-3'. The plasmids were then co-transfected into SIMS cells using the Lipofectamine 3000 Reagent kit with a reverse transfection procedure in 12-well plates. Briefly, SIMS wild-type cells were incubated with trypsin-EDTA at 37°C for ~20 minutes. During this incubation, 1 µg of each Cas9-sgRNA expressing plasmid was mixed with 55.5 µl of Opti-MEM for each well of the 12-well plate followed by 4 µl of P3000 reagent (DNA mixture). In a separate tube, 3.3 µl of Lipofectamine 3000 reagent was mixed with 55.5 µl of Opti-MEM for each well of the 12-well plate (Lipofectamine mixture). Next, 58 µl of the Lipofectamine mixture was added to each tube of the DNA mixture, mixed thoroughly, and incubated at RT for 15 minutes. During the Lipofectamine-DNA incubation, the trypsin-treated cells were resuspended to a concentration of  $4 \times 10^5$  cells/ml. After the incubation, 111 µl of the Lipofectamine-DNA mixture were added to each well of the 12-well plate followed by 1.25 ml of cell suspension. The cells were incubated for 48 hours before switching the growth media to selective growth media containing 2 µg/ml puromycin.

Standard single cell cloning techniques were used to produce multiple clones for genotyping. Genotyping was performed using the following PCR probes oSW84, 5'-TGCCATCAAACCCCTTACTTG-3', oSW233, 5'-GATGCTGTGGCAAACCTTCAA-3',

oSW338, 5'-CTCTGTGGGGAGTCTCAGTG-3' with expected PCR products of 1,365 bp wild-type band, 1,755 bp inversion, or no product for a complete deletion.

After genotyping, clones with at least one knockout allele were selected for another round of CRISPR/Cas9 transfections using two new sgRNA sequences: sgRNA3 5'- GTGAATTTCCCCCTACCCTT-3' and sgRNA4 5'- TATCATCAGTAATGAATGCT-3'. Transfections, clonal selection, and genotyping was performed as above. Absence of functional *Btbd7* mRNA was confirmed with RT-PCR using the probes oSW336, ACATGACCAGGGCAGAAGAA oSW337, CACACAGGAAATGCACAGCT.

## **Antibodies**

The following antibodies were used for immunofluorescence and western blots (see below for concentrations): Anti-BTBD7 (Novus, NBP2-14364), anti-E-cadherin (Thermo Fisher, 13-1900), anti-GFP (MilliporeSigma, MAB3580), Anti-ZCCHC8 (Abcam, ab181152; Proteintech, 23374-1-AP; MilliporeSigma, HPA037483, HPA037484, SAB1407675), Amersham ECL Mouse IgG, HRP-linked whole Ab (Cytivia, NA931), Amersham ECL Rabbit IgG, HRP-linked whole Ab (Cytivia, NA934).

Production of a custom, polyclonal anti-BTBD7 antibody was performed by Covance in the following manner. A KLH-conjugated, 31-amino acid peptide (YDVVLSFSSDSELVEAFGGNQNCLDEELKAH) was synthesized and subsequently immunized into two specific pathogen-free rabbits over the course of a 77-day protocol. On day 0, pre-bleed serum (5 ml) was collected from each rabbit followed by primary subcutaneous (SC) injection of 250 µg of immunogen emulsified in Freund's Complete



Adjuvant. On days 21, 42, and 63, the rabbits received an SC booster injection of 125 µg of immunogen emulsified in Freund's Incomplete Adjuvant. Ten days after each booster injection production bleeds of 20 ml serum were collected from each rabbit and a small aliquot was used for ELISA screening. On day 77, a termination bleed of 50 ml serum was collected from each rabbit. The serum was then pooled and affinity purified for anti-BTBD7 antibodies.

Antibody	Primary/Secondary	WB	IF
Anti-BTBD7 NBP2-14364	Primary	0.04 µg/ml	0.5 µg/ml
Anti-BTBD7 Covance	Primary	0.178 µg/ml	0.5 µg/ml
Anti-ZCCHC8 ab181152	Primary	0.1 µg/ml	0.5 µg/ml
Anti-ZCCHC8 23374-1-AP	Primary	0.025 µg/ml	0.5 µg/ml
Anti-ZCCHC8 HPA037483	Primary	0.04 µg/ml	0.5 µg/ml
Anti-ZCCHC8 HPA037484	Primary	0.12 µg/ml	0.5 µg/ml
Anti-ZCCHC8 SAB1407675	Primary	0.04 µg/ml	0.5 µg/ml
Anti-E-cadherin 13-1900	Primary	N/A	1 µg/ml
Anti-GFP MAB3580	Primary	N/A	1 µg/ml
HRP Anti-Mouse NA931	Secondary	1:5,000	N/A
HRP Anti-Rabbit NA934	Secondary	1:5,000	N/A
Anti-Rabbit AF647 711-606-152	Secondary	N/A	1.5 µg/ml
Anti-Rabbit AF488 711-546-152	Secondary	N/A	1.5 µg/ml
Anti-Rat AF488 712-545-150	Secondary	N/A	1.5 µg/ml

## Western blots

Approximately  $1 \times 10^6$  cells from MDCK.2, HeLa, SIMS wild-type, *Btbd7*- and *Zcchc8*-knockout clones were seeded into 10 cm dishes. At ~75% confluence, cells were washed with pre-chilled PBS followed by the addition of 500 µl of pre-chilled RIPA buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1× Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher, 78444). Cells were scraped on ice and collected in pre-chilled 1.5 ml tubes

(Eppendorf, 022363212). The cell suspensions were incubated on ice for 30 minutes followed by centrifugation at 13,000 rpm for 15 min at 4°C. Supernatants were transferred to pre-chilled 1.5 ml tubes and stored at -20°C. Lysate protein concentrations were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher, 23227). Aliquots of 25 µg lysate were denatured in 1X Laemmli sample buffer (Bio-Rad, 1610747) at 99°C for 5 min. Using a Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell, lysates and 5 µl of Precision Plus Protein Kaleidoscope Standards (Bio-Rad, 1610375) were run on Bio-Rad 7.5% Mini-PROTEAN TGX precast gels with Tris/Glycine/SDS (Bio-Rad, 1610732) running buffer at 115 V followed by transfer to Bio-Rad Trans-Blot Turbo 0.2 µm nitrocellulose via a Bio-Rad Trans-Blot Turbo Transfer system. Membranes were then incubated in blocking solution consisting of 5% nonfat dry milk in TBST (Tris Buffered Saline with 0.5% Tween-20; Quality Biological, 351-086-101; MilliporeSigma, P2287) for 1 hour at room temperature followed by incubation with primary antibodies diluted in blocking solution overnight at 4°C. Membranes were washed 3 times for 15 min each in TBST and incubated with secondary antibodies diluted in Blocking Solution for 1 hour at room temperature. The membranes were then washed 3 times for 15 min each in TBST at room temperature followed by chemiluminescence detection using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher, 34095). The blots were imaged using an Amersham Imager 680 (GE Healthcare, 29270769).

## **Immunofluorescence**

Approximately  $1 \times 10^5$  cells from MDCK.2, HeLa, SIMS wild-type, *Btbd7*- and *Zcchc8*-knockout clones were seeded into 35 mm MatTek dishes (MatTek, P35G-1.5-20-C) and grown for 48 hours. Cell medium was removed, and the cells were fixed with 2 ml of 4% PFA in PBS (Electron Microscopy Sciences, 15710) for 20 min at room temperature. Cells were then quickly washed with 2 ml of PBS followed by permeabilization with 1 ml of PBS with 0.1% Triton X-100 (Thermo Fisher, 28314) for 10 min at room temperature. Cells were then washed with 2 ml of Wash Buffer (PBS + 0.5% Tween-20) and then blocked for 1 hour with 1 ml of blocking buffer (Wash Buffer containing 3% fatty acid-free BSA (Thermo Fisher, 126609)). The cells were then incubated with 200  $\mu$ l of primary antibodies diluted in blocking buffer overnight at 4°C. The cells were then washed with 1 ml of Wash Buffer three times for 5 min each and incubated with 100  $\mu$ l of secondary antibodies diluted in blocking buffer for 1 hour at room temperature. Additionally, cells were stained with a 1:200 dilution of rhodamine phalloidin (Thermo Fisher, R415) during the secondary antibody incubation step. Cells were then washed with 1 ml of Wash Buffer three times for 5 min each. After the final wash, 12  $\mu$ l of Fluoro-Gel II with DAPI mounting medium (Electron Microscopy Sciences, 17985-50) was added to the cells on the glass surface of the MatTek dish. A coverslip was then sealed over the glass surface to protect the cells from damage.

## **Salivary gland isolation and culture**

Mouse submandibular salivary glands were isolated at embryonic day 11 through 17 (E11-E17) as previously described (42). Briefly, a scalpel (Fine Science Tools, 10011-

00 and 10003-12) was used to decapitate the mouse embryo. While the detached head was held on its side with one prong of forceps (Fine Science Tools, 11251-20) pierced through the top, a scalpel was used to slice across the mouth opening to isolate the mandible and tongue, between which the submandibular glands were sandwiched. Under a dissecting microscope, the detached mandible tissue was placed on a glass plate with the tongue facing down. A pair of forceps was used to slice through the midline of the mandible tissue to expose the tongue and the two submandibular glands attached to the base of the tongue. After surrounding tissues were removed, glands were detached using forceps and collected into a 35-mm dish with 3 ml DMEM/F-12 (Thermo Fisher, 11039047) media until all embryos were dissected. Isolated salivary glands were cultured on 13 mm diameter 0.1  $\mu\text{m}$  pore polycarbonate filters (MilliporeSigma, WHA110405) floating on 200  $\mu\text{l}$  Organ Culture Medium in the glass bottom area of a 50 mm MatTek dish (MatTek, P50G-1.5-14-F) at 37°C with 5% CO<sub>2</sub>. Organ Culture Medium was DMEM/F-12 supplemented with 150  $\mu\text{g/ml}$  vitamin C (MilliporeSigma, A7506), 50  $\mu\text{g/ml}$  transferrin (MilliporeSigma, T8158) and 1X PenStrep (100 units/ml penicillin, 100  $\mu\text{g/ml}$  streptomycin; Thermo Fisher, 15140163).

### **Immunostaining of salivary glands**

Cultured salivary glands were fixed (4% PFA in PBS; Electron Microscopy Sciences, 15710) on the filter by replacing Organ Culture Medium under the filter with 200  $\mu\text{l}$  fixative for 1 hour at room temperature (RT) or overnight at 4°C. Fixed glands were detached from the filter and transferred into sample baskets (Intavis, 12.440)

in PBS in a 35 mm dish using a pair of forceps (Fine Science Tools, 11251-20) under a dissecting microscope. Fixed samples in baskets were permeabilized in PBSTx (PBS with 0.2% Triton-X-100; Thermo Fisher, 28314) for 30 min at RT, blocked in 5% donkey serum (Jackson ImmunoResearch, 017-000-121) in PBSTx for 2 hours at RT, incubated in primary antibodies diluted in either PBSTx or 5% donkey serum for 2 days at 4°C, washed 4X 15 min in PBSTx at RT, incubated in secondary antibodies diluted in either PBSTx or 5% donkey serum for 2 days at 4°C, washed 4X 15 min in PBSTx at RT, rinsed in PBS and mounted under a dissecting microscope. To preserve fluorescence and to minimize compression, samples were mounted in 20 or 40 µl antifade mounting media (Thermo Fisher, P36930) supported by one layer of imaging spacers (Grace Biolabs, 654004) attached to a glass slide (Thermo Fisher, 3011-002).

### **Cell spreading assay**

Purified fibronectin was diluted in sterile PBS to a final concentration of 100 µg/ml in a tissue culture hood under sterile conditions. 200 µl aliquots of fibronectin solution were then evenly coated onto the surface of 35 mm MatTek dishes (MatTek, P35G-1.5-20-C). The fibronectin-coated dishes were then incubated at either 37°C for 1 hour or 4°C overnight before washing 3X with sterile PBS at RT.

MDCK.2 cells grown to ~70% confluency were detached using trypsin-EDTA and resuspended with cell-culture medium into 15 ml conical tubes. The cells were centrifuged for 2 minutes at 1,000 RPM. The media was discarded, and the cell pellet was washed with 10 ml of HBSS. The centrifugation and wash steps were repeated, and the cells were resuspended in complete DMEM. Cell concentration was

determined using an automated cell counter (Nexcelom Cellometer Auto 2000), and an aliquot of cell suspension was used to produce a final concentration of  $5.0 \times 10^3$  cells/ml. 2 ml of this cell suspension was added to either non-coated or fibronectin-coated MatTek dishes. The cells were then incubated at 37°C for 10-, 20-, 30-, 60-, or 120-minute intervals followed by fixation with 4% PFA and immunostaining.

## **Microscopy**

Immunofluorescence imaging was performed using the following microscopes and objectives.

Spinning disk confocal microscopy was performed using a system consisting of a Yokogawa CSU-22 scan head (CSU-21: modified by Spectral Applied Research, Inc.) on an automated Olympus IX-81 microscope using a 150X UApo N (NA 1.45) objective equipped with a custom laser launch with 442-, 488-, 568-, and 642-nm laser lines, a back-thinned EM-CCD camera (Photometrics), a motorized Z-piezo stage (ASI Imaging, Inc., Eugene, OR). All components were controlled by MetaMorph (Molecular Devices, Downingtown, PA).

Laser scanning confocal microscopy was performed using either a Nikon A1R+ MP system consisting of 405-, 488-, 561-, and 640-nm solid state lasers, two GaAsP and two PMT detectors, and 20X Plan Apo (NA 0.75) or 60X Plan Apo (NA 1.4) objectives controlled by Nikon NIS-Elements software or a Zeiss LSM 880 Airyscan system consisting of 405- and 561-nm solid state; 458-, 488-, and 514-nm argon; 633 nm HeNe lasers, one GaAsP and two PMT detectors, and 40X C-Apo (NA 1.1) or 63X C-Apo (NA 1.2) objectives controlled by Zeiss ZEN Black software.

Super-resolution imaging was performed using either a GE DeltaVision OMX SR system consisting of an enclosed DeltaVision OMX microscope with 405-, 488-, 568-, and 640-nm solid state lasers, 60X (1.42 NA) objective, and three CMOS cameras controlled by OMX SR Acquisition software or a Nikon N-STORM microscope consisting of a Nikon Ti-2 microscope frame, a laser launch with 405-, 488-, 561-, and 642-nm solid state lasers, a 100X SR Apo (NA 1.49) TIRF objective, and a CMOS (Flash 4 v3, Hamamatsu) or a back-thinned EM-CCD camera (Photometrics). Temperature, humidity and CO<sub>2</sub> were maintained constant using a Tokai stage-top incubator. The system was controlled by Nikon NIS-Elements software.

### **Epitope mapping and peptide competition assays**

The 104-amino acid BTBD7 recombinant protein antigen was purchased from Novus (NBP2-14364PEP). Four overlapping 31-amino acid subsequences were synthesized by LifeTein:

Peptide 1: LLHYLYTGEFGMEDSRFQNVLDILVQLSEEFV

Peptide 2: LSEEFVTPNSLDVDMRGLFDYMCYYDVVLSF

Peptide 3: YDVVLSFSSDSELVEAFGGNQNCLDEELKAH

Peptide 4: DEELKAHKAVISARSPFFRNLLQRRIRTVGEV

Peptide competition assays were performed by incubating anti-BTBD7 antibody (NBP2-14364) with 10X excess amount of peptide in either western blot or immunofluorescence blocking buffer for 30 minutes at RT on a rotating mixer. The antibody + peptide mixture was then used in western blot and immunofluorescence assays as described above.

## **RNA extraction, cDNA preparation, and qRT-PCR**

Approximately  $1 \times 10^5$  HeLa cells were seeded into 60 mm dishes and grown until cells reached ~75% confluence. Cell medium was removed, and 0.5 ml of TRIzol Reagent (Thermo Fisher, 15596026) was added directly to the dishes. After brief trituration, the lysates were collected into 1.5 ml tubes and incubated for 5 min at room temperature. Next, 0.1 ml of chloroform was added to each lysate and mixed thoroughly by inverting the tubes multiple times. The lysates were incubated at room temperature for three minutes and then centrifuged at 13,000 rpm for 15 min at 4°C. The aqueous phase of each lysate was then transferred to separate, fresh 1.5 ml tubes and 0.25 ml of 70% ethanol was added to each tube. The samples were then transferred to RNeasy spin columns and 2 ml collection tubes from an RNeasy Mini Kit (Qiagen, 74104) and centrifuged at 13,000 rpm for 15 sec. The flow-through was discarded and on-column DNase digestion was performed using the RNase-Free DNase Set (Qiagen, 79254) and the following protocol from Qiagen: 350  $\mu$ l of buffer RW1 was added to each column followed by centrifugation at 13,000 rpm for 15 sec. The flow-through was discarded. 80  $\mu$ l of DNase I incubation mix (10  $\mu$ l DNase I stock solution in 70  $\mu$ l buffer RDD) was added to each column membrane and incubated for 15 min at room temperature. 350  $\mu$ l of buffer RW1 was added to each column followed by centrifugation at 13,000 rpm for 15 sec. The flow-through was discarded and 500  $\mu$ l of buffer RPE was added to each column followed by centrifugation at 13,000 rpm for 15 sec. Again, 500  $\mu$ l of buffer RPE was added to each column followed by centrifugation at 13,000 rpm for 2 min. The RNeasy columns were then placed in new 2 ml collection tubes and centrifuged for 13,000 rpm for 1 min. The RNeasy spin columns were then placed in new 1.5 ml



collection tubes and 50  $\mu$ l of RNase-free water was added to the column membranes followed by centrifugation at 13,000 rpm for 1 min. The RNA samples were then stored at -80°C.

A Bio-Rad iScript cDNA synthesis kit (1708890) was used to produce cDNA according to the manufacturer's protocol. qRT-PCR was performed using an iTaq Universal SYBR Green One-Step kit (Bio-Rad, 172-5150) and the following primers:

*Btbd7* forward 5'-GAAGCCATGGAACTTTACCAC-3'

*Btbd7* reverse 5'-GAGCCATATGGATGAGAACTCC-3'

*Gapdh* forward 5'-TGTCTGGCACATTGGACATT-3'

*Gapdh* reverse 5'-GCACCGTCAAGGCTGAGAAC-3'.

### **Mass spectrometry**

Lysate collected from SIMS wild type cells was electrophoresed on a Bio-Rad 7.5% Mini-PROTEAN TGX precast gel with Tris/Glycine/SDS running buffer at 115 V. The gel was removed from the cassette and stained with SimplyBlue SafeStain (Thermo Fisher, LC6065). A narrow region surrounding a 105 kD band was excised and placed in a 1.5 ml Eppendorf microfuge tube then sent to MS Bioworks, Ann Arbor, MI for mass spectrometry analysis.

Sample digestion was performed using a robot (ProGest, DigiLab) to wash with 25 mM ammonium bicarbonate followed by acetonitrile, reduce with 10 mM dithiothreitol at 60°C followed by alkylation with 50 mM iodoacetamide at RT, digest with trypsin (Promega) at 37°C for 4 hours, and then quench with formic acid. The supernatant was then analyzed directly without further processing by nano LC/MS/MS with a Waters

NanoAcquity HPLC system interfaced to a Thermo Fisher Q Exactive. Peptides were loaded on a trapping column and eluted over a 75 $\mu$ m analytical column at 350nL/min; both columns were packed with Luna C18 resin (Phenomenex). The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 70,000 FWHM and 17,500 FWHM resolution, respectively. The fifteen most abundant ions were selected for MS/MS.

Data were searched using a local copy of Mascot with the following parameters: Enzyme: Trypsin; Database: Swissprot Mouse (forward and reverse appended with common contaminants); Fixed modification: Carbamidomethyl (C); Variable modifications: Oxidation (M), Acetyl (Protein N-term), Deamidation (NQ), Pyro-Glu (N-terminal Q); Mass values: Monoisotopic; Peptide Mass Tolerance: 10 ppm; Fragment Mass Tolerance: 0.02 Da; Max Missed Cleavages: 2. Mascot DAT files were parsed into the Scaffold software for validation, filtering and to create a non-redundant list per sample. Data were filtered using a minimum protein value of 90%, a minimum peptide value of 50% (Prophet scores) and requiring at least two unique peptides per protein.

## **2D electrophoresis**

SIMS wild-type cell lysates were used for two-dimensional gel electrophoresis by Kendrick Laboratories, Madison, WI. 2D electrophoresis was performed according to the carrier ampholyte method of isoelectric focusing. Isoelectric focusing was carried out in a glass tube of inner diameter 3.3 mm using 2.0% pH 3-10 Isodalt Servalytes (Serva, Heidelberg, Germany) for 20,000 volt-hrs. The enclosed tube gel pH gradient plot for this set of Servalytes was determined with a surface pH electrode.

