THE UNIVERSITY OF CHICAGO

THE GENETIC AND TRANSCRIPTIONAL BASIS OF CUX1 ISOFORMS IN HEMATOPOIESIS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

COMMITTEE ON CANCER BIOLOGY

BY

MANISHA KRISHNAN

CHICAGO, ILLINOIS

DECEMBER 2021

To my family, partner, and friends

This achievement is dedicated to all of you

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Megan McNerney. This project ended up not panning out at all in the way we thought it would, and I felt discouraged a lot by the direction it was moving in. However, Megan always kept my morale up, and her vision has helped us get to the point of submitting a first author-paper on this work. She graciously took me into her lab at the beginning of my third year, when I had to leave my previous lab, no questions asked, for which I am endlessly grateful for. She has afforded me several professional development opportunities, such as the opportunity to present a poster at the Keystone Symposium on Hematopoiesis, and allowing me to mentor our newest graduate student in the lab. Megan is constantly an example of commitment and dedication to one's work, and I feel like I have blossomed as a scientist under her tutelage. I am confident that I have gained several hard and soft skills in graduate school that will equip me for a successful career, and this could not have been possible without Megan.

I would also like to thank my committee, Dr. Kay Macleod, Dr. James Labelle and Dr. Alex Ruthenburg. Your feedback and guidance through the years are what have gotten me to this stage. All of you had ideas for what would make the project more convincing. Kay, I am especially so grateful to you for helping me navigate the difficult situation of switching labs early on during my Ph.D. It was incredibly scary, and I could not have gotten here without your guidance and advice.

I would like to thank all the past and present members of the McNerney lab, who made my time at UChicago very enjoyable and productive. Jeff Kurkewich was my mentor during my rotation in the McNerney lab, and was always helpful and available when I needed help getting

my project off the ground. Ningfei An and Saira Khan were lab stalwarts that could answer almost any questions I had for them, and made sure that the lab functioned smoothly. Molly Imgruet and Stephanie Konecki, I am truly grateful for the camaraderie and companionship through much of my time in the McNerney lab as we struggled through experiments together. I am also grateful to both of you for interesting discussions about science and techniques, and lots of coffee hours and nights out after work. Matt Jotte, Weihan Liu, Tanner Martinez and Raven Watson, you all brought so much joy and cheer to the lab in the latter half of my Ph.D., and it has been a pleasure to see all your projects shape up over the year. Jeremy Baeten and Madhavi Senagolage, thank you for all your input with my project, and for your guidance with science and life through the years.

I also want to thank Don Wolfgeher in the Kron lab for his invaluable help with the proteomics aspect of my project, and for meeting with me several times to make sure I understood the intricacies of the data. This project was made that much better with your involvement.

I also want to thank all the people that kept me sane during my Ph.D. To my friends in graduate school, Larischa de Wet, Julian Lutze, Katya Frazier, Nicole Fuller, Matt Funsten, Avelino De Leon, Ryan Brown, Celisa Chan, Peter Flynn, Olivia Gaylord and Jill Rosenberg, thank you for making life so fun outside work and understanding the struggles of graduate school. You all helped me strike a healthy balance. To my college and high school friends, there are too many of you to name, but thank you for ensuring that I still stayed in touch with all of you as life got busier with school work. I am so pleased that we have managed to maintain our friendships after all these years. Finally, to my oldest and closest group of friends, Anila

Alexander, Minorka D'Souza, Hiba Ahmad and Soujanya Srinivasan, thank you for always being there for me through the good and bad times.

Finally, I would like to thank my family for being my strongest pillar of support and strength through graduate school. Amma and Appa, thank you for tolerating my infrequent visits home and listening to me always when I needed to vent. Thank you for your guidance and advice over the years – this accomplishment is dedicated to both of you. Roshan, thank you for the movie and TV show recommendations and random pieces of trivia or ideology that you always share with me. I will always cherish our shared Thanksgivings over the years. I would also like to thank my aunts, uncles and cousins for always being supportive and helping me become more independent as I embarked upon this journey far away from home. Visiting all of you over the years made me feel less alone and always helped me reconnect with my roots and feel more grounded.

Finally, to my partner and my rock, Dave Klawon, I literally could not have made it through graduate school without your support. You made it so much more manageable, and helped me reframe situations and adopt a better perspective. Thank you for the fun adventures, interesting conversations, being a pet parent with me and being my home away from home. I am so excited for us to start this next chapter of our lives together.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ATI - alternative transcription initiation

AML – acute myeloid leukemia

CAGE-seq – Cap analysis of gene expression-sequencing

CDR – commonly deleted region

CUX1 – CUT-like homeobox 1

EMT – epithelial-to-mesenchymal transition

EPD – eukaryotic promoter database

gRNA – guide RNA

HD – homeodomain

HDAC – histone deacetylase

HSPC – hematopoietic stem and progenitor cells

IP - immunoprecipitation

mRNA – messenger RNA

MDS – myelodysplastic syndrome

MPN – myeloproliferative neoplasm

POL2RA – RNA polymerase II

PCR – polymerase chain reaction

PIC - pre-initiation complex

TCGA- The Cancer Genome Atlas

t-MN – therapy-related myeloid neoplasm

TNBC – triple negative breast cancer

TSG – tumor suppressor gene

TSS – transcription start site

RACE – Rapid amplification of 5’ cDNA ends

RBP – RNA-binding protein

ABSTRACT

CUX1, a homeodomain-containing transcription factor, is recurrently deleted or mutated in multiple tumor types. In myeloid neoplasms, *CUX1* deletion or mutation carries a poor prognosis. We have previously established that CUX1 functions as a tumor suppressor in hematopoietic cells across multiple organisms. Others, however, have described oncogenic functions of CUX1 in solid tumors, often attributed to truncated CUX1 isoforms, p75 and p110, generated by an alternative transcriptional start site or post-translational cleavage, respectively. Given the clinical relevance, it is imperative to clarify these discrepant activities. Herein, we sought to determine the CUX1 isoforms expressed in hematopoietic cells, and find that they express the full-length p200 isoform. Through the course of this analysis, we found no evidence of the p75 alternative transcript in any cell type examined. Using an array of orthogonal approaches, including biochemistry, proteomics, CRISPR/Cas9 genomic editing, and analysis of functional genomics datasets across a spectrum of normal and malignant tissue types, we found no data to support the existence of the CUX1 p75 isoform previously described. Based on these results, prior studies of p75 require reevaluation, including the interpretation of oncogenic roles attributed to CUX1.

CHAPTER 1

INTRODUCTION

1.1 ALTERNATIVE SPLICING AND PROTEIN ISOFORMS

Eukaryotic gene expression is a complex process that is orchestrated by multiple regulatory steps. DNA, the genetic code in eukaryotes is processed to form messenger RNA (mRNA) through a process called transcription. mRNA is the molecular recipe for the synthesis of proteins, through a process called translation. The process of transcription is initiated at regulatory regions in the DNA called promoters. Promoters contain the regulatory sequences and epigenetic marks that together dictate the level of transcription. Promoters recruit the transcription machinery, such as RNA polymerase II (POL2RA) and other transcription factors, thereby leading to the formation of the pre-initation complex (PIC) [1]. mRNA is typically made up of a coding sequence (CDS), a 5' untranslated region (UTR), and a 3' UTR. The process of removal of introns from messenger RNA (mRNA), called pre-mRNA splicing, is an important step for the expression of eukaryotic genes. These splicing patterns can be altered in different tissues, or in cancerous cells, that can contribute to almost every hallmark of cancer described by Hanahan and Weinberg [2].

Alternative transcription initiation (ATI) is when one gene has multiple transcription start sites (TSS) [3]. ATI is a way to regulate isoform expression pre-transcriptionally, unlike alternative splicing, which regulates isoform expression post-transcriptionally [4]. This phenomenon is thought to be adaptive, but the functional advantage of this is unclear. It is also unclear how pervasive this phenomenon is across mammalian genomes. There have been some attempts at measuring the usage of alternative transcription start sites. In a study of genes expressed in

mouse fibroblasts, 4153 genes out of 9951 genes showed significant initiation at multiple sites, although only 754 of these transcripts were actually translated into distinct proteins [5]. Genome-wide analyses have demonstrated that around half of human and mouse genes have one or more alternative promoters, and on average, a typical human gene has 4 TSSs [6–9]. Computational discrimination of alternative promoters also discovers that almost half of all mammalian genes have alternative promoters that are evolutionarily conserved [8]. This suggests that there is a strong evolutionary selective pressure to retain alternative promoters, as a means of diversifying the transcriptome and proteome.