After equilibration for 10 minutes in buffer "O" (10% glycerol, 50 mM dithiothreitol, 2.3% SDS and 0.0625 M tris, pH 6.8), each tube gel was sealed to the top of a stacking gel that overlaid a 10% acrylamide slab gel (1.0 mm thick). SDS slab gel electrophoresis was carried out for about 5 hours at 25 mA/gel. The following proteins (MilliporeSigma) were used as molecular weight standards: myosin (220,000), phosphorylase A (94,000), catalase (60,000), actin (43,000), carbonic anhydrase (29,000), and lysozyme (14,000). The gels were stained with Coomassie Brilliant Blue R-250 and dried between sheets of cellophane paper with the acid edge to the left.

After slab gel electrophoresis, the duplicate gels for blotting were placed in transfer buffer (10 mM CAPS, pH 11.0, 10% MeOH) and transblotted onto PVDF membrane overnight at 225 mA and approximately 100 volts/two gels. The same proteins as above were used as molecular weight standards.

### **Karyotyping**

Karyotyping of SIMS cells was performed by KaryoLogic, Research Triangle Park, NC.

# Bibliography

1. Antibodypedia. (January 16, 2021) antibodypedia. *Website*.
2. Research, G.V. (January 16, 2021) Research Antibodies Market Size & Share, Industry Report. *Website*.
3. Bordeaux, J., Welsh, A., Agarwal, S., Killiam, E., Baquero, M., Hanna, J., Anagnostou, V. and Rimm, D. (2010) Antibody validation. *Biotechniques*, **48**, 197-209.
4. Hoff, F.W., Lu, Y. and Kornblau, S.M. (2019) Antibody Screening. *Adv Exp Med Biol*, **1188**, 149-163.
5. Antoni, F.A., Wiegand, U.K., Black, J. and Simpson, J. (2006) Cellular localisation of adenylyl cyclase: a post-genome perspective. *Neurochem Res*, **31**, 287-295.
6. Drucker, D.J. (2013) Incretin action in the pancreas: potential promise, possible perils, and pathological pitfalls. *Diabetes*, **62**, 3316-3323.
7. Everaerts, W., Sepúlveda, M.R., Gevaert, T., Roskams, T., Nilius, B. and De Ridder, D. (2009) Where is TRPV1 expressed in the bladder, do we see the real channel? *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 421-425.
8. Hamdani, N. and van der Velden, J. (2009) Lack of specificity of antibodies directed against human beta-adrenergic receptors. *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 403-407.
9. Jensen, B.C., Swigart, P.M. and Simpson, P.C. (2009) Ten commercial antibodies for alpha-1-adrenergic receptor subtypes are nonspecific. *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 409-412.
10. Jositsch, G., Papadakis, T., Haberberger, R.V., Wolff, M., Wess, J. and Kummer, W. (2009) Suitability of muscarinic acetylcholine receptor antibodies for immunohistochemistry evaluated on tissue sections of receptor gene-deficient mice. *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 389-395.
11. Kirkpatrick, P. (2009) Specificity concerns with antibodies for receptor mapping. *Nat Rev Drug Discov*, **8**, 278.
12. Lu, X. and Bartfai, T. (2009) Analyzing the validity of GalR1 and GalR2 antibodies using knockout mice. *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 417-420.
13. Michel, M.C., Wieland, T. and Tsujimoto, G. (2009) How reliable are G-protein-coupled receptor antibodies? *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 385-388.
14. Pradidarcheep, W., Stallen, J., Labruyère, W.T., Dabhoiwala, N.F., Michel, M.C. and Lamers, W.H. (2009) Lack of specificity of commercially available antisera against muscarinergic and adrenergic receptors. *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 397-402.
15. Pyke, C. and Knudsen, L.B. (2013) The glucagon-like peptide-1 receptor--or not? *Endocrinology*, **154**, 4-8.
16. Weller, M.G. (2016) Quality Issues of Research Antibodies. *Anal Chem Insights*, **11**, 21-27.

17. Zhou, Y., Waanders, L.F., Holmseth, S., Guo, C., Berger, U.V., Li, Y., Lehre, A.C., Lehre, K.P. and Danbolt, N.C. (2014) Proteome analysis and conditional deletion of the EAAT2 glutamate transporter provide evidence against a role of EAAT2 in pancreatic insulin secretion in mice. *J Biol Chem*, **289**, 1329-1344.
18. Bodei, S., Arrighi, N., Spano, P. and Sigala, S. (2009) Should we be cautious on the use of commercially available antibodies to dopamine receptors? *Naunyn Schmiedebergs Arch Pharmacol*, **379**, 413-415.
19. Couchman, J.R. (2009) Commercial antibodies: the good, bad, and really ugly. *J Histochem Cytochem*, **57**, 7-8.
20. Parseghian, M.H. (2013) Hitchhiker antigens: inconsistent ChIP results, questionable immunohistology data, and poor antibody performance may have a common factor. *Biochem Cell Biol*, **91**, 378-394.
21. Perkel, J.M. (2014) The antibody challenge. *Biotechniques*, **56**, 111-114.
22. Pradidarcheep, W., Labruyère, W.T., Dabhoiwala, N.F. and Lamers, W.H. (2008) Lack of specificity of commercially available antisera: better specifications needed. *J Histochem Cytochem*, **56**, 1099-1111.
23. Rhodes, K.J. and Trimmer, J.S. (2006) Antibodies as valuable neuroscience research tools versus reagents of mass distraction. *J Neurosci*, **26**, 8017-8020.
24. Schonbrunn, A. (2014) Editorial: Antibody can get it right: confronting problems of antibody specificity and irreproducibility. *Mol Endocrinol*, **28**, 1403-1407.
25. Taussig, M.J., Fonseca, C. and Trimmer, J.S. (2018) Antibody validation: a view from the mountains. *N Biotechnol*, **45**, 1-8.
26. Kapingidza, A.B., Kowal, K. and Chruszcz, M. (2020) In Hoeger, U. and Harris, J. R. (eds.), *Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and other Body Fluid Proteins*. Springer International Publishing, Cham, pp. 465-497.
27. Ma, H. and O'Kennedy, R. (2015) The Structure of Natural and Recombinant Antibodies. *Methods Mol Biol*, **1348**, 7-11.
28. Van Regenmortel, M.H. (2014) Specificity, polyspecificity, and heterospecificity of antibody-antigen recognition. *J Mol Recognit*, **27**, 627-639.
29. Schroeder, H.W., Jr. and Cavacini, L. (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol*, **125**, S41-52.
30. Alving, C.R. (2006) Antibodies to lipids and liposomes: immunology and safety. *J Liposome Res*, **16**, 157-166.
31. Kappler, K. and Hennet, T. (2020) Emergence and significance of carbohydrate-specific antibodies. *Genes Immun*, **21**, 224-239.
32. Paul, E., Manheimer-Lory, A., Livneh, A., Solomon, A., Aranow, C., Ghossein, C., Shefner, R., Offen, D., Pillinger, M. and Diamond, B. (1990) Pathogenic anti-DNA antibodies in SLE: idiotypic families and genetic origins. *Int Rev Immunol*, **5**, 295-313.
33. Sherman, E.M., Holmes, S. and Ye, J.D. (2014) Specific RNA-binding antibodies with a four-amino-acid code. *J Mol Biol*, **426**, 2145-2157.
34. Swaak, T. and Smeenk, R. (1987) Clinical significance of antibodies to double stranded DNA (dsDNA) for systemic lupus erythematosus (SLE). *Clin Rheumatol*, **6 Suppl 1**, 56-73.
35. Ye, J.D., Tereshko, V., Frederiksen, J.K., Koide, A., Fellouse, F.A., Sidhu, S.S., Koide, S., Kossiakoff, A.A. and Piccirilli, J.A. (2008) Synthetic antibodies for

- specific recognition and crystallization of structured RNA. *Proc Natl Acad Sci U S A*, **105**, 82-87.
36. Buus, S., Rockberg, J., Forsstrom, B., Nilsson, P., Uhlen, M. and Schafer-Nielsen, C. (2012) High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Mol Cell Proteomics*, **11**, 1790-1800.
  37. Frank, S.A. (2002), *Immunology and Evolution of Infectious Disease*. Princeton University Press Copyright © 2002, Steven A Frank., Princeton (NJ).
  38. Zollman, S., Godt, D., Privé, G.G., Couderc, J.L. and Laski, F.A. (1994) 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*, **91**, 10717-10721.
  39. Stogios, P.J. and Privé, G.G. (2004) The BACK domain in BTB-kelch proteins. *Trends Biochem Sci*, **29**, 634-637.
  40. Onodera, T., Sakai, T., Hsu, J.C., Matsumoto, K., Chiorini, J.A. and Yamada, K.M. (2010) Btbd7 regulates epithelial cell dynamics and branching morphogenesis. *Science*, **329**, 562-565.
  41. Gustafson, M.P., Welcker, M., Hwang, H.C. and Clurman, B.E. (2005) Zcchc8 is a glycogen synthase kinase-3 substrate that interacts with RNA-binding proteins. *Biochem Biophys Res Commun*, **338**, 1359-1367.
  42. Sequeira, S.J., Gervais, E.M., Ray, S. and Larsen, M. (2013) Genetic Modification and Recombination of Salivary Gland Organ Cultures. *JoVE*, e50060.

## Chapter 2

# **ZCCHC8 is Required for the Degradation of Pervasive Transcripts Originating from Multiple Genomic Regulatory Features**

# Introduction

Historically, most transcriptional studies focused on the protein-coding portion of the mouse or human genome under the conventional wisdom that these 20,000 or so genes were interspersed throughout a largely non-transcribed, non-functional genome. Technical advances that brought about genome-wide analyses of transcription unveiled the reality that the vast majority of the genome is, in fact, transcribed as non-protein-coding RNAs, or non-coding RNAs (ncRNAs). Though intergenic transcription was described as early as 1997 (1), direct study of the phenomenon that would come to be known as “pervasive transcription” would alter transcriptomic dogma nearly a decade later (2-5). Global transcriptomic analyses, like the human ENCODE project (6), revealed the cumulative coverage of transcribed regions in 15 cell lines across the human genome is 62.1% and 74.7% for processed and primary transcripts, respectively. Considering that the accepted estimate of the protein-coding portion of the human genome is just 1.2% (7), it is no wonder that this transcriptional discrepancy has generated intense interest. So unforeseen was this transcriptional disjuncture that van Bakel et al. attempted to show that a majority of pervasive transcription was due to technical artifacts and/or background biological noise (8) thereby garnering much intrigue in the process (9-11). This counterpoint was quickly refuted, however (12), and the scientific community has since accepted the reality of pervasive transcription though the significance of this phenomenon is yet to be fully understood.

Generally, mammalian ncRNAs are categorized as either short or long. Short RNAs, widely accepted as less than 200 nucleotides in length, include the well-known small-nuclear and small-nucleolar RNAs (snRNAs and snoRNAs, respectively), transfer



RNAs (tRNAs), micro-RNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs), among others. As many of these short RNAs have well-described functions they are often not considered to be pervasively transcribed. However, there are many short RNAs that remain functionally undefined. In a 2007 tiling array-based study of the mammalian transcriptome, promoter-associated small RNAs (PASRs) and terminator-associated small RNAs (TASRs) were found to originate from the promoter and terminal regions of protein-coding genes (4). Shortly thereafter, Seila et al. described divergent transcription (defined as non-overlapping transcription initiation events that proceed in opposite directions from the transcription start site) at protein-encoding gene promoters that resulted in the production of so-called transcription start site-associated RNAs (TSSa-RNAs) (13). Finally, transcription initiation RNAs (tiRNAs) mapping within -60 to +120 nucleotides of the TSS and having a modal length of 18 nucleotides were discovered to reside on the same strand as the TSS and preferentially associate with G+C-rich promoters (14).

While variation in definition exists, long ncRNAs (lncRNAs) are commonly defined as being autonomously transcribed RNAs longer than 200 nucleotides with minimal coding potential (15-19) (ribosomal RNAs are generally excluded from the lncRNA classification). LncRNAs display a wide variety of functions and consist of intergenic and intronic transcripts, sense or antisense transcripts overlapping known genes (15,17,20), and enhancer RNAs (eRNAs) (21,22) that reportedly make up an estimated 19% of lncRNAs in human ESCs (23). The bulk of lncRNAs, however, likely consist of long intergenic/intervening non-coding RNAs (lincRNAs) (defined as autonomously transcribed non-coding RNAs longer than 200 nucleotides that do not

overlap annotated coding genes) (19,24). These lincRNAs share features with the other transcripts of the lincRNA family and may constitute more than half of lincRNA transcripts in humans (19). There is some uncertainty in this estimate, however, as many lincRNAs can be difficult to distinguish from eRNAs and there is apparent overlap with other RNA classes (16).

Of particular interest to the work presented here are divergent transcripts located upstream of the transcription start site known as promoter upstream transcripts (PROMPTs) and upstream antisense RNAs (uaRNAs) (25-28). While the lower bound of uaRNAs ranges from a few tens of nucleotides, the upper bound of greater than one thousand (25) more suitably places these RNAs within the lincRNA family. Early estimates suggest PROMPTs are ~200-600 nucleotides in length (27). PROMPTs and uaRNAs are highly unstable and short lived, are produced further upstream of active TSSs than TSSa- and tRNAs, and have both sense and anti-sense orientations (25,27,28).

Most certainly some pervasive transcription is functional and future research will need to clarify its significance. There is currently considerable debate regarding how much pervasive transcription is attributable to incidents of promiscuous RNA polymerase binding, initiation, and transcription. What seems certain, regardless of origin, is that unchecked pervasive transcription will have severe developmental and homeostatic consequences. Excess pools of unnecessary RNA can perturb transcription and initiate inappropriate translation fracturing the finely tuned kinetics of proper gene expression and regulation (29). Additionally, excess RNA can disrupt genome stability through the formation of hybridized RNA-DNA complexes, known as R

loops, that result in displaced single-stranded DNA that is highly susceptible to damage (29).

Obviously, pervasive transcription, and transcription in general, must be managed. As such, cells display remarkable systems of RNA surveillance. RNA surveillance mechanisms serve to monitor and manage short-lived, long-lived, and nonfunctional RNAs in a manner conducive to cellular and organismal well-being. While many RNA surveillance pathways have been detailed for protein-coding transcripts, less is known about the surveillance of ncRNAs. Yet, a few foundational concepts have been established. First, exo- and endo-ribonucleases serve to degrade unnecessary ncRNAs (16,18,30,31). Second, these ribonucleases usually function in concert with cofactors, helicases, and adapters that serve to recognize, unwind, and present the dispensable ncRNAs to the ribonucleases for degradation (16,18,30,31). Finally, these RNA surveillance pathways often have redundant partners (16,18,30,31).

Of particular note are the nuclear exosome-targeting (NEXT) complex and the poly(A) exosome-targeting (PAXT) connection (32,33). Both NEXT and PAXT function in the nucleoplasm and serve as adapter complexes that facilitate recognition and presentation of RNAs to the nuclear exosome (32-34). The nuclear exosome is a multi-subunit complex that includes the 3'-5' exoribonucleases, DIS3 and EXOSC10 (35). NEXT and PAXT both commonly include the RNA helicase MTR4/SKIV2L2 which serves to unwind the targeted RNA and directly interacts with the RNA exosome (36,37). The trimeric NEXT complex further consists of the zinc-knuckle scaffold ZCCHC8 and the RNA binding protein RBM7 (32). Though not fully elucidated, the PAXT connection is presumably comprised of the Zn-finger proteins ZFC3H1 and

ZC3H3, the poly(A) binding protein PABPN1, and the RNA binding proteins RBM26 and RBM27 (33,34,38).

The NEXT complex is responsible for exosomal targeting of PROMPTs, eRNAs, 3' extended snRNAs, 3' extended histone RNA, and intronic RNAs (32,34,39). Under normal conditions these RNAs are rapidly degraded and therefore difficult to detect. Depletion of NEXT components results in polyadenylation, and relative stabilization, of NEXT RNA substrates (32,34,39). PAXT, on the other hand, is responsible for targeting longer poly(A) transcripts that include diverse lncRNAs and prematurely terminated RNAs produced by intronic poly(A) sites within protein-coding genes (33,34,40). Recently, NEXT and PAXT have been shown to act redundantly with PAXT serving as a failsafe in the case of NEXT disruption (34). In the absence of NEXT complex subunits, the now polyadenylated NEXT substrates are targeted by the PAXT connection (34), albeit with likely altered kinetics.

Recently, in experiments regarding developing mouse salivary glands and salivary gland cells, we encountered an unexpected nuclear immunofluorescence staining pattern with intriguing spatiotemporal dynamics. This immunofluorescence staining was the result of antibody cross-reactivity to what we ultimately determined to be the NEXT component ZCCHC8. Because of its interesting nuclear localization pattern in developing salivary glands, we hypothesized that *in vitro* gene ablation experiments and RNA sequencing analyses could provide new insights into the function of ZCCHC8. We present comprehensive and comparative analyses of its functions *in vitro* and *in vivo*. In addition to its known RNA substrates, we have discovered that ZCCHC8 and/or the NEXT complex are responsible for the targeted degradation of

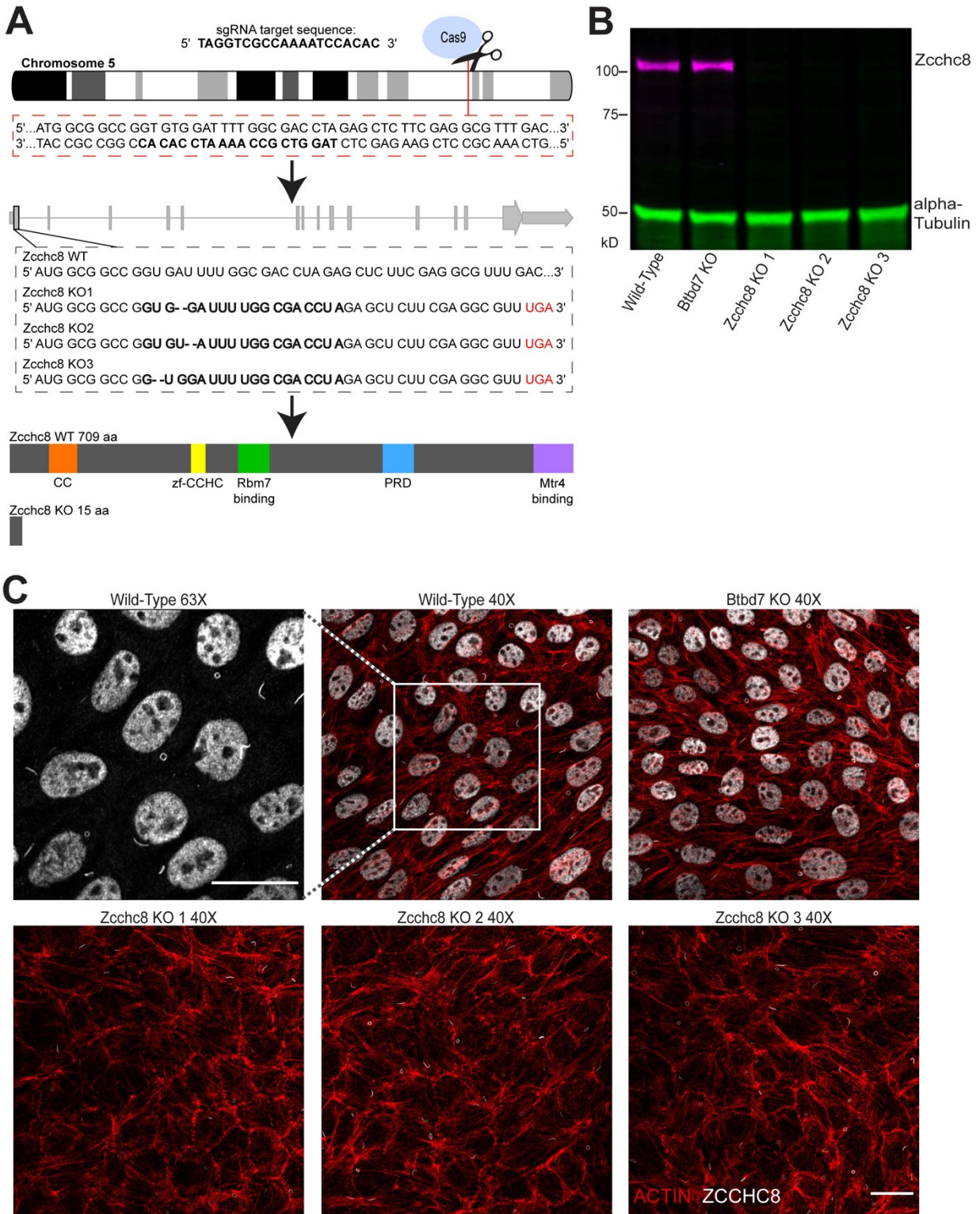
pervasive transcripts produced at CTCF binding sites, open chromatin regions, promoters, promoter flanking regions, and transcription factor binding sites. Further, besides identifying roles in suppressing levels of pervasive, spatially widespread non-coding transcripts, these in-depth analyses reveal that a surprising number of current RIKEN cDNAs (41) and predicted genes appear to be PROMPTs that are unlikely to be functional genes.

# Results

## ***Zcchc8* knockout in mouse salivary gland cells**

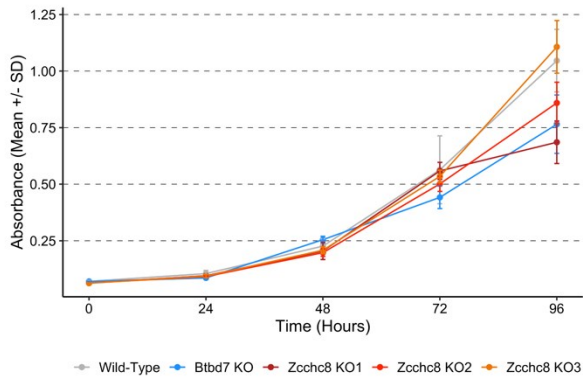
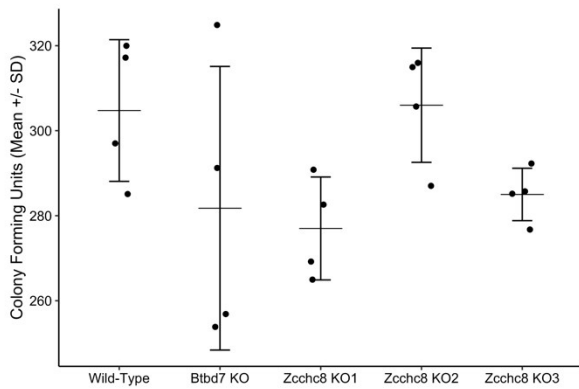
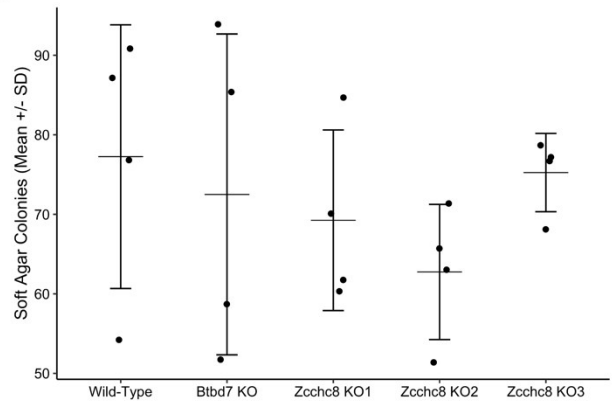
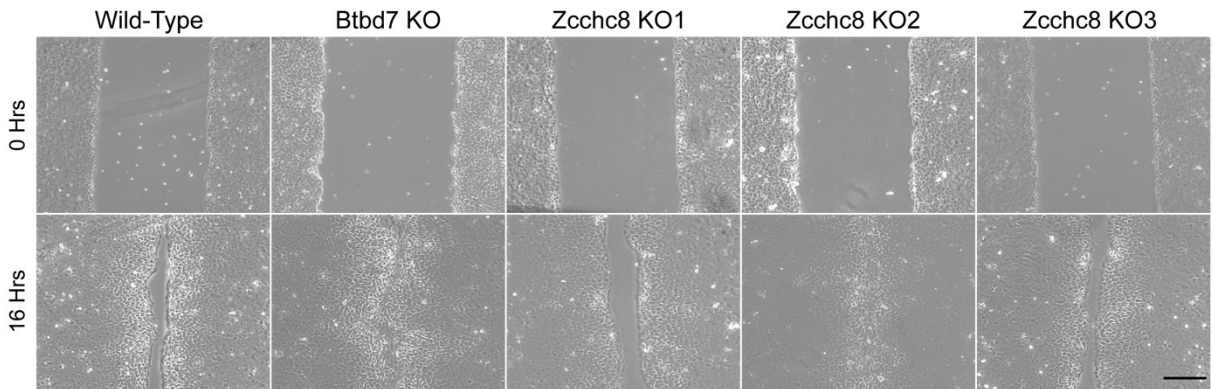
In order to test the hypothesis that ZCCHC8 has one or more alternative functions to that of its scaffolding role in the NEXT complex, we used the CRISPR/Cas9 system to knockout *Zcchc8* in the SIMS mouse salivary gland cell line (Figure 2.1A). We then isolated >20 clonal populations using standard single cell cloning techniques. After western blotting for ZCCHC8, we randomly selected three individual cell lines from those with depleted ZCCHC8 expression. Genetic sequencing confirmed differing two-base-pair deletions near the CRISPR/Cas9 target sequence that resulted in newly formed stop codons shortly downstream from the deletion site in each clone (Figure 2.1A). Follow-up western blotting and immunofluorescence analyses confirmed the absence of ZCCHC8 protein expression in the selected clones (Figure 2.1B-C). Gross morphological examination did not reveal any noticeable differences between the control and deletion cell lines.

We then performed a series of biological assays to assess the effects of *Zcchc8* knockout in this cell line. Somewhat remarkably, these *Zcchc8* knockout clones showed negligible differences in MTT proliferation, Matrigel invasion, colony formation, spheroid formation, soft agar, and scratch assays when compared to wild-type control (Figure 2.2A-E).



**Figure 2.1** (A) Schematic depicting the CRISPR/Cas9 strategy for disrupting *Zcchc8* in SIMS mouse salivary gland cells. Single cell cloning produced three separate knockout clones with two-base-pair deletions ~12-15 bp downstream of the TSS. The resulting frameshift produced a translational stop codon at amino acid position 16. (B) Western blot confirmation of ZCCHC8 ablation in three separate clones. Western blots were probed with a rabbit, polyclonal anti-ZCCHC8 antibody from Proteintech and a mouse, monoclonal anti- $\alpha$ -Tubulin antibody from Sigma. (C) Immunofluorescence confirmation of ZCCHC8 ablation in three separate clones. The same anti-ZCCHC8 antibody as in (B) was used. Note the presence of antibody cross-reactivity to primary cilia before and after *Zcchc8* knockout. Actin fibers were stained with a rhodamine-labeled phalloidin from Thermo Fisher. Scale bar = 20 $\mu$ m. Note: Figure 2.1 and Figure 1.19 on page 34 are identical.



**A****B****C****D**

**Figure 2.2** Biological assays. (A) MTT proliferation assay results. All groups n = 6. (B) Colony forming assay results. All groups n = 4. (C) Soft agar colony forming assay results. All groups n = 4. (D) Representative images from wound healing assays. All groups n = 3.

## RNA sequencing of *Zcchc8* knockout cells

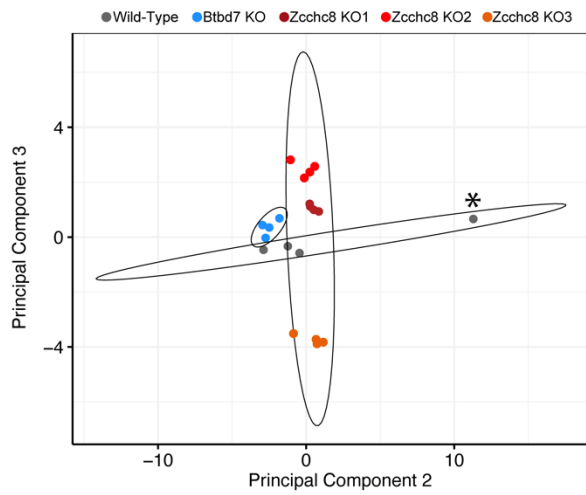
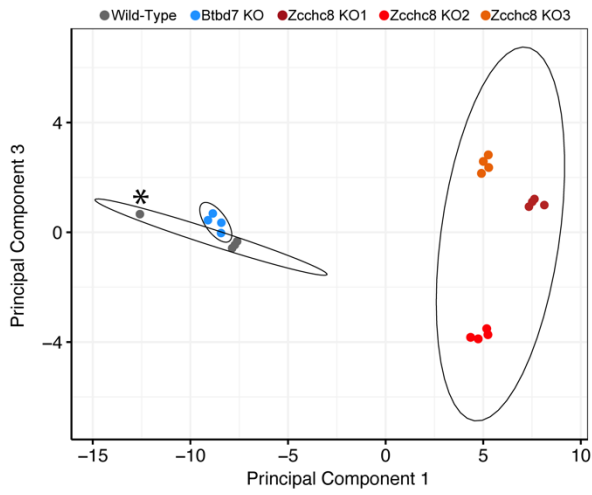
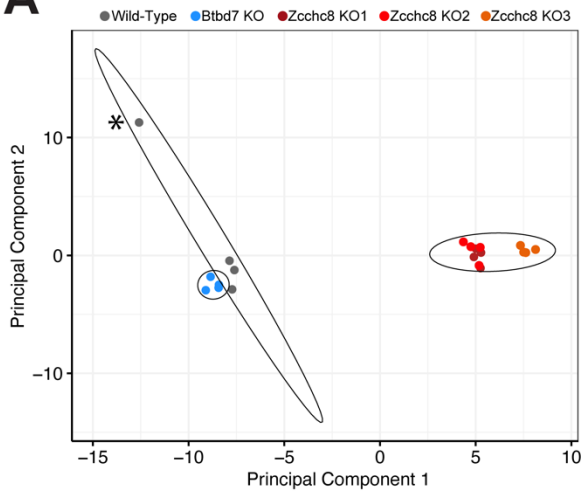
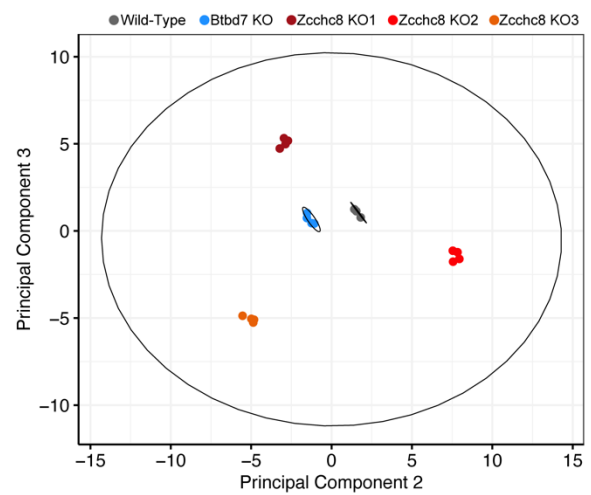
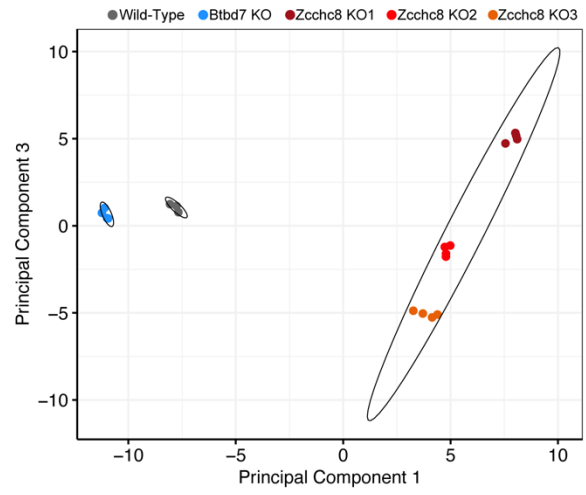
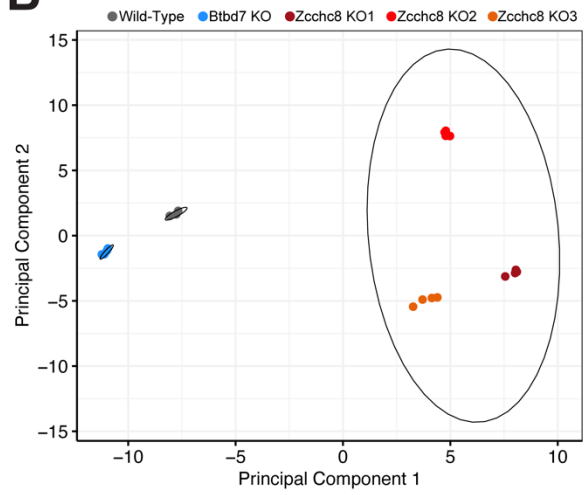
In the absence of any observed in vitro biological phenotype related to morphology, proliferation, survival, or migration, we proceeded to perform RNA sequencing on wild-type and *Zcchc8* knockout SIMS cells. As an additional negative control, we also included SIMS cells in which we previously disrupted the gene *Btbd7* using the CRISPR/Cas9 system. Considering the role of ZCCHC8 and the NEXT complex in RNA degradation pathways, we used the Illumina TruSeq Stranded Total RNA library preparation kit in order to ultimately detect differential expression of small RNAs, lncRNAs, and mRNAs. Four replicate libraries were prepared from each cell line. The samples were analyzed on an Illumina NextSeq500 configured for 40 paired-end reads and the quality of the reads was evaluated using FastQC software. Read mapping against the recent GENCODE mouse release M24 was performed using the STAR 2.7.3a aligner (see Materials and Methods).

We then performed differential expression analysis of genomic features using three independent statistical tests to assess differential expression between groups: DESeq2, edgeR, and Limma-Voom. These methods differ in normalization of feature expression across samples, assumptions about the distribution of the underlying data and statistical test used, and are widely accepted in the current literature.

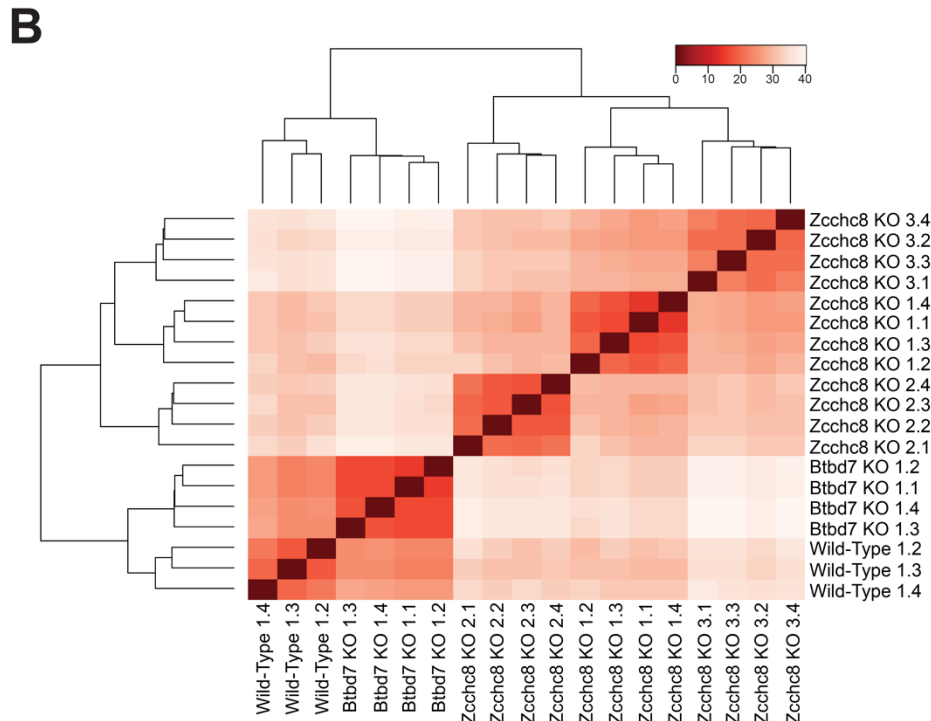
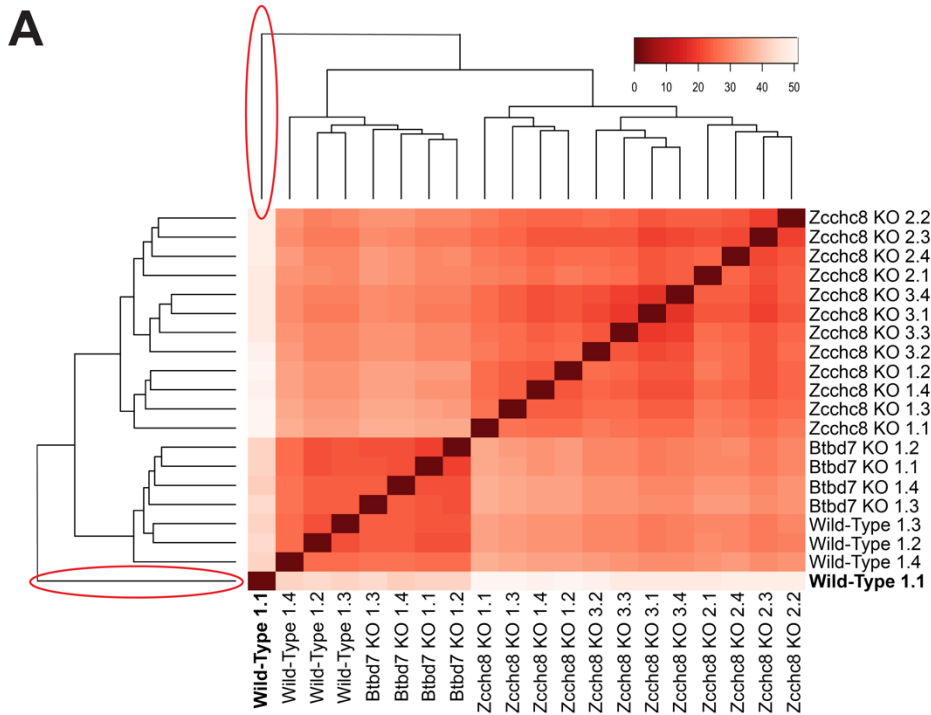
Principal component analysis (PCA), multidimensional scaling (MDS), and Euclidean distance clustering (EDC) of gene expression data revealed a single wild-type SIMS replicate that differed significantly from the other wild-type replicates (Figure 2.3A, Figure 2.4A). This replicate was removed from all subsequent analyses. The replicates of all other clonal groups were highly similar and were included in all analyses

(Figure 2.3B, Figure 2.4B). Importantly, unsupervised EDC revealed two major clusters: a control cluster consisting of wild-type and *Btbd7* knockout cells and a *Zcchc8* knockout cluster (Figure 2.4B). Further, the replicates of each clone clustered together into related subclusters (Figure 2.4B).

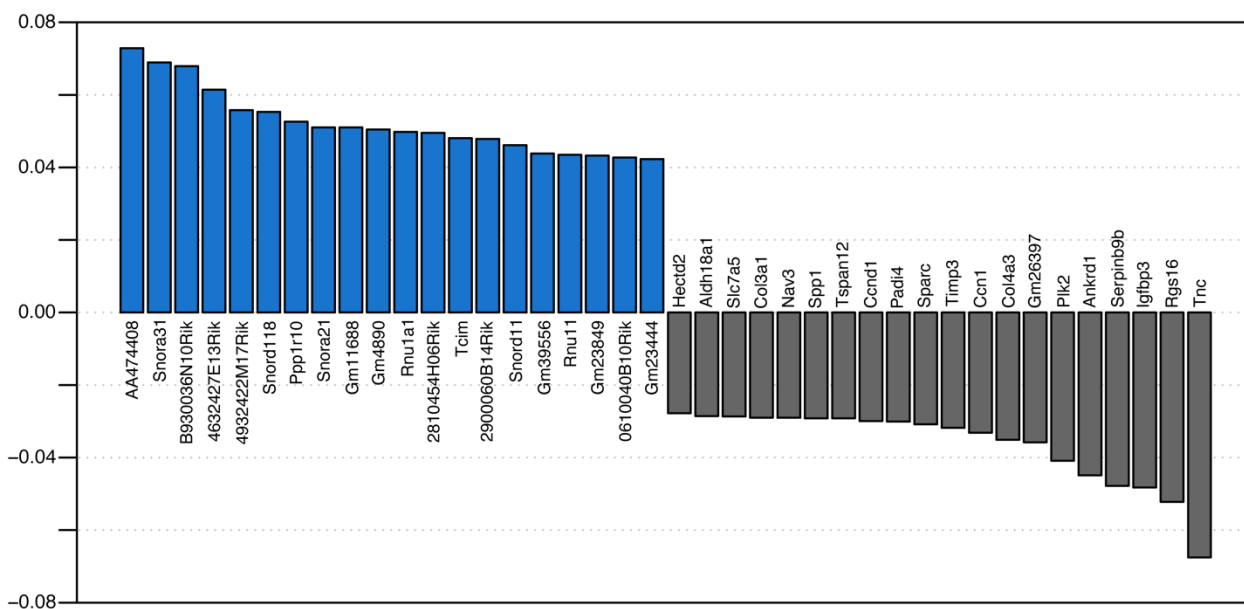
Strikingly, a plot of the top 20 genes both positively and negatively correlated with the variance of the first principal component revealed a large number of RIKEN cDNAs and predicted genes (with the Mouse Genome Informatics naming convention of *Gmxxxxx*) positively correlated with this variance (Figure 2.5). In fact, of the top 100 genes positively correlated with the variance of the first principal component, 62 are RIKEN cDNAs and predicted genes (29 and 33, respectively) (Table 2.1). Only 11 are snRNAs or snoRNAs—two RNA species that are known substrates for ZCCHC8 and the NEXT complex.

**A****B**

**Figure 2.3** (A) Principal component analysis plots depicting a single outlier sample from the wild-type group. Asterisks mark the outlier sample. Principal component variance: PC1 = 69.19%, PC2 = 10.81%, PC3 = 5.26%. (B) Principal component analysis plots after removal of the outlier sample from analysis. Principal component variance: PC1 = 58.41%, PC2 = 20.08%, PC3 = 11.61%. Ellipses represent 95% confidence regions in both (A) and (B).