There are two schools of thought when it comes to the purpose for the existence of ATI. The first one, named the 'adaptive hypothesis', postulates that ATI is a way to expand the transcriptional and proteomic diversity in organisms from a limited number of genes. This is supported by the fact that TSS choice can vary among tissues, and among different physiological processes, and that half of all alternative promoters are evolutionarily conserved. The second one, called the 'error hypothesis', suggests that alternative transcriptional initiation from multiple TSSs may predominantly result from molecular errors that are deleterious. This school of thought is supported by evidence that the TSS diversity of a gene reduces with its expression level (i.e., the higher a gene is expressed, the less likely it is to have alternative TSSs [10]. The 'error hypothesis' highlights the need for a better understanding about how ATI is regulated.

In support of the adaptive hypothesis, transcriptomics experiments reveal that many splice events and alternative TSSs are tissue-specific, and may play a role in tissue differentiation [11]. In this case, alternative promoters can be used to regulate the expression level of a gene in different tissues. For example, the α-amylase gene has two promoters: a weaker one that directs liver-

specific transcription, and a 30-fold stronger promoter that exclusively initiates transcription in the parotid gland [12]. The liver promoter is also active in the parotid gland, but not vice versa [12]. Multiple alternative TSSs within a gene can also expand the 5' UTR repertoire, allowing for more diverse translational regulation [5]. Using alternative TSSs can also serve to enhance or diminish protein synthesis rate. One of the best-known examples of this is the gene *BRCA1,* which has two different isoforms with alternative TSSs. The shorter isoform is more efficiently translated, and is expressed in both cancerous and non-cancerous tissue, whereas the longer isoform is expressed only in cancerous tissue and is translationally inactive, leading to decreased BRCA1 protein expression in breast and ovarian cancer tissue [5,13]. Variations in the coding sequence because of alternative transcriptional initiation can also affect protein function. For instance, *LEF1*, a gene that regulates the transcription of Wnt/β-catenin genes, produces two different protein isoforms; the longer isoform recruits β-catenin to Wnt target genes to transcriptionally activate them, while the shorter isoform lacks the ability to interact with β-catenin, and therefore is unable to turn on the Wnt transcriptional program [14].

As another layer of regulation in different tissues, alternative promoters can have the same function in different cell types, but can be regulated by different cues in these cell types. The glucokinase gene is an example of this type of regulation. The glucokinase gene possesses two alternative promoters that are utilized in a tissue-specific manner. The upstream promoter leading to exon 1β is active mainly in pancreatic and neuroendocrine cells, whereas the downstream promoter leading to exon 1L is mainly active in hepatocytes. The neuroendocrine promoter is not regulated by hormones, while the hepatocyte promoter is heavily positively and negatively regulated by insulin and glucagon respectively [4,15].

In support of the error hypothesis, transcriptional initiation has limited fidelity, and a majority of transcriptional events have been suggested to lack biological function [16]. Additionally, alternative splicing and ATI are increasingly becoming appreciated in cancer types such as neuroblastoma and leukemia. Dysregulation of splicing can take place on the selection of alternative TSSs, alternative polyadenylation, and alternative intron inclusion or exon composition based on choice of splice sites. Mutations in spliceosome factor genes such as *U2AF1*, *SF3B1* and *SRSF2* are frequently found in myelodysplastic syndrome (MDS), and are being investigated as targets of therapeutic intervention [17,18]. Disturbances in the regulation of alternative TSSs can also lead to disease, such as the disruption of the *MYC* promoter in Burkitt's lymphoma [19]. Normally, MYC has two main promoters, P1 and P2, that contribute in varying degrees to transcription in normal tissue. P2 is the downstream promoter that is predominant, and gives rise to 75-90% of *myc* mRNA, whereas P1, the upstream promoter, gives rise to 10-25% of *myc* mRNA. The P1 and P2 promoters of *MYC* have been shown to be differentially regulated by analyzing cap analysis gene expression (CAGE)-seq data (this technique is described in more detail later in this section) [20]. In Burkitt's lymphoma, a chromosomal translocation of the *MYC* gene locus on chromosome 8 to chromosome 14, under the immunoglobin chain elements, disrupts transcription starting from the downstream promoter P2 to instead start from the upstream promoter P1, leading to aberrant transcription and oncogenesis [4,19]. In Burkitt's lymphoma, the *MYC* gene on the chromosomal translocation loses the ability to maintain POL2RA on the P2 promoter, presumably due to interference by regulatory elements at the adjacent immunoglobin locus [4].

Alternative splicing can also be regulated by epigenetic marks, as many splicing decisions happen when nascent RNA is still associated with chromatin, and splicing in eukaryotes occurs co-transcriptionally [21]. More recent studies show that pre-mRNA itself can act as a guide by recruiting histone modifying enzymes and reprogramming the histone modification landscape to mediate alternative splicing [22]. Additionally, there are also studies arguing that both small and long noncoding RNAs can guide local chromatin remodeling to regulate alternative splicing [23,24].

Studies of promoter regions over the years have led to the establishment of several key characteristics associated with them [1]. First, most promoter regions are marked by increased chromatin accessibility. Most genomic DNA has limited accessibility, as it is wrapped around histone subunits to form nucleosomes; however, promoters that are actively involved in transcription initiation are devoid of nucleosomes, and therefore more accessible. From more recent work, it is now appreciated that nucleosome depleted regions are not actually devoid of nucleosomes, but instead have less common histone subunits incorporated into their nucleosomes, such as H3.3 and H2A.Z [25–27]. These alternative histone subunits ensure increased accessibility to chromatin at these regions.

Promoters also generally tend to have increased deposition of activating histone marks such as H3K4me3 and H3K27ac [28]. H3K4me3 has been suggested to provide memory of recent transcriptional activity, thereby facilitating new rounds of transcription [29]. H3K4me3 is written by a variety of histone methyltransferases, such as MLL and SETD1A/B [30,31]. H3K4me3 aids in transcription by recruiting reader proteins that bind to the core promoter motifs (described below) to form the PIC [32]. The histone marks alone are not sufficient to initiate transcription, but have been found to be increasingly correlated with promoters in multiple studies.

Additionally, there are several core promoter motifs that are associated with TSSs, such as the TATA-box motif and the initiator (Inr) motif in vertebrates. Another feature of mammalian promoters is that they overlap with evolutionarily conserved CpG islands, suggesting an important function that needs to be preserved [33]. Promoters also tend to be associated with RNA polymerase II (POL2RA) binding, and the binding of other transcription factors.

TSSs can be mapped using approaches such as cap analysis of gene expression (CAGE)-seq, which utilizes the 5' cap structure of mRNA transcripts to detect the start position and abundance of the transcript (Figure 1.1) [34]. There are also some newer adaptations to CAGE-seq that are useful for mapping transcripts with alternative TSSs [35,36]. Rapid amplification of 5'-cDNA ends (RACE) is another method that is commonly used to identify the 5' ends of mRNAs [37]. Additionally, RAMPAGE and GRO-CAP are newer, sequencing-based methods to map TSSs [38,39].

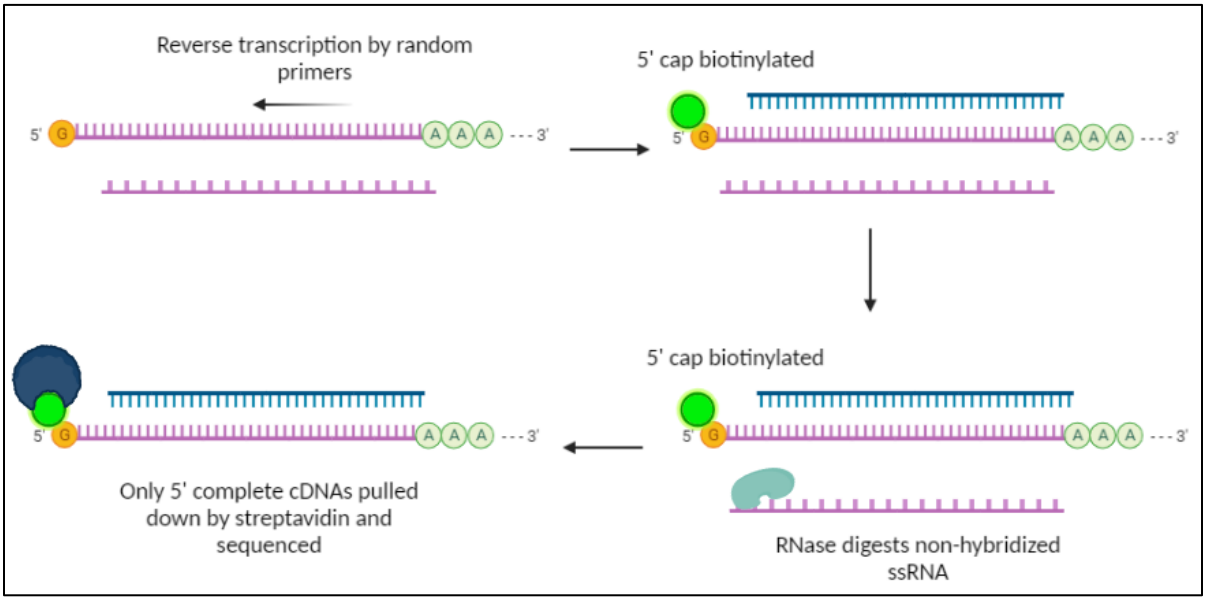


Figure 1.1: Schematic of CAGE-seq workflow. cDNA is reverse-transcribed from mRNA using random or oligo dT primers, such that only mRNAs are converted to cDNA. The 5' cap of

Figure 1.1: Schematic of CAGE-seq workflow (Continued)
these hybridized mRNAs is then labeled with biotin to ensure selection of full-length cDNA. Non-hybridized single stranded-RNAs are then digested with an RNase, leaving 5' complete cDNAs intact in the mixture, which are then pulled down using streptavidin that binds the biotin tag. A double-stranded CAGE linker that is also biotinylated is ligated to the 5' end of the 5' complete cDNAs, and the second strand of this tagged cDNA is synthesized. This is then digested with the Mme1 endonuclease, cleaving the CAGE linker and producing a CAGE tag. A second linker is added to the 3' end of this tag, amplified using PCR, and then sequenced to map exact TSS locations in promoter regions.