**Figure 2.4** (A) Heatmap of sample-to-sample distances after unsupervised clustering of variance-stabilizing-transformed data. The wild-type 1.1 sample clustered individually on a single node (red circles, bold-type). (B) Heatmap produced using the same data analysis as in (A) after the removal of the outlier wild-type 1.1 sample.



**Figure 2.5** Loadings bar chart of the top 20 genes correlating positively and negatively with the variance of the first principal component.

**Table 2.1** Top 100 positively correlated loadings for the first principal component

---

[1] AA474408	[51] <b>Gm43182</b>
[2] Snora31	[52] <b>Gm37352</b>
[3] <b>B930036N10Rik</b>	[53] <b>Gm17767</b>
[4] <b>4632427E13Rik</b>	[54] <b>4930402H05Rik</b>
[5] <b>4932422M17Rik</b>	[55] <b>Gm15706</b>
[6] Snord118	[56] Phtf1os
[7] Ppp1r10	[57] <b>4632404H12Rik</b>
[8] Snora21	[58] <b>4930581F22Rik</b>
[9] <b>Gm11688</b>	[59] Trmt12
[10] <b>Gm4890</b>	[60] <b>Gm44064</b>
[11] Rnu1a1	[61] <b>Gm10209</b>
[12] <b>2810454H06Rik</b>	[62] Bmf
[13] Tcim	[63] Lrrc26
[14] <b>2900060B14Rik</b>	[64] Chdh
[15] Snord11	[65] <b>Gm37305</b>
[16] <b>Gm39556</b>	[66] <b>Gm10382</b>
[17] Rnu11	[67] Neat1
[18] <b>Gm23849</b>	[68] Mrpl17
[19] <b>0610040B10Rik</b>	[69] <b>Gm13523</b>
[20] <b>Gm23444</b>	[70] Snord13
[21] <b>1700007L15Rik</b>	[71] <b>D730045B01Rik</b>
[22] Msh5	[72] <b>1700096K18Rik</b>
[23] <b>Gm25360</b>	[73] <b>9330162G02Rik</b>
[24] Snhg4	[74] Grin1
[25] <b>Gm23971</b>	[75] <b>Gm17259</b>
[26] Lncppara	[76] <b>Gm27680</b>
[27] Mttp	[77] <b>E030037K01Rik</b>
[28] <b>1700052K11Rik</b>	[78] <b>Gm10461</b>
[29] <b>6430511E19Rik</b>	[79] <b>B330016D10Rik</b>
[30] Dleu2	[80] BC037704
[31] AI480526	[81] <b>Gm22422</b>
[32] <b>Gm32699</b>	[82] <b>Gm26766</b>
[33] <b>4921531C22Rik</b>	[83] <b>5031415H12Rik</b>
[34] <b>Gm20633</b>	[84] <b>2810402E24Rik</b>
[35] <b>Gm38391</b>	[85] Nabp1
[36] <b>8030456M14Rik</b>	[86] <b>Gm26670</b>
[37] <b>Gm38140</b>	[87] <b>4933417C20Rik</b>
[38] <b>Gm29170</b>	[88] <b>D730044K07Rik</b>
[39] <b>Gm24830</b>	[89] <b>A430078I02Rik</b>
[40] <b>Gm9958</b>	[90] <b>Gm4419</b>
[41] <b>Gm22265</b>	[91] Cd33
[42] <b>2700046G09Rik</b>	[92] <b>D430020J02Rik</b>
[43] Mir99ahg	[93] Rnu5g
[44] Uap111	[94] <b>Gm44686</b>
[45] <b>1700001G17Rik</b>	[95] Ankle1
[46] Snhg15	[96] <b>1700063J08Rik</b>
[47] Rdm1	[97] Khlh24
[48] Snhg20	[98] Nupr1l
[49] <b>4933406P04Rik</b>	[99] H3f3aos
[50] Lncpint	[100] <b>Gm11335</b>

---

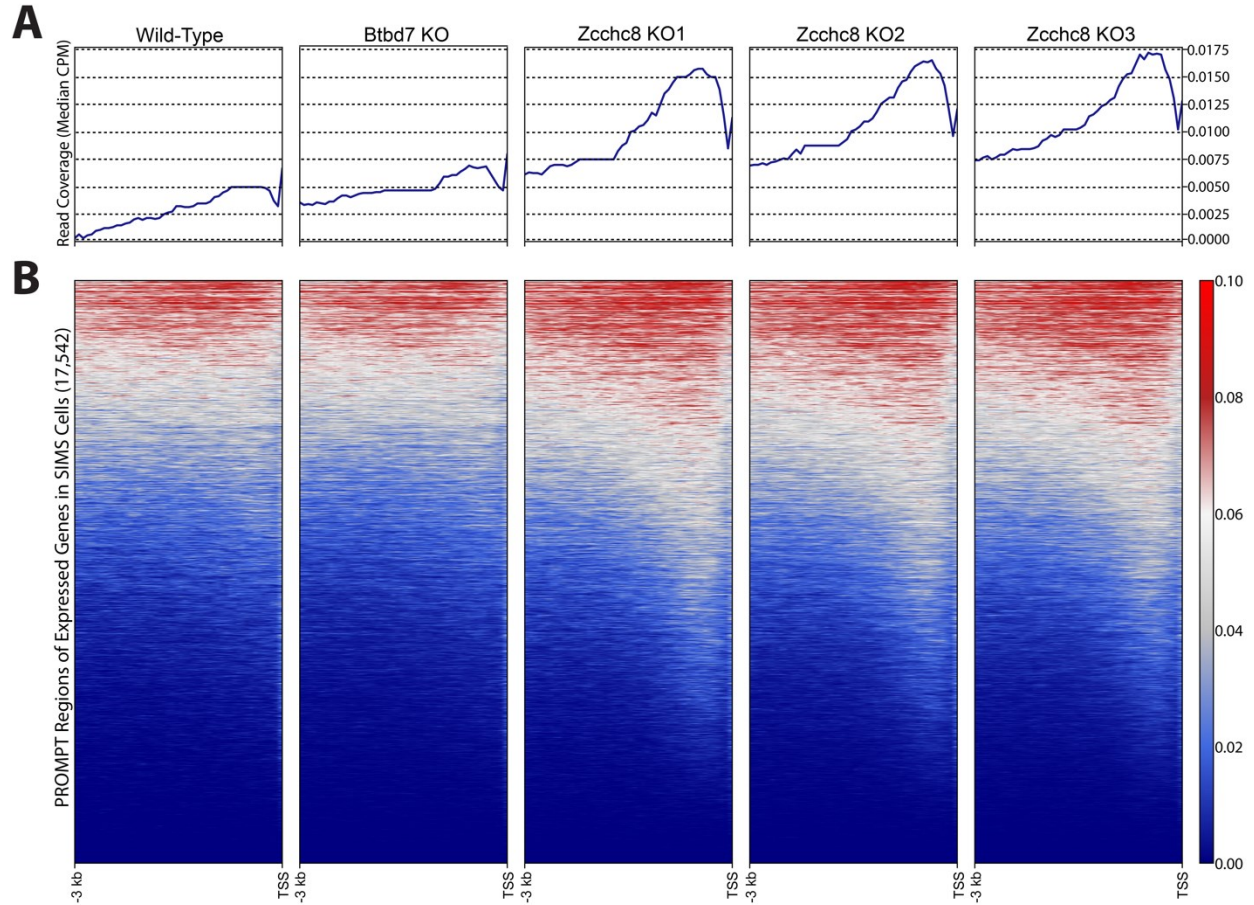
Top 100 genes positively correlated with the variance of the first principal component in order of variance contribution. RIKEN cDNAs and predicted genes are in bold.



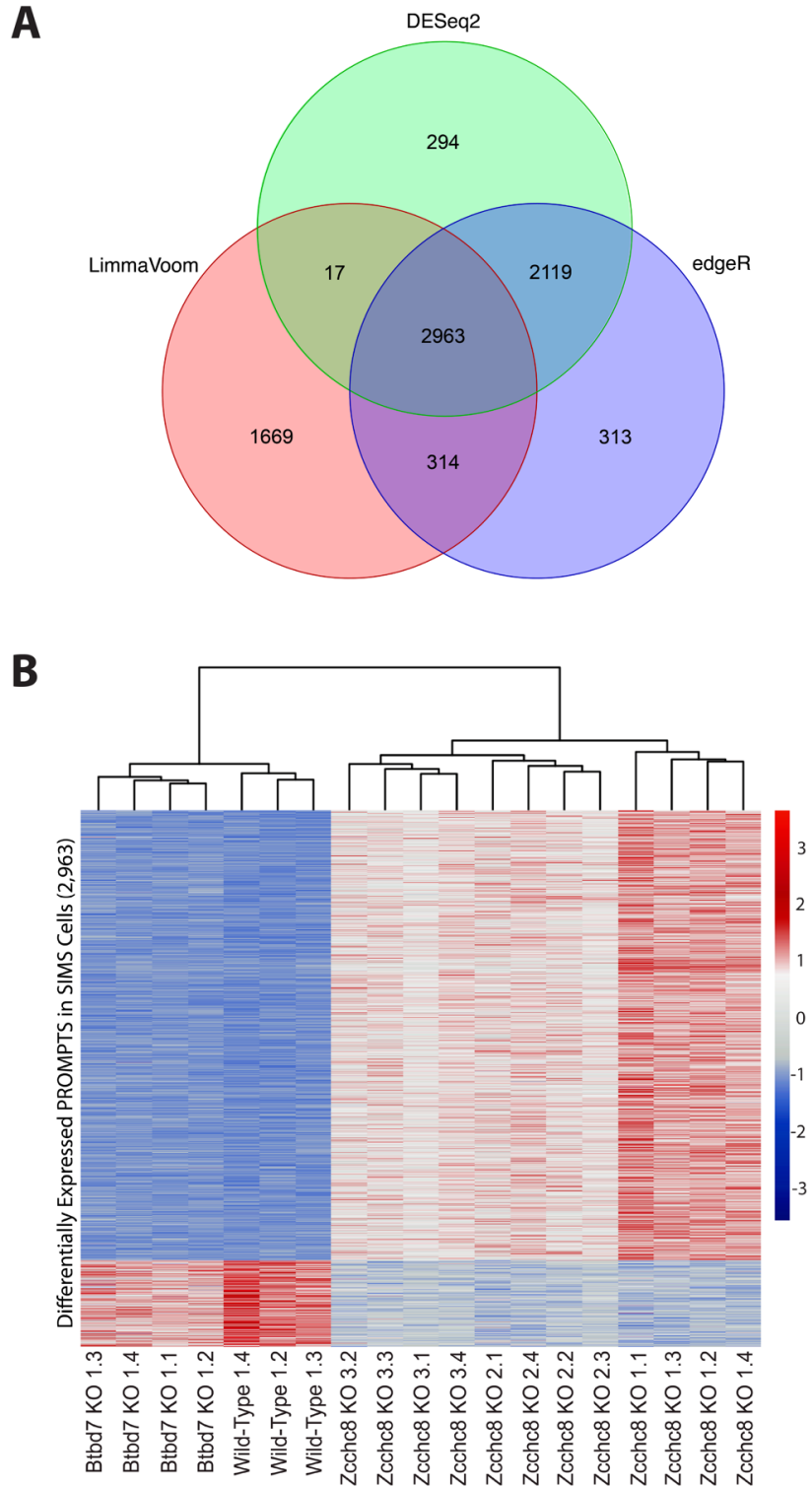
## Effects of *Zcchc8* knockout on genomic regulatory features

In order to confirm the role of ZCCHC8 in the degradation of PROMPT RNAs, we generated metagene plots and read-coverage heat maps of the 3 kb region upstream of the TSS for all expressed genes using the Python package deepTools. As expected, knockout of *Zcchc8* resulted in the appearance of PROMPTs upstream of the TSS for numerous genes (Figure 2.6A-B). Metagene plots revealed the majority of PROMPTs to be less than 1.5 kb upstream of the TSS (Figure 2.6A). *Btbd7* knockout had no effect on transcription upstream of TSSs (Figure 2.6A-B, Figure 2.7B).

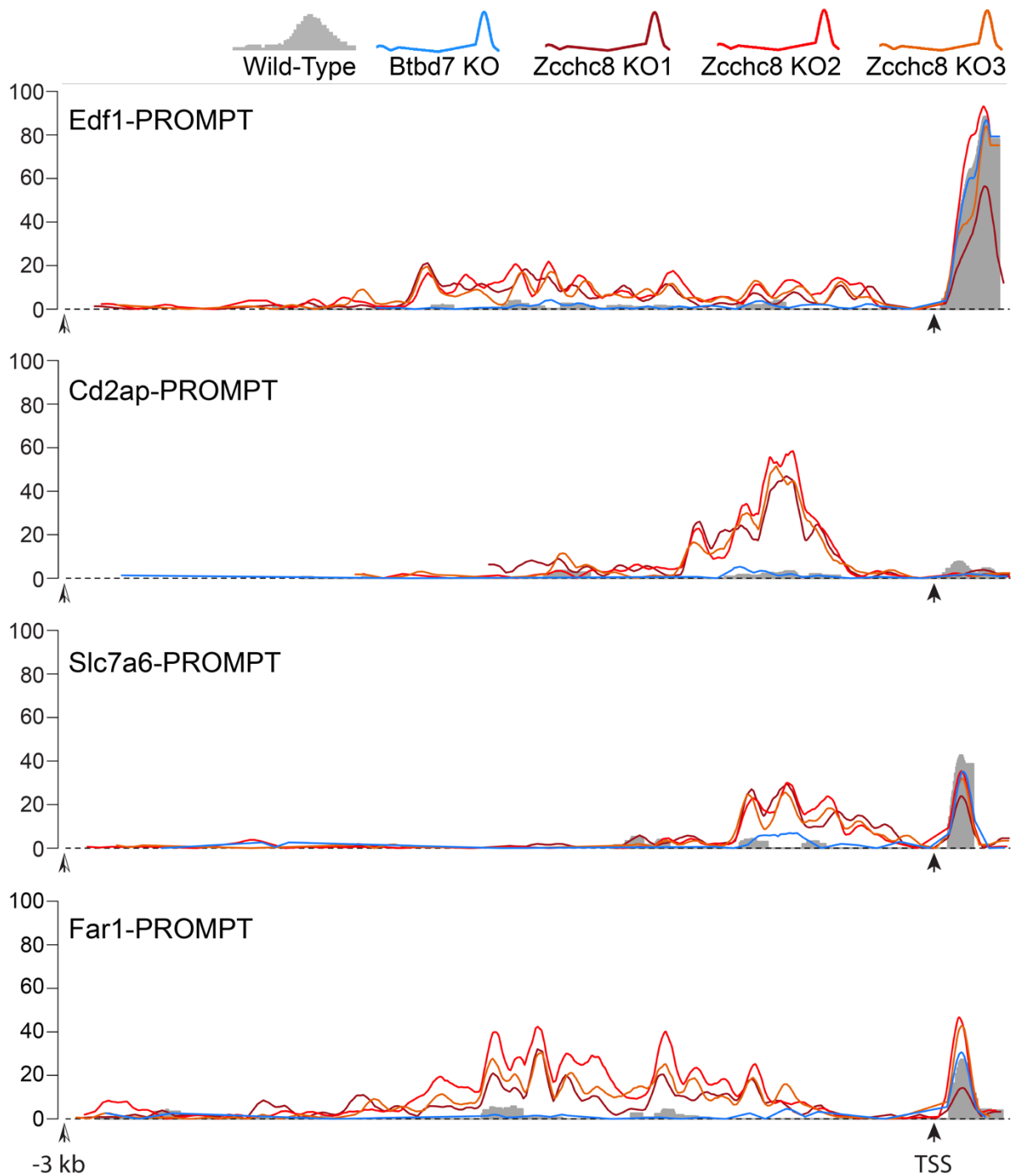
We next performed differential expression analysis of all PROMPT regions 3 kb upstream of annotated gene start sites. Using cut-off values of >1.5-fold expression difference and <0.01 adjusted p-value, DESeq2, edgeR, and Limma-Voom revealed the presence of 5,393, 5,709, and 4,963 differentially expressed PROMPT regions, with 2,963 in common (Figure 2.7A, Table 2.2). Hereafter, we will only refer to those differentially expressed PROMPTs, genes, regulatory features, etc., that are shared according to all three statistical methods. A heat map of PROMPT region expression is shown in Figure 2.7B. Read coverage plots for selected differentially expressed PROMPTs were generated using the Gviz package in R and are shown in Figure 2.8. These detailed plots reveal the remarkable similarity of PROMPT transcription read coverage in three separate knockout clones as compared to wild-type and *Btbd7* knockout controls (Figure 2.8).



**Figure 2.6** (A) Metagenes plots showing the median read coverage for CPM-normalized read counts in 3-kb regions upstream of the TSS for all expressed genes. Genes were considered to be expressed if there were >5 reads in at least one sample. (B) Read coverage heatmaps of the 3-kb upstream regions in (A). In order to create a single plot of all clonal replicates in (A) and (B), read counts from replicate bam files were merged and indexed using the SAMtools software package followed by CPM-normalization.



**Figure 2.7** (A) Venn diagram showing the number of differentially expressed PROMPTs determined via DESeq2, edgeR, and Limma-Voom analysis in SIMS *Zcchc8* knockout cells. (B) Heatmap using k-means clustering ( $k = 2$ ) of the 2,963 PROMPTs shared among the statistical analysis groups in (A).



**Figure 2.8** Read coverage plots indicating the differential expression of PROMPTs within the 3 kb upstream region of the TSS for the genes *Edf1*, *Cd2ap*, *Slc7a6*, and *Far1*. In order to create a single plot of all clonal replicates, read counts from replicate bam files were merged and indexed using the SAMtools software package followed by CPM-normalization prior to plotting using the Gviz R package. Black arrows indicate the TSS for each gene followed by the first 250 bp of transcription coverage. Open arrows indicate the -3 kb position relative to the TSS.

**Table 2.2** Differentially expressed genomic regulatory features

	DESeq2		edgeR		LimmaVoom		Shared	
	Up	Down	Up	Down	Up	Down	Up	Down
<b>PROMPTs</b>								
<i>Zcchc8</i> KO	4618	775	4353	1356	2533	2430	2486	477
<i>Btbd7</i> KO	42	26	40	36	10	12	5	8
<b>Enhancers</b>								
<i>Zcchc8</i> KO	439	50	400	78	258	68	243	28
<i>Btbd7</i> KO	9	4	10	6	11	5	3	3
<b>Promoters</b>								
<i>Zcchc8</i> KO	338	42	327	63	112	38	105	22
<i>Btbd7</i> KO	6	3	8	7	4	2	1	2
<b>PFR</b>								
<i>Zcchc8</i> KO	839	75	771	126	487	157	471	51
<i>Btbd7</i> KO	17	11	14	6	13	8	2	3
<b>CTCF sites</b>								
<i>Zcchc8</i> KO	469	38	425	67	269	71	247	19
<i>Btbd7</i> KO	9	4	4	2	9	9	2	2
<b>TFBS</b>								
<i>Zcchc8</i> KO	86	10	72	13	48	15	45	6
<i>Btbd7</i> KO	3	2	3	0	1	0	1	0
<b>OCR</b>								
<i>Zcchc8</i> KO	239	27	224	35	151	48	142	9
<i>Btbd7</i> KO	3	4	1	4	3	1	0	0
<b>Genes</b>								
<i>Zcchc8</i> KO	86	10	72	13	48	15	45	6
<i>Btbd7</i> KO	3	2	3	0	1	0	1	0

Differentially expressed genomic regulatory features as determined by three different analyses using cut-off values of >1.5-fold expression difference and adjusted p-value <0.01. The number of features that were commonly found within all three statistical analyses are indicated in the Shared column. PFR = Promoter Flanking Regions, TFBS = Transcription Factor Binding Sites, OCR = Open Chromatin Regions.

Considering that divergent transcription at TSSs is presumed to occur at most active protein coding genes (13,23,26,28,42,43), we expected to find a larger number of PROMPTs in our *Zcchc8* knockout cells. We wondered if NEXT complex redundancy through the PAXT connection would be evident through increased transcription or if the remaining NEXT complex components or RNA exosome were up-regulated as a compensatory mechanism for loss of ZCCHC8 function. Upon further examination, we found no evidence of such regulation at the transcriptional level (Table 2.3). We also mapped the chromosomal distribution of PROMPTs and found they were widespread and well-distributed across all chromosomes (Figure 2.9).

Differential expression analysis also detected ~15% PROMPT regions that were down-regulated in *Zcchc8* knockout cells (Figure 2.7B). As the function of the NEXT complex is to degrade PROMPTs, we examined how ZCCHC8 disruption could result in decreased transcription in these regions. Deeper examination revealed portions of full-length expressed genes overlapping these PROMPT regions in either head-to-head or head-to-tail orientation. Thus, these regional transcriptional differences are due to differential gene expression rather than to authentic PROMPT expression. Hereafter, we will refer to PROMPTs as those 3 kb upstream regions that are up-regulated in *Zcchc8* knockout cells. Accordingly, we found 2,486 PROMPTs in our SIMS *Zcchc8* knockout cells.