1.2 CUX1 IN DEVELOPMENT

CUX1, also known as Cutl1, CDP or Cut, is a homeodomain-containing transcription factor that is evolutionarily conserved. It is involved in several cellular processes such as proliferation, differentiation, and lineage decision. The structure of the CUX1 protein consists of four DNA binding domains (3 CUT repeats, each about ~70 amino acids long, and one CUT homeodomain (HD) [40,41]. Another conserved protein domain is the coiled-coil (CC) domain, predicted to be involved in protein-protein interactions. In addition, the amino-terminal (N-terminal) region contains an auto-inhibitory domain thought to inhibit DNA binding, and the carboxy-terminal (C-terminal) region of the protein also contains two active repressive domains that are thought to repress transcription by recruiting histone deacetylases (HDACs) to target promoters [42,43]. The *CUX1* gene also encodes a second protein called CASP, which shares the coiled-coil domain with CUX1, but none of the other DNA-binding domains [44–46]. CASP is a golgi-associated protein, and has so far not been implicated in human disease (Figure 1.2) [47,48].

CUX1 requires interaction between 2 cut-repeat domains (CR domains) or a CR domain and the homeodomain (HD) for DNA binding. CUX1 preferentially recognizes AT-rich DNA sequences [40,49]. Cut repeat 1, or CR1 was shown to bind DNA very rapidly but very transiently compared to CR2, CR3 and the HD [42,50]. Following these trends, p200 CUX1, which contains all

3 CR domains and the HD, exhibits very rapid DNA binding kinetics in vitro and a preference for "CCAAT" motifs, while the p75, p90 and p110 isoforms, which lack CR1, exhibit much slower DNA binding kinetics and an affinity for "ATCRAT" motifs (R = A or G) [49,51,52].

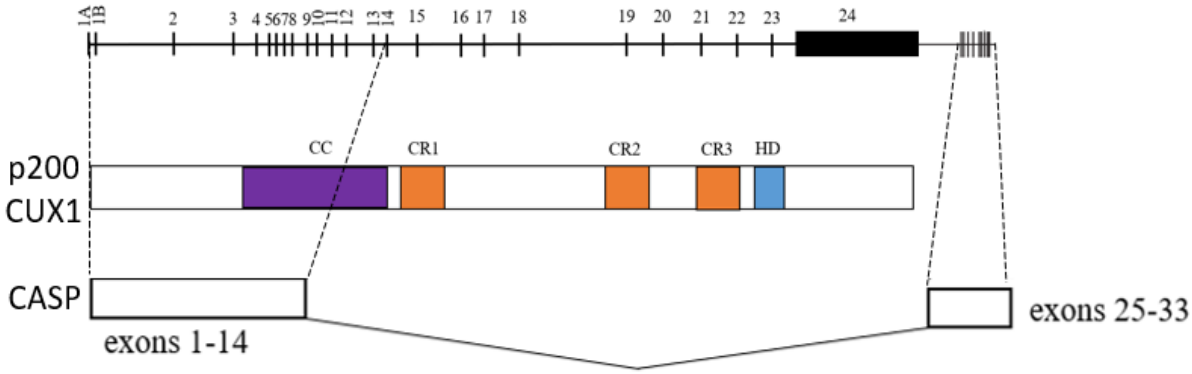


Figure 1.2: CUX1 gene structure and proteins encoded by the CUX1 gene

CUX1 has been shown to be required in the differentiation of several tissues in *Drosophila*, including the wing margin, legs, tracheal system and Malpighian tubules [53]. A study in chicks shows that *CUX1* plays a role in cell type specification downstream of the Notch pathway [54]. CUX1 expression was significantly upregulated with constitutive activation of Notch 1 in a rat epithelial cell line, and CUX1 was shown to interact with a Groucho homolog, TLE-1, in the Notch pathway, suggesting this regulation of CUX1 by the Notch pathway is conserved across several species [55]. Cux1 expression has been reported in the murine cerebellum, specifically in the granule cell lineage, and is thought to play a developmental role in this cell type by repressing dendritic arborization [56,57].

In humans, CUX1 expression has been observed in the central nervous system, where it was first described as a marker for upper layer cortical neurons [58]. In the kidney, *CUX1* expression is

inversely correlated to the degree of differentiation [59]. Indeed, this is true for multiple cell types, where CUX1 expression is highest in precursor cells, and the expression of genes predominantly expressed in terminally differentiated cells are downregulated [60–62]. CUX1 expression is also silenced as cells begin to differentiate down their lineages, as seen by higher transcript levels in HSC and progenitor cells, and lower transcript levels in more differentiated blood cells, and this process is still not completely understood [63]. Two possible mechanism that have been incompletely explored by which CUX1 represses transcription of genes involved in differentiation are by (i) competition for CCAAT binding site occupancy, preventing activation by other transcription factors, or (ii) active repression on the TSS by the two repressive domains in the C-terminal region [64]. One group showed that tissue-specific microRNAs, such as miR-122 in the liver and miR-208a in the heart, can contribute to post-transcriptional silencing of CUX1 during differentiation in mice [65]. *Cux1* mutant mice have shown a requirement for Cux1 in the differentiation of the lung epithelium, hair follicle maturation, and proper development of lymphoid and myeloid cell subsets [66,67]. CUX1 binding sites have been identified within myeloid and lymphoid specific genes, such as gp91-phox promoter, neutrophil collagenase promoter, lactoferrin promoter, neutrophil gelatinase promoter, TCRβ enhancer, TCRδ enhancer, and the immunoglobin heavy chain intron enhancer, indicating that its expression is required for the proper development of several hematopoietic cell types [61,68–70].

CUX1 is a transcription factor, and there have been several attempts to identify signaling pathways and genes transcriptionally regulated by CUX1. CUX1 has been shown to upregulate genes related to embryonic development and differentiation, genes involved in adhesion and cell motility, and genes encoding secreted extracellular matrix proteins [71]. It has also been reported to regulate the expression of genes involved in the regulation of the cell cycle, such as $p21^{WAF}$ (*45*).

In human CD34+ hematopoietic stem and progenitor cells (HSPCs), CUX1 knockdown led to downregulation of genes preferentially expressed in quiescent HSPCs using gene set enrichment analysis, further supporting the fact that CUX1 is involved in maintaining quiescence and controlling induction of genes involved in differentiation [63]. Loss of CUX1 also led to downregulation of genes involved in negative regulation of myeloid cell differentiation, suggesting that CUX1 represses differentiation in this cell type along the myeloid lineage [63]. All these lines of evidence suggest that CUX1 plays a very important role in regulating expression of genes involved in development and differentiation as a transcription factor across several cell types, and this function is evolutionarily conserved across several species.

1.3 ROLE OF CUX1 IN SOLID CANCERS

CUX1 is located on chromosome 7 at locus q22, a region that is amplified in certain cancer types and deleted in other cancers. It is amplified in medulloblastoma [73]; however it has been described as having a tumor suppressor role in this cell type by being involved in DNA repair [56].

In p200 CUX1-transgenic mice, CUX1 was shown to synergize with a *Kras* activating mutation to prevent RAS-induced senescence, enabling prolonged survival of transformed cells [74]. CUX1 was also shown to accelerate repair of ROS-induced DNA damage induced by RAS, and enable RAS-transformed cells to survive more [74]. CUX1 has been shown to function in base excision repair as an ancillary factor for 8-oxoG-DNA glycoslyase, OGG1; this function of CUX1 was implicated in radioresistance [75]. CUX1 levels are increased in carcinomas resistant to chemotherapy [76,77]. Taken together, this data suggests that CUX1 helps cells evade external stressors such as DNA damage induced by chemotherapy or irradiation.

CUX1 has been demonstrated to be upregulated during tumor progression of pancreatic neuroendocrine neoplasms [78]. CUX1 expression in pancreatic cancer cell lines is also associated with increased proliferation and reduced TRAIL-induced apoptosis [78,79]. In addition to the effects CUX1 has on tumor progression in pancreatic cells, it also plays a tumorigenic role in tumor-associated macrophages in the tumor microenvironment [80,81]. In this cell type, CUX1 was shown to interact with the p65 subunit of NF-κB and thereby inhibit its binding to the promoters of chemokines such as CXCL10, which are associated with polarization of M1 macrophages that inhibit tumor progression [80].

CUX1 has also been shown to function by regulating NF-κB-mediated chemokine transcription in melanoma, and causing increased proliferation and cell cycle progression in melanoma cells [82,83]. There is also evidence for the role of CUX1 in mediating tumorigenesis in melanoma by regulating the response to inflammation in this cell type. CUX1 has been shown to repress the production of pro-inflammatory cytokines CCL3 and CCL4 in dendritic cells, and CXCL1 in melanoma cells [84–86]

In cohorts of breast and pancreatic cancer patients, *CUX1* expression was elevated in patients with high-grade tumors [71]. In these cohorts, *CUX1* expression also inversely correlated with relapse-free survival and overall survival (*44*). In triple-negative breast cancer patient tissue, both CUX1 and cathepsin L (the protease that cleaves p200 CUX1 to generate the p110 CUX1 isoforms; described in greater detail in Chapter 1.5) were found to be highly expressed compared to tissue from ER-positive samples [87]. In triple negative breast cancer (TNBC) cell lines, CUX1 was found to directly bind the ER-α promoter and repress its transcription, and thus it is hypothesized that CUX1 activated by Snail signaling and cathepsin L cleavage may contribute to TNBC via repression of ER-α [87].

In bladder cancer, this Snail-CUX1 axis was again found to be activated through PIK3CA, and led to upregulated expression of β-catenin, Vimentin, and decreased expression of E-cadherin, providing evidence for a role for CUX1 in EMT and metastasis in this cancer type [88].

In metastatic prostate cancer androgen-sensitive cells, migration and invasion increased upon CUX1 knockdown, suggesting expression of CUX1 keeps cells from undergoing epithelial-to-mesenchymal transition (EMT) [89]. CUX1 expression is elevated in androgen-independent cell lines compared to androgen-sensitive cell lines [89].

CUX1 knockdown in several solid cancer cell lines results in reduced migration and invasion, establishing a role for *CUX1* in potentiating cancer cell motility [71]. CUX1 knockdown cells injected into nude mice also engrafted worse and exhibited decreased metastasis [71,90]. Collectively, all this data indicates that there are multiple reports of CUX1 having an oncogenic role in solid cancers.

There are some reports of CUX1 playing a protective role in colitis and inflammatory bowel disease, which can often be precursor conditions to developing colorectal cancer as a result of chronic inflammation leading to transformation in the intestinal epithelial cells [84,91]. CUX1 expression is induced strongly when cells are exposed to inflammatory stress, such as TNF-α treatment, and CUX1-knockout mice take significantly longer to undergo mucosal healing after experimentally-induced colitis [84]. A later study by the same group showed that CUX1 actively plays a role in intestinal wound healing by transcriptionally inducing the VAV2-RAC1 signaling pathway [91]. These seemingly conflicting reports of the role of CUX1 in either promoting or suppressing oncogenesis, regardless of which cancer types displayed gain or loss of chromosome 7, served as the starting point for the work presented in this thesis. We postulated that a potential

explanation for these conflicting roles of CUX1 could be explained by different physiological effects potentiated by the different CUX1 isoforms.

1.4 ROLE OF CUX1 IN HEMATOLOGICAL MALIGNANCIES

The importance of *CUX1* in hematology was identified through the study of -7 and del(7q) cytogenetic abnormalities in hematological malignancies including de novo acute myeloid leukemia (AML) and therapy-related myeloid neoplasms (t-MN). -7/del(7q) occurs in approximately 8% of de novo AML and upto 50% of t-MN, particularly those associated with alkylating agent treatment [92–94]. Myelodysplastic syndrome (MDS) is another clonal hematopoietic disorder characterized by -7/del(7q) abnormalities, and exhibiting ineffective hematopoiesis, dysplasia in myeloid lineage cells, cytopenias in one or more blood lineages, and absence of fully transformed leukemia but increased susceptibility to other hematological malignancies. t-MNs in particular represent a very interesting subset of patients who are largely refractory to the standard-of-care treatments, and thus are in urgent need of specialized, targeted interventions; about 50% of these patients present with -7/del(7q) cytogenetic abnormalities. These -7/del(7q) cytogenetic abnormalities are frequently identified in the HSC and progenitor compartments, compared to more differentiated hematopoietic cell populations [95–99]. MDS/AML patients with -7/del(7q) often have marked changes in the proportion of different hematopoietic progenitor compartments such as long-term HSCs, common myeloid progenitors, and granulocyte-monocyte progenitors [97,100]. -7/del(7q) cytogenetic abnormalities are often associated with therapy resistance and poor outcomes [101–103].

Three commonly deleted regions (CDRs) have been identified on chromosome 7 at 7q22, 7q34, and 7q35-36, and these are thought to encode one or more tumor suppressor genes (TSGs),

loss of which predisposes patients with these cytogenetic abnormalities to develop hematological malignancies [104–106]. 7q22 is the most deleted region in studies of MDS/AML patients with del(7q) [104]. Genomic analyses in several patient cohorts indicate a haploinsufficient role for several genes encoded on 7q22 in leukemogenesis [95,107–112].

There have been several attempts to identify the putative TSGs encoded on chromosome 7 by different groups. However, due to the heterogenous nature of sample sizes, and the complex karyotypes of a large subset of these patients, efforts to identify candidate TSGs in these regions were not fruitful for a long time [106]. *Mll3* was first identified as a haploinsufficient tumor suppressor on 7q36.1 in AML [113]. Missense somatic mutations of *Mll3* have been identified in several cancer types [113–117]. In mice, *Mll3* knockdown on its own is not leukemic; however, in the presence of a cooperating *p53* null mutation and *Nf1* suppression, these mice develop AML. Loss of MLL3 alone impaired differentiation in mice and had features reminiscent of human MDS [113]. However, since loss of MLL3 alone did not lead to death due to disease burden, it remains possible that there are other tumor suppressor genes in these CDRs on chromosome 7. *EZH2* is another gene on 7q36.1 that has been described as a tumor suppressor in T-acute lymphoblastic leukemia and myeloid disease [118–120]. *EZH2* is a complicated candidate, because it has also been described as oncogenic in epithelial tumors [121,122]. However, in hematopoietic malignancies, *EZH2* inactivating mutations are acquired early during disease progression and are associated with a poor prognosis [118,123]. In combination with other cooperating mutations, inactivating *Ezh2* mutations can accelerate progression of leukemia [119] Other genes encoded within the chromosome 7 CDRs whose loss has been associated with leukemogenesis are *MLL5,* a hematopoiesis and cell cycle regulator located on 7q22, and *LUC7L2,* a spliceosomal protein on 7q34 recurrently mutated in myeloid malignancies [111,124–128].

Another gene on chromosome 7 implicated in tumor suppression is *CUX1*. *CUX1* is encoded on the 7q22 CDR, and previous work has identified *CUX1* as a haploinsufficient TSG in hematopoietic cells [112]. *CUX1* inactivating mutations are enriched in hematopoietic malignancies, and have been independently associated with a poor prognosis [103,112,129]. *Cut* haploinsufficiency in *Drosophila* hemocytes (a myeloid like blood cell in this species) leads to hemocyte overgrowth and tumor formation [112]. Decreased *CUX1* expression in human HSPCs also led to an engraftment advantage when transplanted into immunodeficient mice [112]. When a cohort of 1480 patients with myeloid neoplasms (MN) were stratified based on the presence of *CUX1* somatic mutations or deletions ($CUX1^{MT/DEL}$), it was observed that $CUX1^{MT/DEL}$ correlated with worse survival compared to $CUX1^{WT}$ [130]. $CUX1^{MT}$ is also 3.4 times more likely in a secondary MN (MN patients with a primary malignancy not treated with chemotherapy or radiotherapy) compared to a primary MN (no prior cancer) [131]. The mutational landscape of CUX1 in this MN cohort is reminiscent of a tumor-suppressor gene rather than an oncogenic gene, indicated by a high ratio of damaging versus benign mutations as well as the absence of mutational hotspots [130,132]

It is becoming increasingly evident that contrary to Knudsen's 2-hit model, TSGs can lead to increased tumorigenesis merely through haploinsufficiency. Evidence for the haploinsufficient model has been demonstrated in del(5q) and cases of chromosome 20 deletion [133–135]. Thus, it is possible that -7/del(7q) cases are mediated by a similar model.