Similar to our PROMPT analysis, we performed differential expression analyses on regulatory features (as annotated in the Ensembl Regulation database) and found ZCCHC8 ablation resulted in transcriptional up-regulation of 247 CTCF binding sites, 243 enhancers, 142 open chromatin regions, 104 promoters, 471 promoter flanking

regions, and 45 transcription factor binding sites (>1.5-fold expression difference and <0.01 adjusted p-value) (Figure 2.10A). In *Btbd7* knockout cells, there were fewer than 5 differentially expressed regulatory features of each category. The results of this differential expression analysis are summarized in Table 2.2. Further, we evaluated whether any of these regulatory features overlapped with PROMPTs and found only a small number shared intersecting genomic coordinates (Figure 2.10B).

Lastly, we mapped the chromosomal locations of these regulatory regions and found them to be well-distributed along each chromosome with the exception of chromosome X (Figure 2.11).

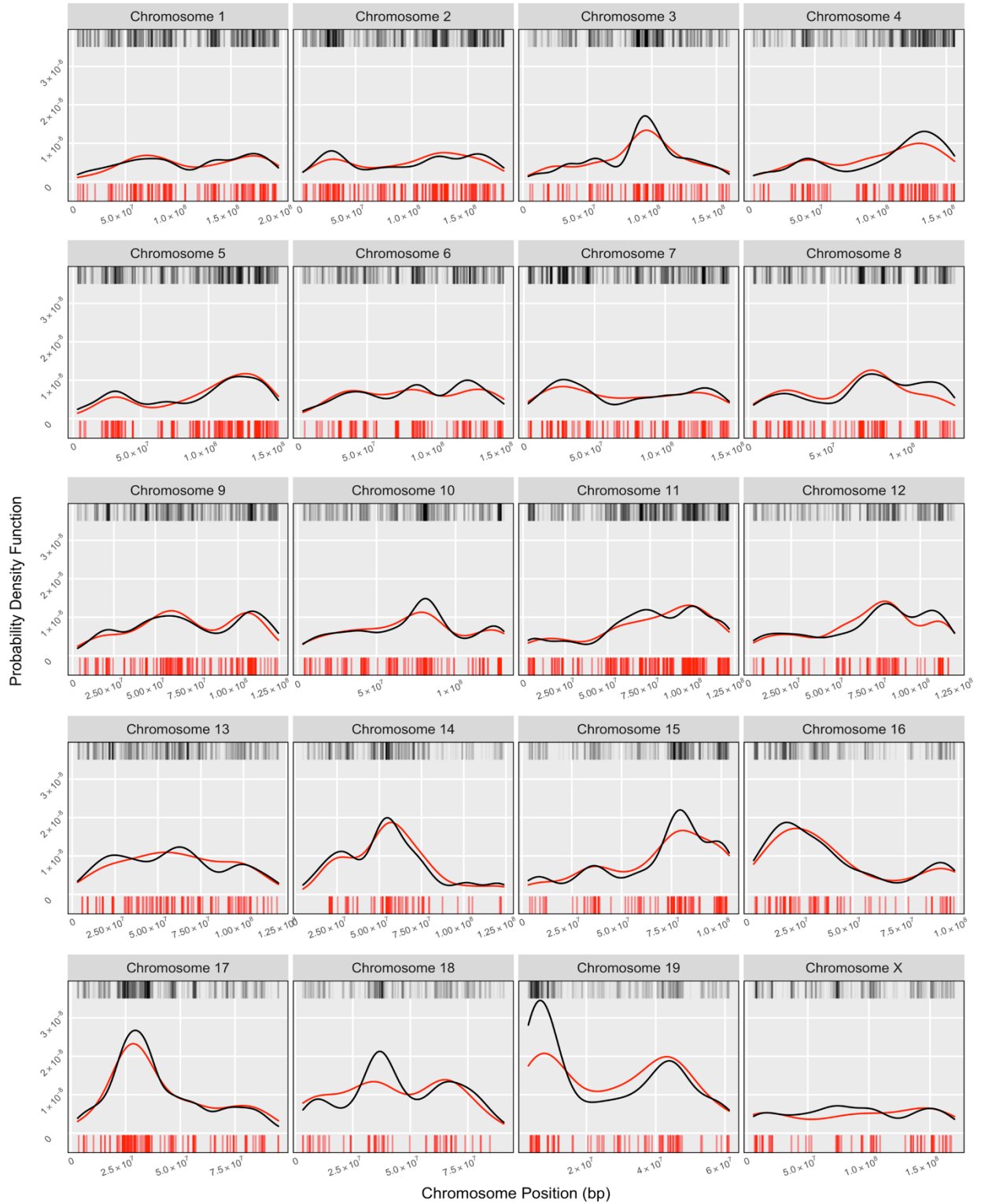
**Table 2.3** Gene expression of RNA degradation complex subunits in *Zcchc8* KO cells

	DESeq2		edgeR		Limma-Voom	
	log <sub>2</sub> (FC)	Adj. p-value	log <sub>2</sub> (FC)	Adj. p-value	log <sub>2</sub> (FC)	Adj. p-value
<b>NEXT</b>						
<i>Mtr4</i>	-0.26	0.013	-0.28	0.015	-0.20	0.075
<i>Rbm7</i>	-0.12	0.371	-0.13	0.354	-0.08	0.503
<b>PAXT</b>						
<i>Pabpn1</i>	0.25	0.347	0.23	0.442	0.16	0.532
<i>Rbm26</i>	-0.17	0.029	-0.18	0.063	-0.12	0.146
<i>Rbm27</i>	0.12	0.231	0.10	0.419	0.17	0.031
<i>Zc3h3</i>	0.18	0.703	0.16	0.789	-0.04	0.923
<i>Zfc3h1</i>	0.14	0.231	0.12	0.352	0.20	0.045
<b>RNA Exosome</b>						
<i>Dis3</i>	-0.31	0.0006	-0.33	0.001	-0.26	0.011
<i>Dis3l</i>	-0.24	0.195	-0.26	0.163	-0.26	0.119
<i>Exosc1</i>	-0.37	0.0004	-0.39	0.001	-0.34	0.004
<i>Exosc2</i>	0.04	0.903	0.02	0.968	-0.02	0.937
<i>Exosc3</i>	0.21	0.430	0.19	0.512	0.04	0.894
<i>Exosc4</i>	-0.27	0.048	-0.28	0.062	-0.30	0.019
<i>Exosc5</i>	0.20	0.312	0.19	0.432	0.14	0.468
<i>Exosc7</i>	-0.18	0.567	-0.20	0.513	-0.19	0.424
<i>Exosc8</i>	-0.35	0.0009	-0.37	0.002	-0.32	0.006
<i>Exosc9</i>	-0.40	0.014	-0.41	0.015	-0.41	0.011
<i>Exosc10</i>	-0.06	0.651	-0.07	0.598	-0.02	0.886

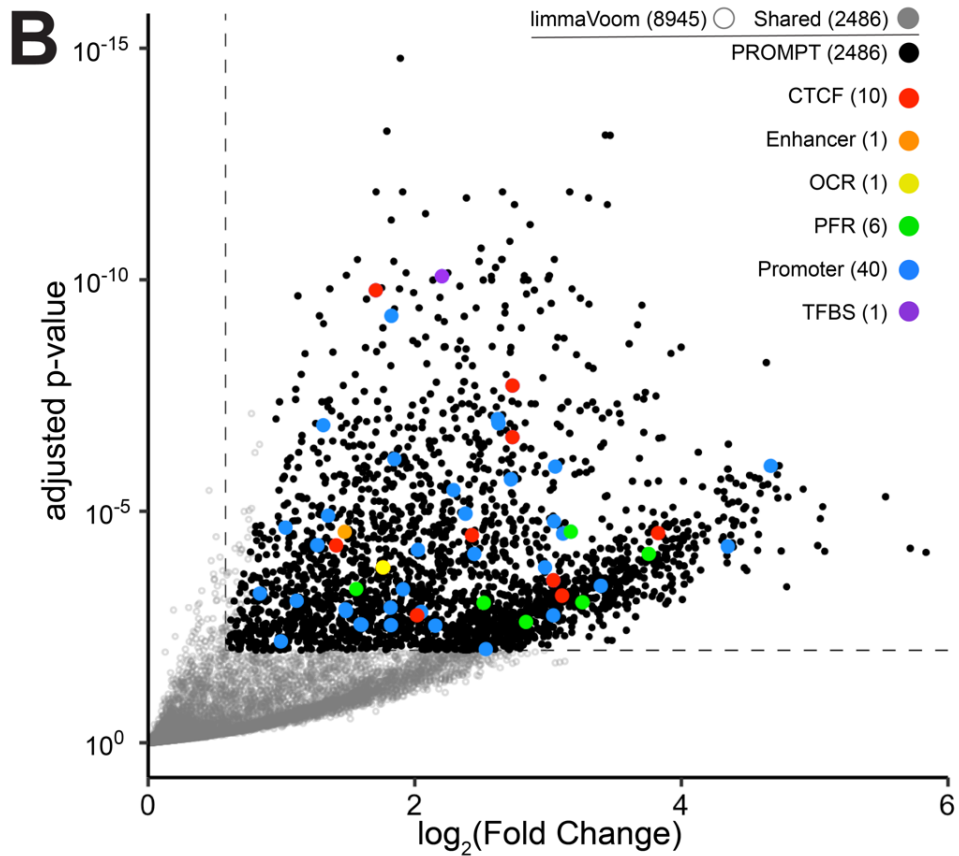
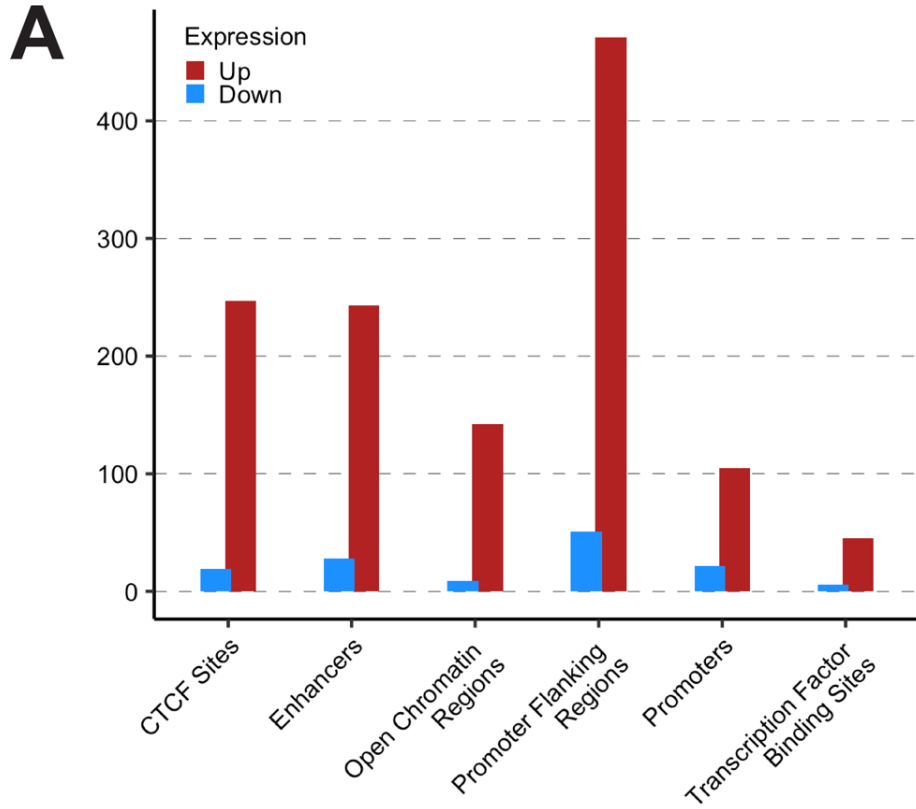
Gene expression data for RNA degradation complex subunits as computed using three different statistical analyses: DESeq2, edgeR, and Limma-Voom. Adj. p-value = adjusted p-value; FC = fold-change.



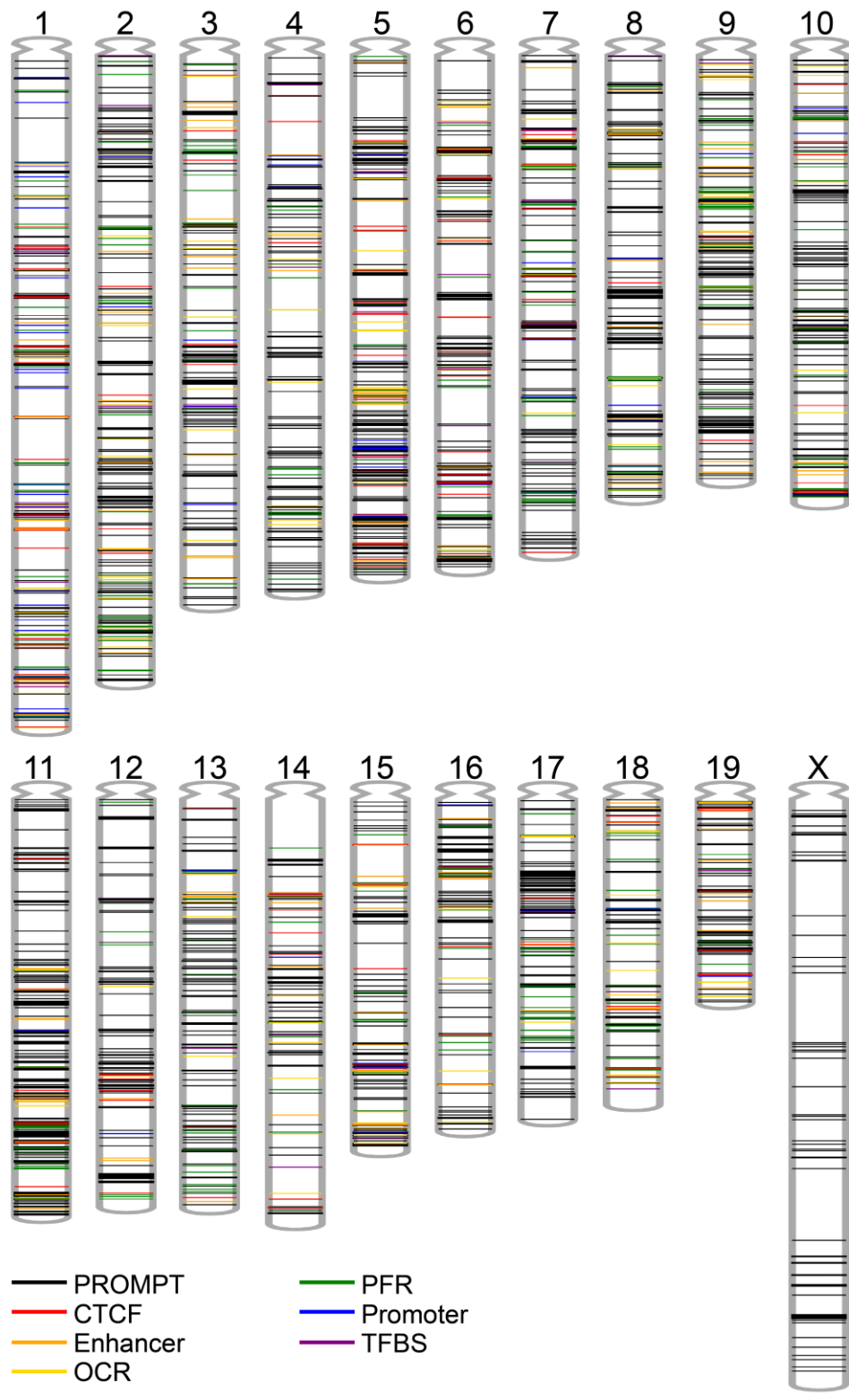
Genes: P.D.F. — Position  
PROMPTS: P.D.F. — Position



**Figure 2.9** Probability density function and rug plots indicating the distribution of PROMPTs closely matches the distribution of expressed genes along each chromosome. Rug plots mark individual genes (black) and PROMPTs (red) as vertical tick marks above and below, respectively. Probability density function curves for genes (black) and PROMPTs (red) were generated using the Gaussian kernel density estimate with Silverman's rule of thumb bandwidth method (Nrd0).



**Figure 2.10** (A) Bar chart showing the number of differentially expressed genomic regulatory features in SIMS *Zcchc8* knockout cells. (B) Scatter plot of PROMPTs and genomic regulatory features with overlapping genomic coordinates. Data were generated using the Limma-Voom statistical analysis. Horizontal and vertical dashed lines demarcate adjusted p-value of 0.01 and fold-change of 1.5 ( $\log_2(1.5) \approx 0.584$ ), respectively. Open circles indicate those PROMPTs that are specific to the Limma-Voom analysis. Closed circles indicate PROMPTs that are shared within DESeq2, edgeR, and Limma-Voom analyses and meet the significance thresholds of >1.5 fold-change and <0.01 adjusted p-value. Colored circles indicate genomic regulatory features with overlapping genomic coordinates to PROMPTs.



**Figure 2.11** Chromosomal distribution of differentially expressed PROMPTs and genomic regulatory features from SIMS *Zcchc8* knockout cells.

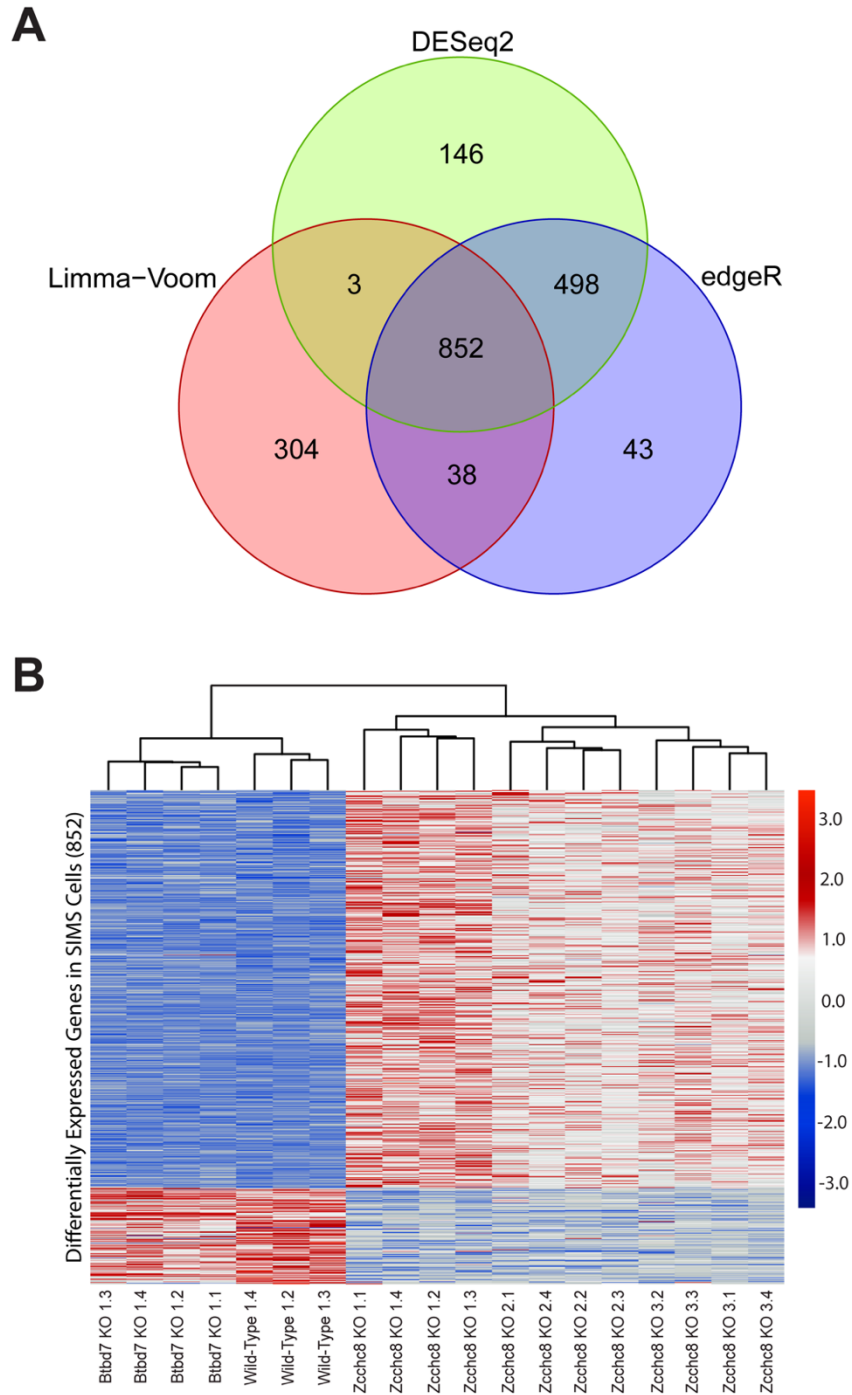
## Differential gene expression in *Zcchc8* knockout cells

We next turned our attention to differential gene expression using the same analyses we used for PROMPTs. Using the same thresholds of >1.5-fold expression change and adjusted p-value of <0.01, we found a total of 852 differentially expressed genes in our SIMS cells (Figure 2.12A). Of these genes, 712 were up-regulated in *Zcchc8* knockout cells while only 140 were down-regulated (Figure 2.12B). Intriguingly, of the 712 up-regulated genes in the *Zcchc8* knockout cells, 486 were RIKEN cDNAs or predicted genes (Figure 2.13A) affirming our exploratory PCA findings. Only 9 down-regulated RIKEN cDNAs and predicted genes were discovered (Figure 2.13A). Conversely, the *Btd7* knockout cells showed only 62 up-regulated and 80 down-regulated genes with only 2 and 8 of these consisting of RIKEN cDNAs or predicted genes, respectively (Figure 2.13B).