Work from our lab sought to identify the role of *CUX1* in hematopoiesis, and resulted in the generation of an inducible shRNA CUX1-knockdown mouse model [63]. Knockdown of CUX1 in these mice resulted in development of MDS/MPN and anemia, thus it is evident that loss of CUX1 alone is sufficient to induce de novo myeloid malignancies [63]. More recent work on

CUX1 haploinsufficiency has shown that CUX1-deficient cells seem to propagate acute myeloid leukemia by activating the CFLAR gene, which causes CUX1-deficient cells to evade apoptosis [136].

CUX1 haploinsufficiency has a huge impact on gene expression regulated by the transcription factor. In the K562 erythroleukemia cell line, *CUX1* knockdown led to striking changes in expression of genes involved in cell cycle regulation, and increased proliferation was observed [63]. This increased proliferation was also observed in human CD34+ hematopoietic stem cell and progenitor cells (HSPCs) upon *CUX1* knockdown [63]. In the K562 cell line, CUX1 acts as both an activator and a repressor; however in human CD34+ HSPCs, RNA-seq after CUX1 knockdown showed downregulation of 81% of differentially expressed genes, suggesting that CUX1 primarily acts as a transcriptional activator in this cell type [63,137]. Similarly, there is also newer evidence in the lab that CUX1 predominantly binds in enhancer regions in the K562 cell line, and predominantly in promoter regions in cells of the erythroid lineage. All this evidence suggests that CUX1 may have different transcriptional programs in different cell types.

The mutational spectrum of -7/del(7q) leukemias is significantly different from that of other AMLs, which might suggest that loss of one of the CDRs, and CUX1 by proxy, might make these leukemias more susceptible to gaining specific passenger mutations [138]. Indeed, in a study of 200 TCGA samples, RAS pathway mutations co-occur with CUX1 deletions at a much higher rate, compared to samples with no CUX1 or 7q deletions [138]. Additionally, there is a strong association between -7/del(7q), and del(5q) and *TP53* mutations [138]. It is possible that understanding the biology behind why CUX1 loss leads to such a unique mutational landscape might allow us to discover better therapeutic interventions for this subset of patients. It is

unknown if the somatic mutation spectrum in -7/del(7q) patients presenting with different diseases is the same, or different depending on the classification of their disease.

There is a growing body of evidence of the role of CUX1 in DNA damage recognition and DNA repair, specifically in hematopoietic tissue. High levels of CUX1 are predicted to enhance p53 functionality in medulloblastoma, by looking at proteins that interact with CUX1 using the iRefIndex protein interaction database, an interaction network database based on experimentally validated protein-protein interactions [56]. CUX1 has been shown to associate with MBD1, which has been previously described in the DNA checkpoint response [139]. In AML, work from our group reveals that CUX1-deficient mice are actually more therapy-resistant [140]. In this t-MN mouse model, CUX1-deficient HSPCs sustain unrepaired DNA damage, and continue to proliferate and expand when exposed to cytotoxic chemotherapy, leading to clonal outgrowth and therapy-resistant erythroleukemia [140].

In summary, loss of even a single allele of CUX1 in multiple mouse models seems to recapitulate the myelo-dysplastic effect seen in -7/del(7q) patients, indicating that CUX1 acts as a haploinsufficient tumor suppressor gene in myeloid malignancies. This loss of CUX1 is thought to be a passenger mutation, either through loss of chromosome 7, or through the acquisition of initial driver mutations, and CUX1 loss has been shown to cooperate with more characteristic oncogene mutations to drive oncogenesis. Loss of CUX1 also seems to potentiate oncogenesis by leading to increased proliferation, cell cycle progression, skewing towards myeloid progenitors and lack of DNA damage recognition and DNA repair, leading to accumulation of DNA damage.

1.5 CUX1 ISOFORMS AND ROLE IN TUMOR PROGRESSION

The *CUX1* gene has been reported to encode several protein isoforms in the literature. p75 CUX1 was described to be initiated from an alternative transcriptional start site (TSS) within intron 20 of the *CUX1* gene [141]. It was found to be expressed in the mouse thymus, CD4+ T cells, breast tumor tissue and breast cancer cell lines [141]. All subsequent studies of the biological effects of this isoform were evaluated in overexpression models, and there are virtually no studies in the literature that study the effect of the endogenous p75 isoform by knocking it down or perturbing it directly. p75 transgenic mice generated by exogenously expressing the CUX1 isoform under the mouse mammary tumor virus-long terminal repeat (MMTV-LTR) promoter develop a myeloproliferative disease-like myeloid leukemia in mixed genetic backgrounds, and malignancies in the mammary gland and lung metastases in the FVB genetic background [90,142].

Cell lines established from the p75 CUX1 transgenic mice expressed high levels of Wnt genes, and this effect was also observed upon ectopic overexpression of p110 CUX1 in human tumor cell lines [143]. Activation of the Wnt pathway is also observed in the p75 transgenic mouse tissue [90]. Ectopic overexpression of short CUX1 isoforms has been implicated in turning on several other oncogenic signaling pathways, such as PI3K/Akt signaling [88,143].

p110 CUX1 is generated by proteolytic cleavage of full-length CUX1 by a serine protease, cathepsin L between the cut repeat 1 (CR1) and the cut repeat 2 (CR2) domains, that takes place at the end of the G1 phase [51,144]. A recent report has also described p200 CUX1 to be hydrolyzed by neutrophil elastase, and generate p110 CUX1, although this remains to be validated rigorously, on account of the use of antibodies that detect CASP to detect short CUX1 isoforms

[145]. It has been shown to stably bind DNA, and act as either a transcriptional repressor or activator, dependent on promoter context [146]. p110 CUX1 was also shown to stimulate proliferation and accelerate entry into the S phase of the cell cycle [147,148]. In the S phase of the cell cycle, there is documented evidence of p110 CUX1 activating transcription of genes involved in DNA replication [148]. The DNA-binding and transcriptional activity of p110 CUX1 is again downregulated in the G2/M phases of the cell cycle by post-translational modifications, which is talked about in more detail in the next section; it is not thought to play any significant role in these later stages of the cell cycle. However, in transformed cells, p110 CUX1 expression is constitutively elevated, and it is presumed that expression of an oncogene can lead to increased proteolytic processing of the p200 isoform to generate the p110 isoform [149]. Indeed, there is evidence of cathepsin levels (the enzyme that cleaves p200 CUX1 to generate p110 CUX1) being elevated by oncogenes such as *ERBB2 (Her2)* [150]. Overexpression of a p110 construct in mouse mammary epithelial cells stimulated more migration, and exhibited increased adhesion compared to a control vector, or a vector expressing p200 CUX1 [151]. A mouse model over-expressing the p110 CUX1 isoform develops mammary tumors [90]. p110 CUX1 was recently identified as a driver of pancreatic cancer progression in a genetic mouse model, leading to increased proliferation, cell cycle progression and tumor burden [152]. Short CUX1 isoform expression, specifically that of p110, is reported to be elevated in human uterine leiomyomas [153].

p110 CUX1 has been shown to induce a transcriptional program of increased migration and invasion. Gene expression analysis in mammary cells stably expressing p110 CUX1 shows induction of Rho-GTPases implicated in cytoskeleton remodeling required for cell motility,

matrix adhesion proteins, such as integrins, focal adhesion kinase, E-cadherin and occludin, and proteins involved in EMT [151].

Aside from these well-studied short isoforms, there are also several other less well characterized isoforms, such as p150, p80, p90, which have been described to be generated by post-translational proteolytic cleavage by a nuclear isoform of cathepsin L, or other types of caspases [52,144,146,154]. For example, CUX1 has been reported to be cleaved by neutrophil elastase to generate a short CUX1 isoform in the MV4;11 AML cell line [155]. Expression of cathepsin caspases have been shown to be elevated in the context of cancer, and can be induced by the expression of oncogenes [156–158]. Cathepsins belong to the family of aspartic peptidases, which are characterized by an acidic active site residue (aspartate or glutamate) and an activated water molecule that perform an acid-base reaction on the peptide bond. Specifically, cathepsin L is one of 11 cysteine cathepsin proteases and is generally found localized in the lysosome. Cathepsin L possesses endoprotease activity and has been shown to be important in protein degradation, and in the development and function of the immune system [159,160].

Despite their lysosomal localization, there has been evidence for a role of cathepsin L proteins lacking a signal peptide in the nucleus. Cathepsin L deficiency is correlated with a global rearrangement of chromatin and a redistribution of histone marks [161]. Peptides generated by proteolytic processing often exhibit distinct half-lifes, cellular localization and biological functions that are distinct from the full-length protein [162,163]. In breast and prostate cancer, the transcription factor Snail, typically associated with EMT, has been shown to regulate its own transcription and upregulate nuclear cathepsin L activity, which subsequently degrades p200 CUX1 to generate p110 and p90 CUX1 isoforms, which further promotes EMT [164].

It appears that the ratio of CUX1 isoforms seems to play a more crucial role in cancer than just the expression level of CUX1. There is also emerging evidence from newer work in our lab that CUX1 may function in a dose-dependent manner, and that too much or too little expression of CUX1 protein levels could lead to different lineage decisions and deleterious consequences, but more investigation along this line of evidence is needed before definitive conclusions can be drawn. In the literature, an overexpression mouse model of p200 CUX1 develops late-onset mammary carcinogenesis, which is a preliminary piece of evidence that over-expressing CUX1 at levels that are physiologically aberrant can also lead to tumorigenesis [74]. Other studies in transgenic mice where CUX1 is ectopically over-expressed report similar findings, such as multiorgan hyperplasia and larger kidneys, heart, liver, and testes, apparently caused by repression of the $p27^{kip1}$ gene [165]. Since there are reports of multiple CUX1 isoforms with seemingly unique functions attributed to each in the literature, it became imperative to understand the nature and role of these isoforms in hematopoiesis better, and understand if short isoforms could contribute to tumorigenesis in a way that p200 CUX1 could not.

1.6 REGULATION OF CUX1

CUX1 has been shown to be regulated by several signaling pathways, and it is unclear if these different signaling pathways work in conjunction to regulate CUX1, or if the manner of regulation is different depending upon different cellular contexts. *CUX1* is positively regulated by TGF-β, and this effect has been demonstrated in both cancer cells and tumor-associated macrophages [71,80]. CUX1 has also been shown to be positively regulated by the PI3K/Akt pathway, when incubation with ligands that activate the PI3K/Akt pathway also upregulates CUX1 levels [79,129]. This relationship between CUX1 and PI3K appears to be a positive feedback

loop, as PI3K has also been shown to negatively regulated by CUX1 [63,129]. CUX1 knockdown in human CD34+ HSPCs induces genes transcriptionally upregulated by the PI3K/Akt pathway, and there is a report of CUX1 transcriptionally activating a PI3K-inhibiting gene, *PIK3IP1* [129]. CUX1 homolog Cut has also been shown to be regulated by Notch to mediate cell type specification in Drosophila [166–168]. Since the Notch signaling pathway has been shown to be highly conserved across several species, it was presumed that this signaling axis would persist in mammals. CUX1 was shown to co-localize with several Notch pathway components in numerous tissues during rat embryogenesis, and co-immunoprecipitated with Grg4, a protein in the Groucho family of co-repressors that functions downstream of Notch effector proteins to mediate gene repression, in a rat kidney epithelial cell line [55,169]. It is intriguing that virtually all of the pathways so far shown to regulate CUX1, a tumor suppressor gene, have been established as oncogenic signaling pathways. This could suggest that inducing CUX1 expression is used as a way of controlling increased proliferation and differentiation in a physiological context, but is hijacked when cells transform and these signaling pathways are upregulated, leading to aberrant CUX1 over-expression and subsequent tumor progression.

CUX1 DNA-binding ability has been reported to be activated by proteinase-activated receptor 2 (PAR_2), a G-protein coupled receptor [170]. This activation is mediated through the CR3 and HD domains [170]. There is also evidence for the DNA-binding ability of CUX1 being regulated through phosphorylation of several key residues by protein kinase C, casein kinase II, protein kinase A, and cyclin A/Cdk1 [72,171–173].

CUX1 expression and function has been shown to be regulated through the cell cycle. CUX1 DNA-binding activity is slowly upregulated as cells progress from G1 to S phase, as a result of

dephosphorylation by the phosphatase Cdc25A [72]. DNA binding is then gradually inhibited following phosphorylation by cyclin A/Cdk1 in G2 phase and cyclin B/Cdk1 in M phase [173,174]. It is also thought that CUX1 can mediate different transcriptional programs through the cell cycle by cooperating with different cell cycle transcription factors at different stages of the cell cycle. For example, p110 CUX1 has been shown to cooperate with E2F1 and E2F2 through indirect protein-protein interaction, recruit them to chromatin and transcriptionally activate genes involved in cell cycle progression, DNA replication and DNA repair [175]. CUX1 was identified as the DNA-binding component of the Rb-containing transcription factor complex HiNF-D, which is regulated by the cell cycle [176]. CUX1 has also been shown to form complexes with cell cycle protein such as cyclin A, Cdk2 and Rb-related proteins [176]. Several putative CUX1 target genes are regulated in a cell-cycle-dependent manner, such as $p21^{WAF1/CIP1}$, c-myc, thymidine kinase and histones. Expression of histone genes is required for progression into the S phase of the cell cycle, as de novo synthesis of histone nucleosomal proteins is needed for the packaging of newly synthesized DNA into chromatin [177]. More work is required to understand what role CUX1 plays in controlling cell cycle progression.

CHAPTER 2

2.1 MATERIALS AND METHODS

Human cell culture

KG-1, Mono7, ML-2, HL-60, Kasumi-1, and UoCM1 were grown at 37^{o}C in Gibco RPMI 1640 media supplemented with 20% fetal bovine serum (FBS) and 1X antibiotic-antimycotic (Gibco). K562, U937 and T-47D cell lines were grown in Gibco RPMI 1640 media supplemented with 10% FBS and 1X antibiotic-antimycotic. NIH-3T3 fibroblasts, MCF7 and MDA-MB-231 breast cancer cell lines were grown in Gibco DMEM media supplemented with 10% FBS and 1X antibiotic-antimycotic. All cell lines were authenticated using STR analysis to confirm their identity.

Human mobilized peripheral blood CD34+ HSPCs from multiple healthy donors were purchased from the Fred Hutchinson Co-operative Center for Excellence in Hematology (Seattle, WA, USA). CD34+ HSPCs were expanded in StemSpan SFEMII base media supplemented with CC110 culture supplement for 1-3 days prior to electroporation.

Generation of CUX1-GFP tagged KG-1 cell line

pCUX1.1.0-gDNA and pCUX1-donor plasmids were a gift from Kevin White (Addgene plasmid #112434; http://n2t.net/addgene:112434 ; RRID:Addgene_112434). The pCUX1-donor plasmid was chemically synthesized by Jingdong Tian (General Biosystems, Inc). KG-1 cells were transfected with 0.5 μg of each plasmid using the Neon® Transfection System (Invitrogen by Life Technologies, Waltham, MA, USA). The electroporation settings used for transfection was 1650 volts, 20 ms pulse width, 1 pulse, and electroporation was performed according to manufacturer instructions. Cells were cultured in 0.5 mg/ml G418 after 7 days for 3 weeks to

select for a transfected population. Primers used to confirm correct integration of the GFP tag are listed below:

EGFP 5' primer: 5' – CATGAAGCAGCACGACTTCT – 3'

EGFP 3' primer: 5' – CTGCTTGTCGGCCATGATATAG – 3'

5' primer spanning homology arm and EGFP: 5'-GGAACCTATCGAATGGGAGTTC-3'

3' primer spanning homology arm and EGFP: 5' – AAGTCGTGCTGCTTCATGT – 3'

Generation of CUX1-mCherry tagged mouse model

CUX1-mCherry tagged mice generated in the lab were used to examine tagged CUX1 isoforms. CUX1-mCherry mice were designed using CRISPR/Cas9-mediated insertion. The guide sequence [sequence 1] 5'-CCATCGAATGGGAGTTCTGA-3' was designed using the Broad Institute design tools to target the exon 24, the final coding exon of CUX1. To facilitate insertion, regions flanking the cut site were amplified to generate 2-kb and 3-kb homology arms. The homology arms were PCR amplified from B6C3F1 mouse genomic DNA using sequences 2 and 3, resulting in a 5097bp product topo cloned into the pCR-XL_2 vector. The mCherry tag was PCR amplified from the pmCherry-N1 plasmid and then Gibson cloned using sequences 4 and 5 in the middle of the 5097bp genomic sequence. The plasmid was then transformed into NEB 5-α competent *E. coli* and plated on ampicillin plates (100 μg/ml) for selection. Colonies were selected and expanded overnight in 100 ml LB broth with ampicillin (100 μg/ml). The donor plasmid was purified using FosmidMax kit (epicenter #FMAX046). The donor sequence and orientation were confirmed by Sanger sequencing using sequences 6-8, which span both homology arms and the mCherry insert. CRISPR-Cas9 gRNA and Cas9 nuclease were purchased

from Synthego. A ribonucleoprotein complex was assembled for microinjection as described in manufacturer's instructions with a final concentration of 20 ng/µL gRNA, 50 ng/µL Cas9 nuclease and 12.6 ng/µL donor plasmid. Mixes were then injected into the nuclei of C57BL/6J embryos. PCR spanning the insert was used to identify successful insertions which were validated by Sanger sequencing using primer sets [sequences 9-14].