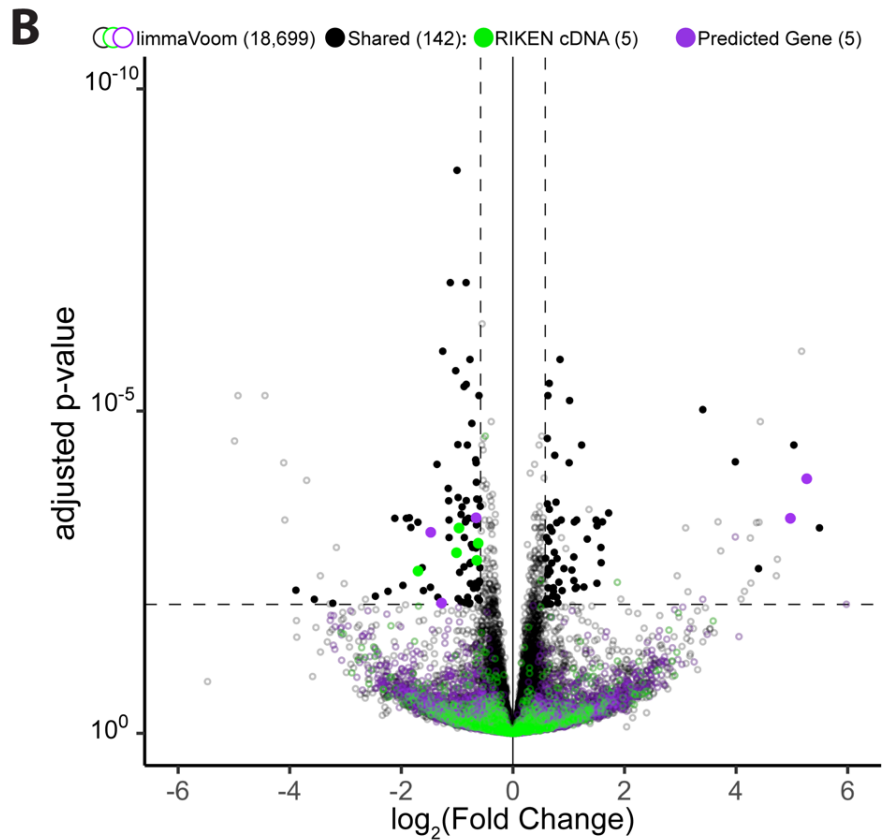
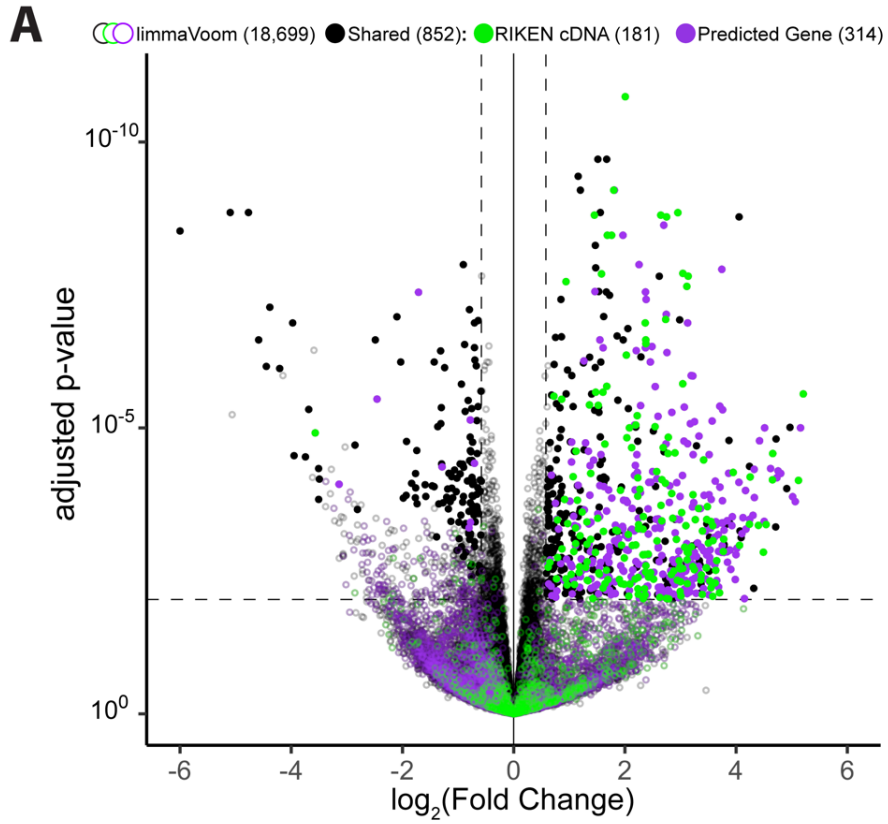
This unexpected finding stimulated us to examine the relationship between these RIKEN cDNAs and predicted genes and the genomic regulatory features governed by ZCCHC8 and the NEXT complex. Surprisingly, 334 of these predicted genes had intersecting genomic coordinates with the differentially expressed PROMPTs in our data set. Read coverage plots for selected RIKEN cDNAs and predicted genes revealed expression patterns that display the trademarks of PROMPTs (Figure 2.14).

Lastly, we asked if ZCCHC8 and/or the NEXT complex might directly regulate specific genes or pathways. We subjected the remaining differentially expressed genes to Signaling Pathway Impact Analysis (SPIA), PathNet, and Ingenuity Pathway Analysis in order to generate candidate genes and pathways. Our efforts at comparing these

candidates did not reveal any selective effects on the pathways we singled out for validation.

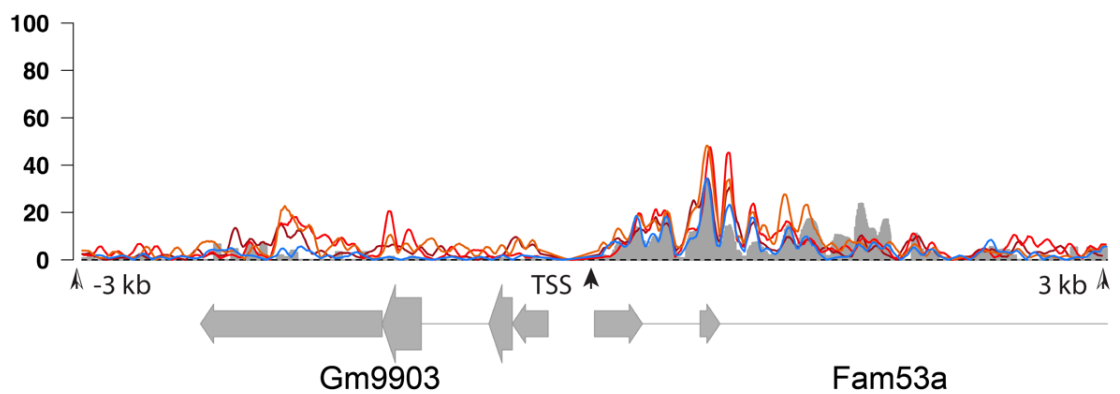
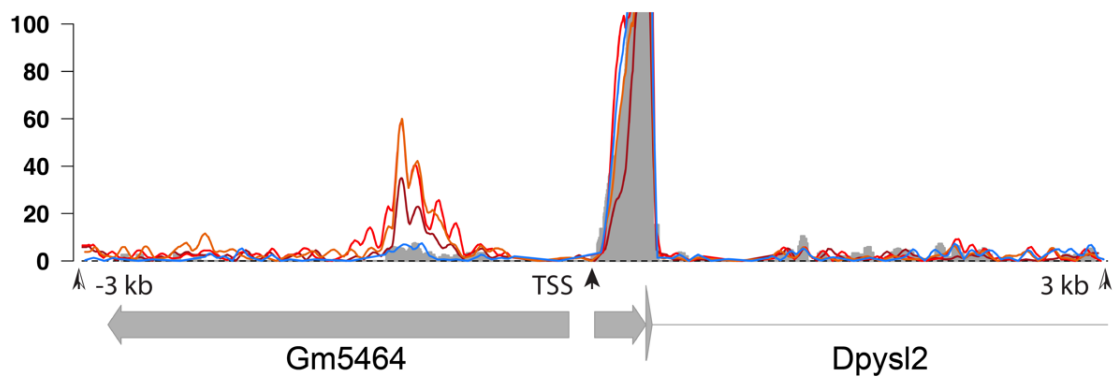
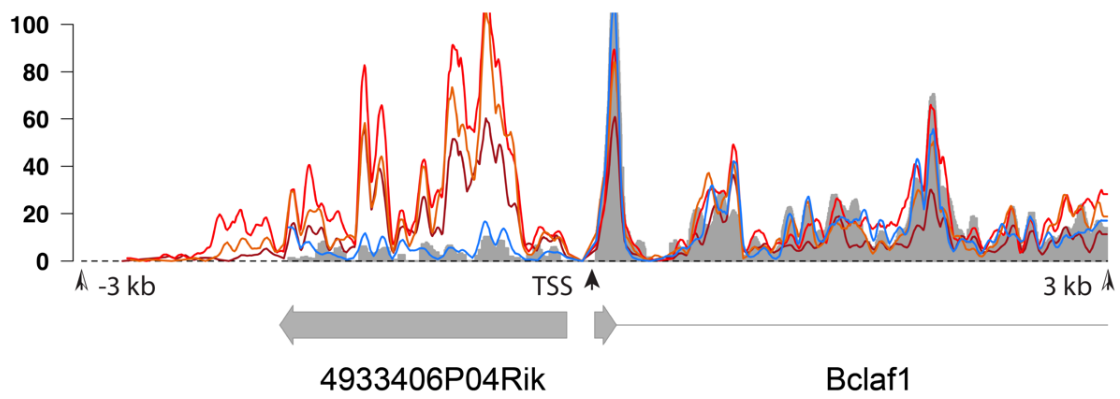
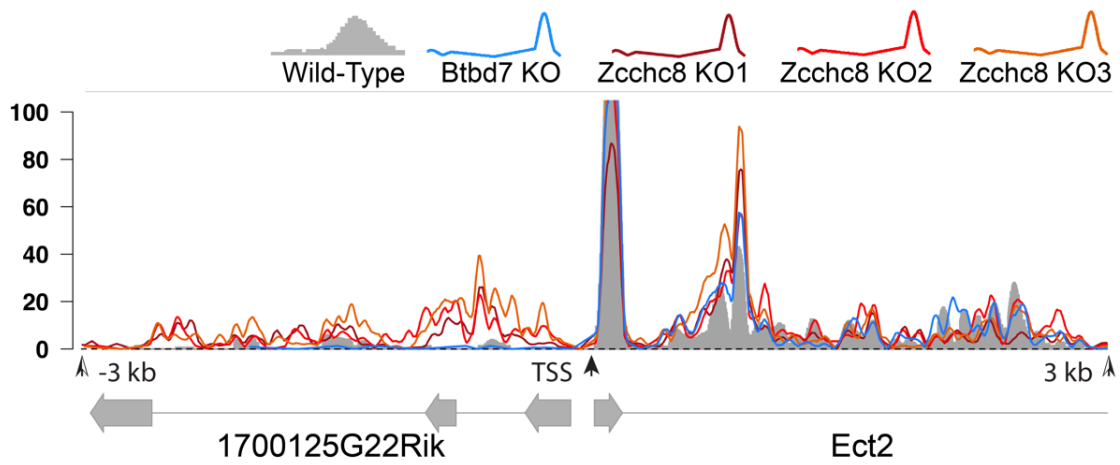


**Figure 2.12** (A) Venn diagram showing the number of differentially expressed genes determined via DESeq2, edgeR, and Limma-Voom analysis in SIMS *Zcchc8* knockout cells. (B) Heatmap using k-means clustering ( $k = 2$ ) of the 852 genes shared among the statistical analysis groups in (A).





**Figure 2.13** (A) Scatter plot of gene expression data in SIMS *Zcchc8* knockout cells. Data were generated using the Limma-Voom statistical analysis. Horizontal and vertical dashed lines demarcate adjusted p-value of 0.01 and fold-change of 1.5 ( $\log_2(1.5) \approx 0.584$ ), respectively. Open circles indicate those genes that are specific to the Limma-Voom analysis. Closed circles indicate genes that are shared within DESeq2, edgeR, and Limma-Voom analyses and meet the significance thresholds of >1.5 fold-change and <0.01 adjusted p-value. Colored circles indicate RIKEN cDNAs (green) and predicted genes (magenta). (B) Scatter plot of gene expression data in SIMS *Btbd7* knockout cells as in (A).



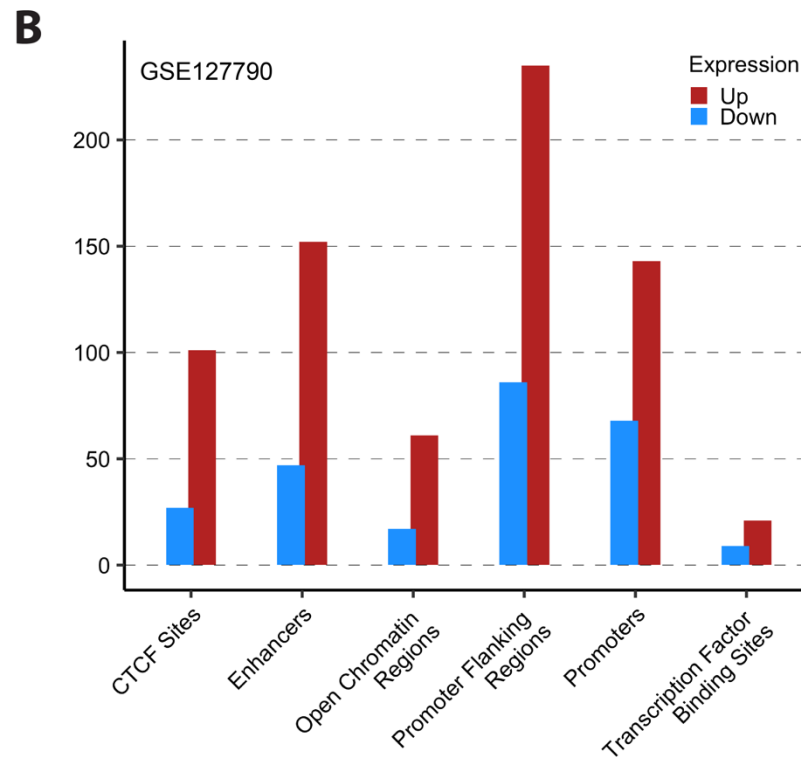
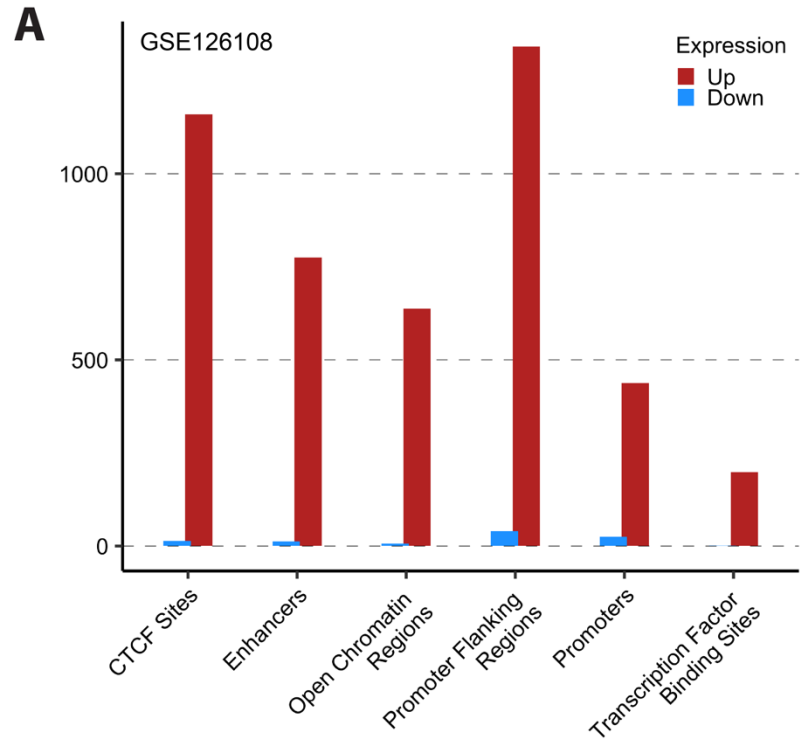
**Figure 2.14** Read coverage plots indicating the differential expression of RIKEN cDNAs and predicted genes upstream of protein coding genes. In order to create a single plot of all clonal replicates, read counts from replicate bam files were merged and indexed using the SAMtools software package followed by CPM-normalization prior to plotting using the Gviz R package. Black arrows indicate the TSS for each protein coding gene. Open arrows indicate the +/-3 kb position relative to the TSS. Note that the full length of the RIKEN cDNAs and predicted genes are shown while only the first 3 kb of the protein coding genes are shown. Read coverage y-axes were truncated at 100 reads in order to maintain scale.

## Effects of *Zcchc8* knockout in mice

Two other research groups recently developed *Zcchc8* knockout mice (44,45). Gable et al., performed RNA seq on E12.5 brains while Wu et al., performed RNA seq on ES cells derived from knockout mice. Using the publicly available raw data (NCBI GEO: GSE126108 and GSE127790) from these experiments, we performed our same differential expression analyses to assess the global effects of *Zcchc8* disruption.

Interestingly, we found 4,740 PROMPTs in E12.5 brains and only 422 PROMPTs in ES cells. We found our SIMS cells had 1,772 PROMPTs in common with E12.5 brains and only 138 PROMPTs in common with ES cells. There were 132 PROMPTs common to all three groups.