Reference Number	Sequence
1	CCATCGAATGGGAGTTCTGA
2	CATGGTGGTTCTCAGCCATA
3	ACACTGAGGGGCAATAGTGG
4	ATCGAATGGGAGTTC GGCGGCGGCGGCAGCGTACCGGTCGCCACCATGGTGAGCAAGGGC
5	GGCCCGGCGCCCTCACTTGTACAGCTCGTCCATGCCGCCG
6	TAATACGACTCACTATAGGG
7	ACCTGGTGCGGAAGAAGAAG
8	CGTTAAGTGCGCAACACG
9	GCCCATCGAATGGGAGTT, GATTGGGCTTAATGCTCCTTTG
10	GCGAACTTGAACAGCATCATC, GCGTTAAGTGCGCAACAC
11	ATCCACCGCCTGGAGAA, CCTTGGCCTATGGCGATTT
12	GACGGCGAGTTCATCTACAA, GGAGGTGATGTCCAACTTGAT
13	GGCCATCATCAAGGAGTTCA, GGGAAGGACAGCTTCAAGTAG
14	GACTACTTGAAGCTGTCCTTCC, GATGGTGTAGTCCTCGTTGTG

Western blots

Cells were lysed in RIPA buffer (50 mM Tris-HCl, 150 mM NaCl, 0.5% sodium deoxycholate, 0.1% SDS, 5 mM EDTA and 1% NP-40 substitute). 1% Halt™ Protease Inhibitor Cocktail (Thermo Fisher) was added to the lysis buffer before use. The lysates were passed through a 25-gauge needle, incubated on ice for 20 minutes, with frequent vortexing, and clarified by centrifugation (5000g, 10 minutes at 4°C). Total protein of the resulting supernatant was quantified using the Bradford assay at 595 nm wavelength, with BSA used to generate the

standard curve. 10-15 µg of protein was subjected to SDS-PAGE and probed with anti-CUX1 antibody conjugated to HRP (B-10-HRP, mouse mAb derived against aa 1308-1332, Santa Cruz, 1:1000 in 5% milk/TBST) and visualized using ECL substrate. β-actin was detected with anti-β-actin-HRP (C4, Santa Cruz, 1:3000 in 5% milk/TBST). Other antibodies used to probe for CUX1 expression include ABE217 (rabbit polyclonal antibody derived against aa 861, Millipore Sigma, 1:1000 in 5% milk/TBST), and PUC (rabbit polyclonal antibody generated in-house that recognizes aa 1223-1242, 1:1000 in 5% milk/TBST). GFP (D5.1) rabbit mAb #2956 (Cell Signaling Technology, Product #2956S, 1:1000 in 5% FBS/TBST) was used to probe for GFP-tagged CUX1.

gRNA design

All gRNAs were designed using the Broad Institute's sgRNA designer tool, and generated gRNA sequences were verified using Synthego's Verify Guide Design tool. gRNA sequences used in this study were purchased from Synthego, and the sequences are listed below:

CUX1 exon 4 gRNA: 5' – UGCACUGAGUAAAAGAAGCA – 3'

CUX1 intron 20 5' gRNA 1: 5' – GUAUUUCACGAUUCAGCCAA – 3'

CUX1 intron 20 3' gRNA 1: 5' – CUUUGGGUCAUACAUUGGCA – 3'

CUX1 intron 20 5' gRNA 2: 5' – AUGGCACAAAUCCACGCCAC – 3'

CUX1 intron 20 3' gRNA 2: 5' – AUACUAAUUAAACGCUCUGU – 3'

CUX1 exon 23.1 (NLS) gRNA: 5' – GCUGUGCCGCCGCUUCAUGU – 3'

CUX1 exon 23.2 (HD) gRNA: 5' – CCAGCUGAAGAAACCCCGG – 3'

HPRT gRNA: 5’ – GCAUUUCUCAGUCCUAACA – 3’

gRNA transfections

gRNA:Cas9 RNP complexes were formed by mixing 1.8 µL of each gRNA at 100µM and 1.5 µL Cas9 nuclease at 20 µM in 15 µL of electroporation buffer R. RNP complexes were incubated at room temperature for 10 minutes. 2 x 10^5 KG-1 cells were pelleted and washed twice with PBS. The cells were then suspended in the RNP complex, and electroporation was performed using the Neon® Transfection System (Invitrogen by Life Technologies, Waltham, MA, USA). The optimized electroporation settings used for transfecting the KG-1 cell line was 1700 volts, 20 ms pulse width, 1 pulse, and electroporation was performed according to manufacturer instructions. For CD34+ HSPCs, 0.71ul Cas9 nuclease at 20uM was mixed with 2.39ul gRNA at 30uM and 0.9 ul of electroporation buffer T. RNP complexes were formed by incubating for at least 15 minutes at room temperature. 200,000 CD34+ cells in 8ul of Buffer T were then added to the RNPs, and 10ul of the mixture electroporated and immediately cultured in SFEMII + CC110. AAVS1 gRNA was used as a negative control. Electroporation settings used for CD34+ HSPCs was 1600 volts, 10ms pulse length, 3 pulses.

To examine the clonality of transfected cell lines, a serial dilution approach was adapted from Corning Life Sciences. Briefly, the cells are suspended at 2 x 10^4 cells/ml, and 200 µl of this solution is added to well A1 in a 96- well plate. The cells are then diluted 1:2 with fresh media in each well down the first column of the 96-well plate. The first column is then also diluted 1:2 across the rows of the plate using fresh medium and a multi-channel pipette. Plates are then incubated at 37 degrees for 10-14 days to isolate clonal populations. Edited single cell clones were also established by sorting single cells into a 96-well plate on the AriaIIIu cell sorter.

Clonal populations are identified by looking at the indel spectrum of DNA extracted from cells within a single well.

PCR confirmation of gRNA editing

The following primers were used to amplify the gRNA cut site from genomic DNA to verify editing efficiency:

CUX1 intron 20 5' primer for first cut site: 5' – AGCGCCCCTGTTTAGTTCTC – 3'

CUX1 intron 20 3' primer for first cut site: 5' – GAGCCACCACGAAACTCAGA – 3'

CUX1 intron 20 5' primer for second cut site: 5' – CCCTCATGTTAAGCCTTCCGA – 3'

CUX1 intron 20 3' primer for second cut site: 5' – CACTGGAACTCATCGGGGAC – 3'

CUX1 exon 4 gRNA 5' primer: 5'-CCCTCCTAGACCCTGAGCTT-3'

CUX1 exon 4 gRNA 3' primer: 5'-TTCATGTGTCCTGCACTCCC-3'

CUX1 exon 23 gRNA 5' primer: 5' – GGTGAGGACCTGACATTCCG – 3'

CUX1 exon 23 gRNA 3' primer: 5′ – GGACCAAGGAACGGACCAAT – 3′

PCR products of the cut site were amplified for the exon4 and exon 23 gRNA editing and purified using the Qiagen PCR Purification Kit. The products were submitted for Sanger sequencing. DNA editing efficiency was assessed using TIDE analysis that estimated the percentage of indels in the pool.

For the intron 20 gRNAs, we devised a PCR confirmation strategy where we designed two primers spanning each cut site in the intron 20 region. The primers for each cut site were used to

amplify each cut site region to determine if each individual edit had taken place in a given single-cell clone. The 5' primer from the first cut site and the 3' primer from the second cut site were used to determine if the 2.56 kb region spanning the putative p75 TSS had been successfully deleted in a given single-cell clone.

Reverse-transcriptase PCR

RNA was extracted from 500,000 cells of the K562, Kasumi-1, KG-1, T47D, MDA-MB-231 and MCF-7 cell lines using Trizol, precipitated using chloroform and 70% ethanol, and purified using the RNeasy Mini kit (QIAGEN cat no 74104). Rneasy kit columns were pre-treated with DNase I to avoid contamination from DNA in the qPCR. DNase I was resuspended 1:8 in buffer RDD and incubated on the Rneasy column membrane for 15 minutes at room temperature. 500 ng of RNA from each cell line was used to synthesize cDNA using the Thermo Scientific Maxima™ H Minus cDNA Synthesis Master Mix Kit (Thermo Scientific cat no M1661). 1 µL of this synthesized cDNA was then used in the qPCR reaction. Primers were used specific to the p200 Cux1 isoform, the p75 Cux1 isoform, and GAPDH primers as a housekeeping control. The primers for the p75 transcript were obtained from the paper that originally described the existence of this isoform [141]. In addition, we also designed our own primers spanning various regions of intron 20 of CUX1. Controls tested include a no template water control, and a no RT enzyme control. 30 cycles of PCR were performed. Primer sequences are listed below:

GAPDH forward primer: 5' – ACCACAGTCCATGCCATCAC – 3'

GAPDH reverse primer: 5' – TCCACCACCCTGTTGCTGTA – 3'

p200 + p75 *CUX1* forward primer: 5' – CCGGAGGAGAAGGAGGCGCT – 3'

p200 + p75 *CUX1* reverse primer: 5’ – AGCTGTCGCCCTCCGAGCTG – 3’

CUX1 forward primer (primer 1): 5’ – CCACTCCGTGACATCGCTC – 3’

p75 *CUX1* forward primer (primer 2): 5’ - CCCTCATGTTAAGCCTTCCGA – 3’

p75 *CUX1* forward primer (primer 3): 5’-GCTATTTTCAGGCACGGTTTCTC – 3’

p75 *CUX1* reverse primer (primer 4): 5’ -TCCACATTGTTGGGGTCGTTC - 3’

We also performed qPCR using the primers listed above to look at delta CT values for the putative p75 transcript. qPCR was performed using SYBR Green reagent (Thermo Fisher) on a 7500 Fast Real-time machine (Applied Biosystems). Results were normalized to GAPDH for primers and the K562 cell line for template.