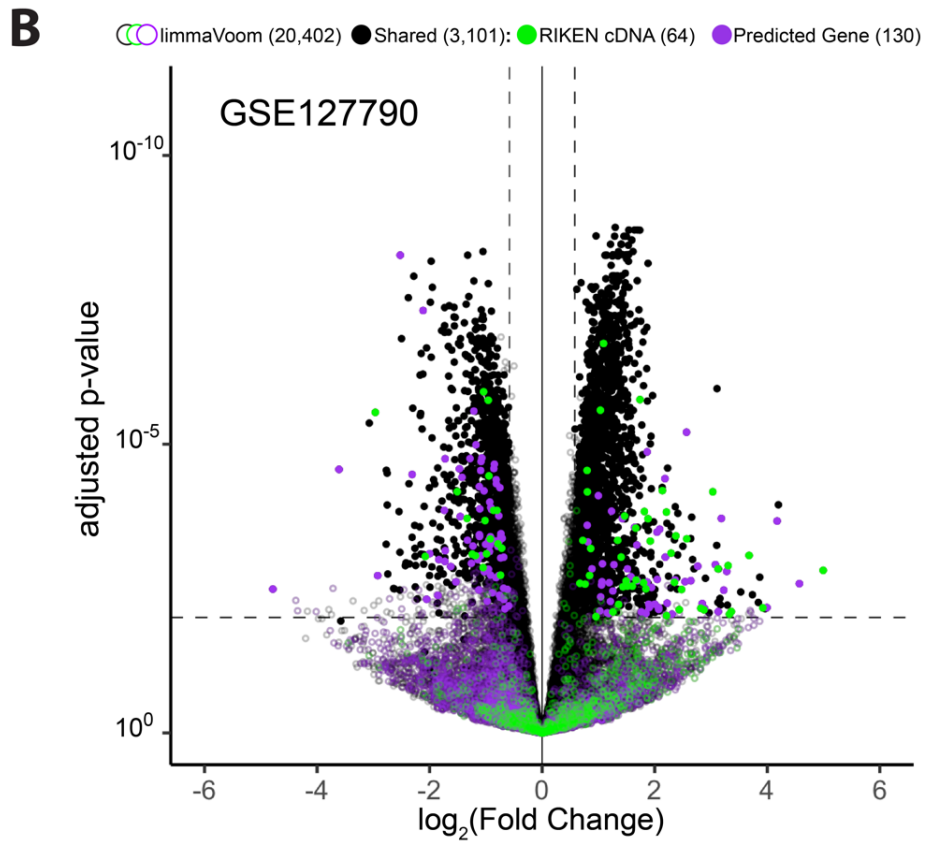
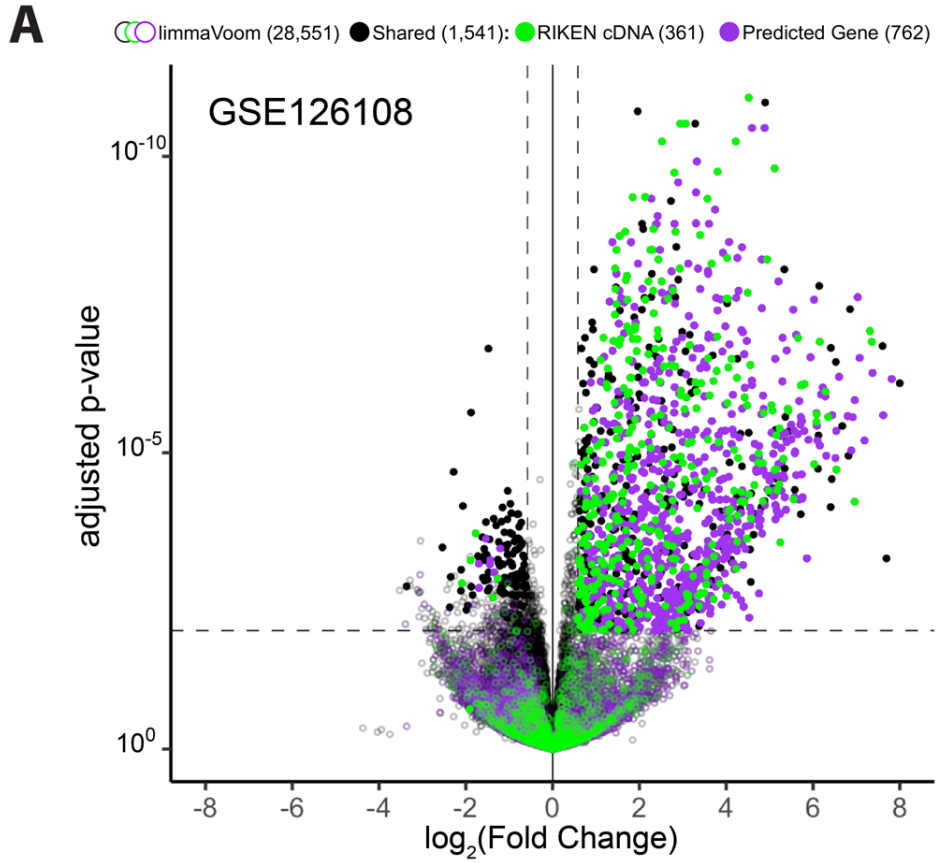
Considering our establishment of the role of ZCCHC8 and/or the NEXT complex in regulating pervasive transcription of genomic regulatory features in SIMS *Zcchc8* knockout cells, we asked if our findings would be recapitulated in the knockout mouse datasets. Indeed, in E12.5 brains (GSE12618) we found ZCCHC8 ablation resulted in the transcriptional up-regulation of 1,159 CTCF binding sites, 775 enhancers, 638 open chromatin regions, 438 promoters, 1,342 promoter flanking regions, and 198 transcription factor binding sites (>1.5-fold expression difference and <0.01 adjusted p-value) (Figure 2.15A). Likewise, in ES cells (GSE127790) we found 101 CTCF binding sites, 152 enhancers, 61 open chromatin regions, 143 promoters, 235 promoter flanking regions, and 21 transcription factor binding sites (>1.5-fold expression difference and <0.01 adjusted p-value) (Figure 2.15B).



**Figure 2.15** (A) Bar chart showing the number of differentially expressed genomic regulatory features in E12.5 brains taken from *Zcchc8* knockout mice (GSE126108). (B) Bar chart showing the number of differentially expressed genomic regulatory features in ES cells taken from *Zcchc8* knockout mice (GSE127790).

We further explored if our discovery of the relationship between PROMPTs, RIKEN cDNAs and predicted genes would also be recapitulated in the knockout mouse datasets. First, we found 1,541 differentially expressed genes in E12.5 brains with 1,410 up-regulated and only 131 down-regulated in knockouts (Figure 2.16A). We found 3,101 differentially expressed genes in ES cells with 1,804 up-regulated and 1,297 down-regulated (Figure 2.16B). In comparison, our SIMS cells had 450 up-regulated genes and only one down-regulated gene in common with E12.5 brains; 89 up-regulated genes and only seven down-regulated genes in common with ES cells. Among the three knockout groups, there were 69 common up-regulated genes and zero down-regulated genes in common.

Once again closer inspection revealed a large proportion of the up-regulated genes in the *Zcchc8* knockout groups were RIKEN cDNAs and predicted genes. Remarkably, of the 1,410 up-regulated genes in E12.5 brains, 1,123 are RIKEN cDNAs or predicted genes (Figure 2.16A) with 645 having overlapping genomic coordinates with PROMPTs. Further, of the 450 common up-regulated genes between our SIMS cells and the E12.5 brains, 372 were RIKEN cDNAs or predicted genes. In an intriguing contrast, only 102 of the 1,804 up-regulated genes in ES cells were RIKEN cDNAs or predicted genes (Figure 2.16B).



**Figure 2.16** (A) Scatter plot of gene expression data in E12.5 brains taken from *Zcchc8* knockout mice (GSE126108). Data were generated using the Limma-Voom statistical analysis. Horizontal and vertical dashed lines demarcate adjusted p-value of 0.01 and fold-change of 1.5 ( $\log_2(1.5) \approx 0.584$ ), respectively. Open circles indicate those genes that are specific to the Limma-Voom analysis. Closed circles indicate genes that are shared within DESeq2, edgeR, and Limma-Voom analyses and meet the significance thresholds of >1.5 fold-change and <0.01 adjusted p-value. Colored circles indicate RIKEN cDNAs (green) and predicted genes (magenta). (B) Scatter plot of gene expression data in ES cells taken from *Zcchc8* knockout mice (GSE127790) as in (A).



## Discussion

This study provides a comprehensive examination of the effects of ZCCHC8 ablation on the mouse transcriptome. Our results show that ZCCHC8 and/or the NEXT complex regulates a larger family of ncRNA species than previously reported. We show that disruption of *Zcchc8* results in the up-regulation of transcriptional products in multiple genomic regulatory regions including CTCF binding sites, enhancers, open chromatin regions, promoters, promoter flanking regions, and transcription factor binding sites. Further, we have shown >95% of these genomic regions have non-overlapping genomic coordinates with PROMPTs in our SIMS *Zcchc8* knockout cells. Thus, we conclude that, in addition to established substrates, ZCCHC8 and/or the NEXT complex serve to regulate pervasive transcription of CTCF binding sites, open chromatin regions, promoters, promoter flanking regions, and transcription factor binding sites. To our knowledge, this is the first such comprehensive report.

Surprisingly, we found ZCCHC8 ablation resulted in a large increase in the expression of RIKEN cDNAs and predicted genes. Given the transcriptional coverage, profile, and orientation of these transcripts, our results suggest that a significant number of RIKEN cDNAs and predicted genes are, in fact, PROMPTs (though ultimately experimental confirmation will be required). Using rather stringent criteria, we found 334 RIKEN cDNAs and predicted genes in our SIMS cells, and 645 in E12.5 brains, that display the hallmarks of PROMPTs. Considering that ZCCHC8 and the NEXT complex primarily serve to regulate pervasive transcription of a large number of ncRNA families, it stands to reason that many of the remaining up-regulated RIKEN cDNAs and predicted genes that are non-overlapping with PROMPTs may also be pervasive

transcripts from other ncRNA families. Less likely, the possibility remains that many of these RIKEN cDNAs and predicted genes constitute a specific set of functional genes that are negatively regulated in large part by ZCCHC8 and/or the NEXT complex.

As expected, the overall genomic expression profile significantly differed between SIMS cells, E12.5 brains, and ES cells. Interestingly, despite the obvious difference between mouse salivary glands cells and E12.5 brains, these groups shared 1,772 PROMPTs and 372 predicted genes and RIKEN cDNAs while relatively little overlap exists with these groups and ES cells. Additionally, our SIMS cells and the E12.5 brains had ~5.9 and 11.2 times more PROMPTs, respectively, and ~4.8 and 11.0 times more up-regulated RIKEN cDNAs and predicted genes, respectively, than ES cells.

Obviously, these results are indicative of the larger transcriptomic differences among cells of varying differentiation potential and suggest that epigenomic and heterochromatic organization of the genome during these states may play a role in PROMPT expression and/or degradation. Considering that it is generally accepted that ES cells have more open, plastic chromatin with reduced nucleosome density than differentiated cells, it is seemingly counterintuitive that the more differentiated state of SIMS cells and E12.5 brains may better lend itself to conditions suitable for increased transcription of PROMPTs. Nevertheless, recent research using a highly sensitive chemical mapping of nucleosome organization in mouse ES cells has shown that, contrary to the prevailing model, nearly all genes have a class of “fragile” nucleosomes occupying previously designated nucleosome-depleted regions upstream of transcription start sites (46). Further, this nucleosome mapping showed a high degree of nucleosome occupancy at CTCF sites (46). Though this finding is currently

considered controversial it has clear implications for potentially suppressing PROMPT and genomic regulatory feature transcription in ES cells.

Alternatively, it is speculated that as RNA polymerases elongate the sense transcript, negative supercoiling of the DNA upstream of the TSS can prime antisense transcription initiation (13). Perhaps, the epigenomic and heterochromatic landscape of more differentiated cells is more conducive to negative supercoiling that would prime such antisense transcription in a narrow window upstream of the TSS.

Regardless of the mechanism, it is clear that ES cells have fewer PROMPTs than our SIMS cells and the E12.5 brains. Perhaps this should not come as a surprise considering the potential detrimental effects pervasive transcripts, like PROMPTs, may have on the tight regulation required to maintain stemness. It should also not be surprising if future research were to uncover a more generalized, global regulatory system for preventing these unwanted transcripts in ES cells. To potentially shed light on the mechanism, it would be interesting to profile the DNA methylation state, nucleosome density, and nucleosome modification status of the PROMPT regions shared in SIMS cells and E12.5 brains versus the ES cells.

# MATERIALS AND METHODS

## Cell Culture and Maintenance

Wild-type, *Btbd7*- and *Zcchc8*-knockout SIMS cells were maintained in Phenol red-free DMEM (GE Healthcare/Cytiva, SH30284.01) supplemented with 10% fetal bovine serum (FBS; GE Healthcare/Cytiva, SH30070.03) and incubated at 37°C with 10% CO<sub>2</sub>. Cells were passaged every three to four days using trypsin-EDTA (Thermo Fisher, 25300120) after rinsing with HBSS (Thermo Fisher, 14170161). Cell density was determined using an automated cell counter (Nexcelom Cellometer Auto 2000).

## Biological Assays

### *MTT proliferation*

Cell proliferation assays were performed using the Cell Proliferation Kit I (MilliporeSigma, 11465007001). Approximately  $1 \times 10^3$  wild-type, *Btbd7*- and *Zcchc8*-knockout SIMS cells were plated in six wells each of five, 96-well plates. At 24 hours intervals post-plating, 10  $\mu$ l of MTT reagent was added to each well of a single 96-well plate followed by incubation at 37°C for 4 hours. 100  $\mu$ l of Solubilization solution was added to each and the plates were incubated overnight at 37°C. Absorbances were measured on a Cytation 5 (BioTek) imaging reader at 570 nm wavelength.

### **Colony formation assays**

Approximately 500 wild-type, *Btbd7*- and *Zcchc8*-knockout SIMS cells were seeded into four wells of a 24-well plate. The cell media was replaced every other day. After 10 days, colonies were fixed with 0.5 ml of 4% PFA for 20 min, washed once with PBS, and stained with 0.5 ml of 0.5% crystal violet for 10 min. Colonies were washed 5 times with 2 ml of distilled H<sub>2</sub>O. Colonies with more than ~50 cells were counted under a microscope.

### **Soft agar colonization assays**

0.5 ml of solubilized 0.7% sterile noble agar (VWR, 90000-772) in complete DMEM was added to each well of 24-well plate and allowed to solidify at room temperature in the tissue culture hood. wild-type, *Btbd7*- and *Zcchc8*-knockout SIMS cells were suspended in 0.3% sterile noble agar in complete DMEM at concentrations of  $\sim 1 \times 10^4$  cells per 0.5 ml. Quickly, 0.5 ml of cell suspension was added to four wells of 24-well plate containing the solidified agar. The top layer was allowed to solidify at room temperature in the tissue culture hood followed by addition of 250  $\mu$ l of DMEM supplemented with 30% FBS. Colonies were counted under a dissecting microscope after three weeks growth in a humidified incubator at 37°C with 10% CO<sub>2</sub>.

### **Scratch assays**

Approximately  $1 \times 10^5$  cells from SIMS wild-type, *Btbd7*-, and *Zcchc8*-knockout clones were seeded in triplicate into 35 mm MatTek dishes (MatTek, P35G-1.5-20-C) and grown until confluent. A p20 pipette tip was used to make a scratch in a straight

line through the center of the glass surface of the MatTek dish. The cells were washed once with HBSS and fresh media was added to the dishes. An orientation mark perpendicular to the scratch was placed on the bottom surface of the dishes and images were taken at the 0-hour timepoint. The cells were placed in a humidified incubator at 37°C with 10% CO<sub>2</sub> and grown overnight. At the 16-hour timepoint the dishes were oriented under a microscope and wound healing progress was imaged.

### **CRISPR/Cas9 Knockout of *Zcchc8* in SIMS Cells**

The *Zcchc8* KN2.0, non-homology mediated mouse gene knockout kit (KN519669) was purchased from OriGene. Either the pCas-Guide CRISPR vector (OriGene, KN519669G1) containing a single guide RNA target sequence 5'-TAGGTCGCCAAAATCCACAC-3' or the pCas-Guide CRISPR vector (OriGene, KN519669G2) containing a single guide RNA target sequence 5'-CGAGGCGTTTGACCCACCAG-3' or a combination of the two was transfected into the mouse submandibular salivary gland cell line SIMS using Thermo Fisher's Lipofectamine 3000 Reagent kit in the following manner. Approximately  $1 \times 10^5$  SIMS cells were plated in each well of a 6-well plate the day prior to transfection. On the day of transfection, either 3.75 or 7.5  $\mu$ l of Lipofectamine 3000 reagent was diluted into 125  $\mu$ l of serum-free Opti-MEM in separate 1.5 ml microfuge tubes for each guide RNA or the combination. In separate 1.5 ml microfuge tubes, 1.0  $\mu$ g of either pCas-Guide CRISPR vector or the combination, 1.0  $\mu$ g of linear donor cassette with EF1a promoter followed by eGFP-P2A-Puromycin resistance (OriGene, KN519669D) and 2.0  $\mu$ l of P3000 reagent per  $\mu$ l DNA was diluted into 125  $\mu$ l of serum-free Opti-MEM. The diluted

DNA mixtures were then added to the respective Lipofectamine microfuge tubes and the reactions were incubated for 15 minutes at room temperature. After incubation, the DNA-lipid mixtures were added drop-wise to the SIMS cells in the respective individual wells. After 48 hours post-transfection, the cells were split 1:10 into DMEM (Thermo Fisher, 11965118) + 10% FBS every three days for a total of 4 passages. The cells were then grown in the selective medium DMEM + 10% FBS + 2  $\mu$ g/ml puromycin (MilliporeSigma, P8833) for approximately one month. A subset of cells was harvested to check for *Zcchc8* knockout efficiency via western blot, immunofluorescence, and PCR. At this time, it was determined that the greatest knockout efficiency had been achieved using the pCas-Guide CRISPR vector with guide RNA sequence of 5'-TAGGTCGCCAAAATCCACAC-3' (OriGene, KN519669G1). These cells were chosen to produce individual clones using standard single cell cloning techniques in 96-well plates.

## Western Blots

Approximately  $1 \times 10^6$  cells from SIMS wild-type, *Btbd7*<sup>-</sup>, and *Zcchc8*-knockout clones were seeded into 10 cm dishes. At ~75% confluence, cells were washed with pre-chilled PBS followed by the addition of 500  $\mu$ l of pre-chilled RIPA buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1 $\times$  Halt Protease and Phosphatase Inhibitor Cocktail (Thermo Fisher, 78444). Cells were scraped on ice and collected in pre-chilled 1.5 ml tubes (Eppendorf, 022363212). The cell suspensions were incubated on ice for 30 minutes followed by centrifugation at 13,000 rpm for 15 min at 4°C. Supernatants were

transferred to pre-chilled 1.5 ml tubes and stored at -20°C. Lysates were quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher, 23227). Aliquots of 25  $\mu$ g lysate were denatured in 1X Laemmli sample buffer (Bio-Rad, 1610747) at 99°C for 5 min. Using a Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell, lysates and 5  $\mu$ l of Precision Plus Protein Kaleidoscope Standards (Bio-Rad, 1610375) were run on Bio-Rad 7.5% Mini-PROTEAN TGX precast gels with Tris/Glycine/SDS (Bio-Rad, 1610732) running buffer at 115 V followed by transfer to Bio-Rad Trans-Blot Turbo 0.2  $\mu$ m nitrocellulose via a Bio-Rad Trans-Blot Turbo Transfer system. Membranes were then incubated in blocking solution consisting of 5% nonfat dry milk in TBST (Tris Buffered Saline with 0.5% Tween-20; Quality Biological, 351-086-101; MilliporeSigma, P2287) for 1 hour at room temperature followed by incubation with primary antibodies diluted in blocking solution overnight at 4°C. Membranes were washed 3 times for 15 min each in TBST and incubated with LI-COR secondary antibodies diluted in Blocking Solution for 1 hour at room temperature protected from light. The membranes were then washed 3 times for 15 min each in TBST at room temperature and imaged on a LI-COR Odyssey CLx imaging system controlled by the LI-COR Image Studio software. The primary antibody dilutions were 1:10,000 anti-Zcchc8 (Proteintech, 23374-1-AP), 1:10,000 anti- $\alpha$ -tubulin (Millipore Sigma, T6199). The secondary antibody dilutions were 1:10,000 680RD goat anti-mouse (LI-COR, 926-68070) and 1:10,000 800CW goat anti-rabbit (LI-COR, 926-32211).



## Immunofluorescence and Confocal Microscopy

Approximately  $1 \times 10^5$  cells from SIMS wild-type, *Btbd7*<sup>-</sup>, and *Zcchc8*-knockout clones were seeded into 35 mm MatTek dishes (MatTek, P35G-1.5-20-C) and grown for 48 hours. Cell media was removed and cells were fixed with 2 ml of 4% PFA in PBS (Electron Microscopy Sciences, 15710) for 20 min at room temperature. Cells were then quickly washed with 2 ml of PBS followed by permeabilization with 1 ml of PBS with 0.1% Triton X-100 (Thermo Fisher, 28314) for 10 min at room temperature. Cells were then washed with 2 ml of wash buffer (PBS + 0.5% Tween-20) and then blocked for 1 hour with 1 ml of blocking buffer (wash buffer containing 3% fatty acid free BSA (Thermo Fisher, 126609)). The cells were then incubated with 200  $\mu$ l of primary antibodies diluted in blocking buffer overnight at 4°C. The cells were then washed with 1 ml of wash buffer three times for 5 min each and incubated with 100  $\mu$ l of secondary antibodies diluted in blocking buffer for 1 hour at room temperature. Cells were then washed with 1 ml of wash buffer three times for 5 min each. After the final wash, 12  $\mu$ l of Fluoro-Gel II with DAPI mounting medium (Electron Microscopy Sciences, 17985-50) was added to the cells on the glass surface of the MatTek dish. A coverslip was then sealed over the glass surface to protect the cells from damage. The cells were then imaged using either a 40 $\times$  1.1 NA or 60 $\times$ , 1.2 NA objective on a Zeiss LSM 880 system controlled by Zeiss ZEN software. The primary antibody dilution was 1:500 anti-Zcchc8 (Proteintech, 23374-1-AP). The secondary antibody dilution was 1:200 Alexa Fluor 647 donkey anti-rabbit (Jackson ImmunoResearch, 711-606-152). Additionally, cells were stained with a 1:200 dilution of Rhodamine Phalloidin (Thermo Fisher, R415) during the secondary antibody incubation step.

## RNA Extraction and Sequencing Library Preparation

Approximately  $1 \times 10^5$  cells from SIMS wild-type, *Btbd7*<sup>-</sup>, and *Zcchc8*-knockout clones were seeded into 60 mm dishes and grown until cells reached ~75% confluence. Cell media was removed and 0.5 ml of TRIzol Reagent (Thermo Fisher, 15596026) was added directly to the dishes. After brief trituration, the lysates were collected into 1.5 ml tubes and incubated for 5 min at room temperature. Next, 0.1 ml of chloroform was added to each lysate and mixed thoroughly by inverted the tubes multiple times. The lysates were incubated at room temperature for three minutes and then centrifuged at 13,000 rpm for 15 min at 4°C. The aqueous phase of each lysate was then transferred to separate, fresh 1.5 ml tubes and 0.25 ml of 70% ethanol was added to each tube. The samples were then transferred to RNeasy spin columns and 2 ml collection tubes from an RNeasy Mini Kit (Qiagen, 74104) and centrifuged at 13,000 rpm for 15 sec. The flow-through was discarded and on-column DNase digestion was performed using the RNase-Free DNase Set (Qiagen, 79254) and the following protocol from Qiagen: 350  $\mu$ l of buffer RW1 was added to each column followed by centrifugation at 13,000 rpm for 15 sec. The flow-through was discarded. 80  $\mu$ l of DNase I incubation mix (10  $\mu$ l DNase I stock solution in 70  $\mu$ l buffer RDD) was added to each column membrane and incubated for 15 min at room temperature. 350  $\mu$ l of buffer RW1 was added to each column followed by centrifugation at 13,000 rpm for 15 sec. The flow-through was discarded and 500  $\mu$ l of buffer RPE was added to each column followed by centrifugation at 13,000 rpm for 15 sec. Again, 500  $\mu$ l of buffer RPE was added to each column followed by centrifugation at 13,000 rpm for 2 min. The RNeasy columns were then placed in new 2 ml collection tubes and centrifuged for 13,000 rpm for 1 min. The

RNeasy spin columns were then placed in new 1.5 ml collection tubes and 50  $\mu$ l of RNase-free water was added to the column membranes followed by centrifugation at 13,000 rpm for 1 min. The RNA samples were then stored at -80°C.

RNA integrity was assessed using a Fragment Analyzer (Advanced Analytical) and sequencing libraries were prepared using the Illumina TruSeq Stranded Total RNA library preparation kit according to the manufacturer's protocols. Library preparation was performed on four replicates of each cell line. The samples were analyzed on an Illumina NextSeq500 configured for 40 paired-end reads and the quality of the reads was evaluated using FastQC software.

## **Read Mapping and Differential Expression Analysis**

In order to perform read mapping of PROMPTs and genomic regulatory features, the respective gtf/gff files were altered in order to generate the genome files for STAR read alignment. The Ensembl regulatory build gff file was filtered for each genomic regulatory feature and a new gff file was created for each. Each genomic feature was given a specific identifier in the feature identifier column of the respective new regulatory build gff files (ctcf, enhancer, ocr, promoter, pfr, tfbs). STAR genome generation was then conducted with the additional parameters --sjdbGTFfeatureExon feature and --sjdbGTFtagExonParentGene ID.

For PROMPTs, the primary assembly gtf file was filtered to remove all feature types except for the "gene" identifier. This identifier was changed to "prompt" and the genomic coordinates were altered to 3 kb upstream of the start site. STAR genome

generation was then conducted with the additional parameters `--sjdbGTFfeatureExon` prompt and `--sjdbGTFTagExonParentGene` ID.

Read mapping against the recent GENCODE mouse release M24 was performed using the STAR 2.7.3a aligner using the standard mode with mapping parameters derived from the GENCODE project. Additionally, read counting was performed using the `quantMode` utility of STAR.

Read counts were filtered to remove low expressing features (genes, PROMPTs, genomic regulatory features, etc., with  $<5$  counts in at least one sample). Differential expression was evaluated by three independent statistical methods (DESeq2, edgeR and Limma-Voom).

## **Bioinformatics and Data Analysis**

Data analysis was performed using the SAMtools (47), deepTools (48), PhenoGram (49), and R software packages. Scripts used for data analysis can be found at <https://github.com/collinsjw/Zcchc8-KO-pervasive-transcripts>.

## **Data Availability**

RNA sequencing data is deposited in NCBI GEO (accession number GSE165689).

# Bibliography

1. Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. (1997) Intergenic transcription and transinduction of the human  $\beta$ -globin locus. *Genes and Development*.
2. Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*.
3. Carninci, P., Wells, C., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K., Pavan, W., Pavesi, G. *et al.* (2005) The Transcriptional Landscape of the Mammalian Genome. *Science*.
4. Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Dutttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*.
5. Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*.
6. Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F. *et al.* (2012) Landscape of transcription in human cells. *Nature*.
7. F.S. Collins, E.S.L.J.R. and *et al.* (2004) International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature*, **431**, 931-945.
8. van Bakel, H., Nislow, C., Blencowe, B.J. and Hughes, T.R. (2010) Most "dark matter" transcripts are associated with known genes. *PLoS Biology*, **8**.
9. Lee Phillips, M. (2010) Existence of RNA 'dark matter' in doubt. *Nature*.
10. Robertson, M. (2010).
11. Robinson, R. (2010) Dark Matter Transcripts: Sound and Fury, Signifying Nothing? *PLoS Biology*.
12. Clark, M.B., Amaral, P.P., Schlesinger, F.J., Dinger, M.E., Taft, R.J., Rinn, J.L., Ponting, C.P., Stadler, P.F., Morris, K.V., Morillon, A. *et al.* (2011) The reality of pervasive transcription. *PLoS Biology*, **9**, 5-10.
13. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A. and Sharp, P.A. (2008) Divergent transcription from active promoters. *Science*.
14. Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R.R., Grimmond, S.M., Schroder, K. *et al.* (2009) Tiny RNAs associated with transcription start sites in animals. *Nature Genetics*.
15. Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, **22**, 1775-1789.

16. Jensen, T.H., Jacquier, A. and Libri, D. (2013) Dealing with pervasive transcription. *Molecular Cell*, **52**, 473-484.
17. Kopp, F. and Mendell, J.T. (2018).
18. Nair, L., Chung, H. and Basu, U. (2020) Regulation of long non-coding RNAs and genome dynamics by the RNA surveillance machinery. *Nature Reviews Molecular Cell Biology*, **21**, 123-136.
19. Ransohoff, J.D., Wei, Y. and Khavari, P.A. (2018).
20. Ma, L., Bajic, V.B. and Zhang, Z. (2013) On the classification of long non-coding RNAs. *RNA Biology*, **10**, 924-933.
21. de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L. and Natoli, G. (2010) A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biology*.
22. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*.
23. Sigova, A.A., Mullen, A.C., Molinie, B., Gupta, S., Orlando, D.A., Guenther, M.G., Almada, A.E., Lin, C., Sharp, P.A., Giallourakis, C.C. *et al.* (2013) Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*.
24. Ulitsky, I. and Bartel, D.P. (2013) lincRNAs: Genomics, Evolution, and Mechanisms. *Cell*, **154**, 26-46.
25. Flynn, R.A., Almada, A.E., Zamudio, J.R. and Sharp, P.A. (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proceedings of the National Academy of Sciences*, **108**, 10460-10465.
26. Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R. *et al.* (2013) Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. *Nature Structural and Molecular Biology*, **20**, 923-928.
27. Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A. and Jensen, T.H. (2011) PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. *Nucleic Acids Research*, **39**, 7179-7193.
28. Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H. and Jensen, T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851-1854.
29. Ogami, K., Chen, Y. and Manley, J.L. (2018) RNA surveillance by the nuclear RNA exosome: mechanisms and significance. *Noncoding RNA*, **4**.
30. Belair, C., Sim, S. and Wolin, S.L. (2018) Noncoding RNA Surveillance: The Ends Justify the Means. *Chemical Reviews*, **118**, 4422-4447.
31. Laffleur, B. and Basu, U. (2019) Biology of RNA Surveillance in Development and Disease. *Trends in Cell Biology*, **29**, 428-445.
32. Lubas, M., Christensen, Marianne S., Kristiansen, Maiken S., Domanski, M., Falkenby, Lasse G., Lykke-Andersen, S., Andersen, Jens S., Dziembowski, A.

- and Jensen, Torben H. (2011) Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex. *Molecular Cell*, **43**, 624-637.
33. Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, Jens S., Sandelin, A. *et al.* (2016) Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. *Molecular Cell*, **64**, 520-533.
  34. Wu, G., Schmid, M., Rib, L., Polak, P., Meola, N., Sandelin, A. and Jensen, T.H. (2020) A Two-Layered Targeting Mechanism Underlies Nuclear RNA Sorting by the Human Exosome. *Cell Reports*, **30**, 2387-2401.e2385.
  35. Schmid, M. and Jensen, T.H. (2019). Springer, Vol. 1203, pp. 113-132.
  36. Schneider, C. and Tollervey, D. (2013) Threading the barrel of the RNA exosome. *Trends in Biochemical Sciences*, **38**, 485-493.
  37. Schuch, B., Feigenbutz, M., Makino, D.L., Falk, S., Basquin, C., Mitchell, P. and Conti, E. (2014) The exosome-binding factors Rrp6 and Rrp47 form a composite surface for recruiting the Mtr4 helicase. *The EMBO Journal*, **33**, 2829-2846.
  38. Silla, T., Schmid, M., Dou, Y., Garland, W., Milek, M., Imami, K., Johnsen, D., Polak, P., Andersen, J.S., Selbach, M. *et al.* (2020) The human ZC3H3 and RBM26/27 proteins are critical for PAXT-mediated nuclear RNA decay. *Nucleic Acids Research*, **48**, 2518-2530.
  39. Lubas, M., Andersen, P.R., Schein, A., Dziembowski, A., Kudla, G. and Jensen, T.H. (2015) The Human Nuclear Exosome Targeting Complex Is Loaded onto Newly Synthesized RNA to Direct Early Ribonucleolysis. *Cell Reports*, **10**, 178-192.
  40. Ogami, K., Richard, P., Chen, Y., Hoque, M., Li, W., Moresco, J.J., Yates, J.R., Tian, B. and Manley, J.L. (2017) An Mtr4/ZFC3H1 complex facilitates turnover of unstable nuclear RNAs to prevent their cytoplasmic transport and global translational repression. *Genes and Development*.
  41. Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H. *et al.* (2001) Functional annotation of a full-length mouse cDNA collection. *Nature*, **409**, 685-690.
  42. Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B. and Sharp, P.A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature*, **499**, 360-363.
  43. Lloret-Llinares, M., Mapendano, C.K., Martlev, L.H., Lykke-Andersen, S. and Jensen, T.H. (2016) Relationships between PROMPT and gene expression. *RNA Biology*, **13**, 6-14.
  44. Gable, D.L., Gaysinskaya, V., Atik, C.C., Conover Talbot, C., Kang, B., Stanley, S.E., Pugh, E.W., Amat-Codina, N., Schenk, K.M., Arcasoy, M.O. *et al.* (2019) ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. *Genes and Development*.
  45. Wu, Y., Liu, W., Chen, J., Liu, S., Wang, M., Yang, L., Chen, C., Qi, M., Xu, Y., Qiao, Z. *et al.* (2019) Nuclear Exosome Targeting Complex Core Factor Zcchc8 Regulates the Degradation of LINE1 RNA in Early Embryos and Embryonic Stem Cells. *Cell Reports*, **29**, 2461-2472.e2466.

46. Voong, L.N., Xi, L., Sebeson, A.C., Xiong, B., Wang, J.-P. and Wang, X. (2016) Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, **167**, 1555-1570.e1515.
47. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
48. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dünder, F. and Manke, T. (2016) deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*, **44**, W160-165.
49. Wolfe, D., Dudek, S., Ritchie, M.D. and Pendergrass, S.A. (2013) Visualizing genomic information across chromosomes with PhenoGram. *BioData Min*, **6**, 18.



# CURRICULUM VITAE

## PERSONAL DATA

Name **Joshua W. Collins**  
Country of Citizenship USA  
Birthplace Gardner, KS

## EDUCATION

2011-2021 Ph.D. in Biology, CMDB Program, NIH-Johns Hopkins Graduate Partnership Program, Baltimore, MD  
2004-2006 M.S. candidate in Mathematics, Pittsburg State University, Pittsburg, KS  
2001-2004 M.S. in Biology, Pittsburg State University, Pittsburg, KS  
1996-2000 B.S. in Biology, Pittsburg State University, Pittsburg, KS

## EMPLOYMENT HISTORY

2015-present Predoctoral IRTA, NIDCR, NIH, Bethesda, MD  
2011-2015 Predoctoral IRTA, NCI, NIH, Bethesda, MD  
2007-2011 Research Associate II, WCS, LMP, NCI, NIH, Bethesda, MD  
2006-2007 Scientist I, Xenotech, Lenexa, KS  
2005-2006 Lecturer/Research Associate, Pittsburg State University, Pittsburg, KS  
2003-2005 Graduate Teaching Assistant, Mathematics Department, Pittsburg State University, Pittsburg, KS  
2000-2003 Graduate Research Assistant, Biology Department, Pittsburg State University, Pittsburg, KS

## HONORS AND AWARDS

2016 NIDCR Fellows Scientific Training Poster Award  
2006 Graduate Award for Excellence of Research in Mathematics  
2004 Distinguished Thesis Award, 1<sup>st</sup> place winner  
2002 Graduate Student Thesis Grant Award  
2001 Graduate Award for Excellence of Research in Biology  
2001 Graduate Research Colloquium, 1<sup>st</sup> place winner  
2001 Graduate Student Thesis Grant Award  
2000 Graduate Student Thesis Grant Award

## Training and Career Development - Scientific Meetings

2021 Poster Presentation – Collins JW, Martin D, Wang S, Yamada KM. ZCCHC8 is a global regulator of pervasive transcription. 2021 RNA Biology Symposium Sponsored by the NCI RNA Biology Initiative; April 14-16, 2021.

2013 Poster contribution – Marino, N Collins J, Shen C, Sledge GW and Steeg PS. Analysis of gene expression patterns downstream of multiple metastasis suppressor genes. 9<sup>th</sup> International Congress of the NDP Kinase/Nm23/awd Gene Family; July 31-Aug 4, 2013. Boston University, Boston, MA.

2010 Poster contribution – Marino N, Marshall JC, Collins J, Veenstra T, Zhou M, Steeg, PS.: The actin-binding protein Gelsolin as a potential binding partner of Nm23. Joint Metastasis Research Society- American Association for Cancer Research Conference: Metastasis and the Tumor Microenvironment; 2010 Sep 12-15; Philadelphia, PA.

2010 Poster contribution – Marshall JC, Nakayama J, Collins J, Liewehr, DJ, Steinberg SM, Vidal-Vanaclocha F, Barbier M, Murone M, Steeg, PS.: An LPA1 antagonist, Debio 0719, acts as a metastasis suppressor in breast cancer. Joint Metastasis Research Society-American Association for Cancer Research Conference: Metastasis and the Tumor Microenvironment; 2010 Sep 12-15; Philadelphia, PA.

2010 Poster contribution – Marino N, Marshall JC, Collins J, Zhou M, Veenstra T, Steeg PS.: The identification of Gelsolin as a potential binding partner of Nm23. 101st Annual Meeting of the American Association for Cancer Research; 2010 Apr 17-21; Washington, DC.

2010 Poster contribution – Marshall JC, Collins J, Murone M, Barbier M, Steinberg S, Steeg PS.: LPA1 antagonist, Debio-0719, acts as a metastasis suppressor in breast cancer. 101st Annual Meeting of the American Association for Cancer Research; 2010 Apr 17 21; Washington, DC.

2009 Poster contribution – Marshall JC, Collins J, Zhou M, Veenstra T, Khanna C, Vidal-Vanaclocha F, Steeg PS.: An unbiased screen for Nm23-H1 interacting proteins identifies the cytoskeletal regulator Ezrin. 100th Annual Meeting of the American Association for Cancer Research; 2009 Apr 18-22; Denver, CO.

2008 Poster contribution - Horak CE, Marshall JC, Lee JH, Collins J, Weiland T, Steeg PS.: EDG2 signaling mediating tumor cell motility in highly metastatic cells. 99th Annual Meeting of the American Association for Cancer Research; 2008 Apr 12-16; San Diego, CA.

2008 Poster contribution - Marshall JC, Collins J, Miller KD, Steeg PS.: High dose medroxyprogesterone acetate (MPA) has anti-angiogenic properties. 99th Annual Meeting of the American Association for Cancer Research; 2008 Apr 12-16; San Diego, CA.

2004 Poster presentation - Collins, J., Long, J., Brooker, N.L. 2004. Plant Lipid Derivative Seed Treatments for Managing Fungal Soybean Diseases. The 16th International Symposium on Plant Lipids in Budapest, Hungary.

## **INTRAMURAL ACTIVITIES OUTSIDE THE NIDCR DIR**

2021	Science Policy Discussion Group, Office of Intramural Training and Education, NIH, Bethesda, MD
2017	Poster Judge for the 2017 NIH Postbac Poster Day, NIH, Bethesda, MD
2014	Poster Judge for the 2014 NIH Postbac Poster Day, NIH, Bethesda, MD
2012	Poster Judge for the 2012 NIH Postbac Poster Day, NIH, Bethesda, MD

## **Training and Career Development - Symposia, Workshops, and Seminars**

2021	Diversity and Inclusion Seminar Series, Office of Intramural Training and Education, NIH, Bethesda, MD
2021	Poster Presentation, NIH Annual Graduate Student Research Symposium, Bethesda, MD
2021	Deep Learning by Example on Biowulf, NIH High Performance Computing, Bethesda, MD
2021	Bash Shell Scripting, NIH High Performance Computing, Bethesda, MD
2020	Poster Presentation, NIH Annual Graduate Student Research Symposium, Bethesda, MD
2020	Python in HPC, NIH High Performance Computing, Bethesda, MD
2018	Poster Presentation, NIH Annual Graduate Student Research Symposium, Bethesda, MD
2018	RNA-Seq Analysis Training, NIH library, Bethesda, MD
2016	ChIP Sequencing Data Analysis NIH library, Bethesda, MD
2016	Poster Presentation, Annual Graduate Student Research Symposium, Bethesda, MD
2015	Data Wrangling in R, NIH library, Bethesda, MD
2015	Introduction to R and RStudio, NIH library, Bethesda, MD
2012-2021	Johns Hopkins-NIH Graduate Student Monthly Program Meeting
2012	Johns Hopkins Graduate Student Retreat, St. Michaels, MD
2008	Gene Expression Microarray Data Analysis, NCI, NIH
2008	Using Multiple Reference Genes and Other Factors to Obtain High-Quality Gene Expression Data, Presented by BioRad at NCI, NIH

## **Animal Study Protocols**

14-064	Evaluation for efficacy of a new therapeutic compound in a model of breast cancer metastasis
13-243	Metastasis suppression function of NM23-H1 in an animal model of breast cancer

## ACTIVITIES WITHIN THE NIDCR DIR

### Training and Career Development - Seminars and Workshops

2018-present NIDCR Cutting-Edge Technology Workshop Series

2016-present NIDCR Data Session Seminar Series

### Training and Career Development - Fellows Scientific Training

2019 Poster Presentation, NIDCR Fellows Scientific Training Annual Meeting, Washington, DC

2016 Poster Presentation, NIDCR Fellows Scientific Training Annual Meeting, Potomac, MD

## BIBLIOGRAPHY

### Peer-reviewed research articles

1. Cruz Walma DA and **Collins JW**. Western Blotting with Solutions containing Nanoliter Volumes of Antibody. **Curr Protoc Cell Biol**. 2019 Sep;84(1):e87.
2. Marino N, **Collins J**, Shen C, Caplen NJ, Merchant AS, Gokmen-Polar Y, Goswami C, Hoshino T, Qian Y, Sledge GW, Steeg PS. Identification and Validation of Genes with Expression Patterns Inverse to Multiple Metastasis Suppressor Genes in Breast Cancer Cell Lines. **Clin Exp Metastasis**. 2014 Oct; 31(7):771-86.
3. Marino N, Marshall JC, **Collins J**, Ming Z, Quian Y, Veenstra T and Steeg PS. Nm23-H1 binds to gelsolin and inactivates its actin-severing capacity to promote tumor cell motility and metastasis. **Cancer Res**. 2013 Oct 1;73(19):5949-62.
4. Marshall JC, **Collins J**, Nakayama J, Horak CE, Liewehr DJ, Steinberg SM, Albaugh M, Vidal- Vanaclocha F, Palmier D, Barbier M, Murone M, Steeg, PS.: Effect of inhibition of the lysophosphatidic acid receptor 1 on metastasis and metastatic dormancy in breast cancer. **J Natl Cancer Inst**. 2012 Sep 5;104(17):1306-19.

### Peer-reviewed review articles

1. Yamada KM, **Collins JW**, Cruz Walma DA, Doyle AD, Morales SG, Lu J, Matsumoto K, Nazari SS, Sekiguchi R, Shinsato Y, Wang S. 2019. Extracellular matrix dynamics in cell migration, invasion and tissue morphogenesis. **Int. J. Exp. Pathol**. 100(3):144-152. PubMed PMID: 31179622; PubMed Central PMCID: PMC6658910.
2. Marshall JC, **Collins J**, Marino N, Steeg PS.: The Nm23-H1 metastasis suppressor as a translational target. **Eur J Cancer**. 2010 May;46(7):1278-82.
3. Marino N, Nakayama J, **Collins J**, Steeg PS.: Insights into the biology and prevention of tumor metastasis provided by the Nm23 metastasis suppressor gene. **Cancer Metastasis Rev**. 2012 Dec;31(3-4):593-603.

### Chapters and proceedings

1. Gril B, **Collins J**, Taylor J, Rinker-Schaeffer C, Steeg PS, Marshall, JC. 2011. The role of metastasis suppressor genes in metastasis. **Cancer Metastasis: Biologic Basis and Therapeutics 1st Edition**. 6:64-76.

### Non-peer-reviewed research articles

1. **Collins JW**, Martin D, Genomics and Computational Biology Core, Wang S, Yamada KM. ZCCHC8 is required for the degradation of pervasive transcripts originating from multiple genomic regulatory features. **BioRxiv: The Preprint Server for Biology**. 2021 Feb; <https://doi.org/10.1101/2021.01.29.428898>.