Immunoprecipitation

CUX1 was immunoprecipitated from the KG-1 cell line, and then blotted for CUX1 again to look at whether short CUX1 isoforms were able to be pulled down by the anti-CUX1 antibody. 100 x 10^6 cells were spun down for a CUX1 pulldown and a control IgG pulldown each. Cells were lysed in hypotonic buffer (5mM EDTA, 5mM EGTA, 5mM Tris-Cl) with protease inhibitor added (Roche complete mini-EDTA free). Pellets were passed through a 20-gauge needle and incubated on ice, then spun down. The supernatant was removed, and the pellet was resuspended in RIPA buffer with protease inhibitor added (Roche Complete). Pellets were again passed through a 27-gauge needle, incubated on ice and subsequently spun down. The supernatant was collected, and then incubated overnight at 4^{o}C on a rocker with either 12 µg of the anti-CUX1 antibody (B-10, Santa Cruz) or the rabbit IgG antibody. Protein A/G beads (Santa Cruz) were then added the following day to the supernatant, and incubated at 4^{o}C on a rocker for 1 hour. The

immunoprecipitated protein is then spun down, washed in cold PBS, resuspended in loading buffer, and subjected to SDS-PAGE.

Sample preparation for LC-MS/MS

For the whole cell lysate samples, 20 µg of whole cell extract from KG-1 cells (determined by Bradford assay, Thermo #1856209 using λ595nm) was loaded onto a 4-12% MOPS buffered 1D SDS-PAGE gel (Invitrogen NP0336BOX) and run at ~200V for ~45 min. For the IP samples, 30% of the IP eluate (30µL/100µL) was loaded. The gel was stained with Imperial Stain (Thermo #24615) for 1 hour at room temperature. For the whole cell lysate samples, the "p200" sections were excised from the gel by sterile razorblade (MW range 150-225kDa) and the "p75" sections were excised in the MW range of ~60-75kDa and in-gel trypsin digested as described below. For the IP samples, gel sections were excised as follows: "p200" sections MW range (~150-250kDa) and "p75" sections MW range (~65kDa-90kDa) and in-gel trypsin digested as described below.

Trypsin Digestion

Gel sections were washed in dH2O and destained overnight using 100 mM NH4HCO3 (Sigma #285099) pH 7.5 in 50% acetonitrile (Fisher A998SK-4). A reduction step was performed by addition of 100 µl 50 mM NH4HCO3 pH 7.5 and 10 µl of 200 mM tris(2-carboxyethyl) phosphine HCl (Sigma #C4706-2G) at 37 °C for 30 min. The proteins were alkylated by addition of 100 µl of 50 mM iodoacetamide (Sigma #RPN6320V) prepared fresh in 50 mM NH4HCO3 pH 7.5 buffer and allowed to react in the dark at 20 °C for 30 minutes. Gel sections were washed in Millipore water, then acetonitrile, and vacuum dried. Trypsin digestion was carried out overnight at 37 °C with 1:50-1:100 enzyme–protein ratio of sequencing grade-modified trypsin

(Promega #V5111) in 50 mM NH4HCO3 pH 7.5, and 20 mM CaCl2 (Sigma #C-1016). Peptides were extracted with 5% formic acid (Sigma #F0507-1L) in aqueous and 75% organic (ACN) combined and vacuum dried. Peptides were C18 cleaned up using C18 spin columns (Thermo #89870) and sent to the Mayo Clinic Medical Genome Facility Proteomics Core for HPLC and LC-MS/MS data acquisition via Q-Exactive Orbitrap (Thermo).

LC-MS/MS via MaxQuant

Peptide samples were re-suspended in Burdick & Jackson HPLC-grade water containing 0.2% formic acid (Fluka #60-006-17), 0.1% TFA (Pierce # 28903), and 0.002% Zwittergent 3–16 (Millipore Sigma #693023), a sulfobetaine detergent that contributes the following distinct peaks at the end of chromatograms: MH+ at 392, and in-source dimer [2 M + H+] at 783, and some minor impurities of Zwittergent 3-12 seen as MH+ at 336. The peptide samples were loaded onto a 100 µm x 40 cm PicoFrit column self-packed with 2.7 µm Agilent Poroshell 120, EC-C18, washed, then switched in-line with a 0.33uL Optimize EXP2 Stem Traps, packed spray tip nano column packed with Halo 2.7 µm Pep ES-C18, for a 2-step gradient. Mobile phase A was water/ acetonitrile/ formic acid (98/2/0.2) and mobile phase B was acetonitrile/isopropanol/water/ formic acid (80/10/10/0.2). Using a flow rate of 350 nl/min, a 90 min, 2-step LC gradient was run from 5% B to 50% B in 60 min, followed by 50%–95% B over the next 10 min, hold 10 min at 95% B, back to starting conditions and re-equilibrated.

Electrospray tandem mass spectrometry (LC-MS/MS) was performed at the Mayo Clinic Proteomics Core on a Thermo Q-Exactive Orbitrap mass spectrometer, using a 70,000 RP (70K Resolving Power at 400Da) survey scan in profile mode, m/z 340–1800 Da, with lockmasses, followed by 20 MSMS HCD fragmentation scans at 17,500 resolution on doubly and triply

charged precursors. Single charged ions were excluded, and ions selected for MS/MS were placed on an exclusion list for 60 seconds. An inclusion list (generated with in-house software) consisting of expected Cux1 sequences was used during the LC-MS/MS runs.

Database Searching

Tandem mass spectra MS/MS samples were analyzed using MaxQuant (Max Planck Institute of Biochemistry, Martinsried, Germany; version 1.6.17.0). MaxQuant was set up to search the 210308_SPROT_Human_UP5640.fasta database assuming the digestion enzyme strict trypsin. MaxQuant was searched with a fragment ion mass tolerance of 20 PPM and a parent ion tolerance of 20 PPM. Carbamidomethyl of cysteine was specified in MaxQuant as a fixed modification. Deamidated of asparagine and glutamine, oxidation of methionine, formyl of the peptide n-terminus (to account for formic acid in the mobile phase), acetyl of the n-terminus, phosphorylation of serine and threonine, and Ubiquitylation of lysine (GlyGly) were specified in MaxQuant as variable modifications. Maxquant 1FDR results files were processed in Perseus (version 1.6.14.0) for the proteingroups.txt in addition to being imported into Scaffold version 5.0.1.

Criteria for Protein Identification

Scaffold (version Scaffold_5.0.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted at 1% FDR by the Peptide Prophet algorithm (Keller, A et al Anal. Chem. 2002;74(20):5383-92) with Scaffold delta-mass correction. Protein identifications were accepted if they could be established at 1% FDR and contained at least 1 identified peptide. Protein probabilities were assigned by the Protein Prophet algorithm [178]. Proteins that contained similar peptides and could not be

differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters. The mass spectrometry proteomic data sets (MK1) and (MK3) were uploaded to the ProteomeXchange consortium via the PRIDE partner repository with the dataset identifier PXD027527.

Analysis of publicly available functional genomics datasets

ENCODE ChIP-seq data for H3K27ac, H3K4me3 and H3K4 me1 in 7 Tier 1 cell lines, and the MCF-7 cell line was mapped to hg19, and visualized on the UCSC Genome Browser. The MCF-7 ChromHMM 18-state model track was downloaded from ENCODE as a bigBed file, and uploaded onto the UCSC Genome Browser. To assess DNase accessibility, we used the uniform DNase hypersensitivity track on ENCODE that incorporates DNase I hypersensitivity sites, formaldehyde-assisted isolation of regulatory elements (FAIRE)-sequencing, and ChIP-seq for several regulatory factors across 125 human cell lines. Peaks in this track were subjected to a 1% FDR cutoff, and are visualized as gray boxes. The extent of darkness signifies the signal intensity. Additionally, we also included separate DNase accessibility tracks in the K562, MCF-7 and T-47D cell lines, especially to assess accessibility in two cell lines previously described to express p75 (MCF-7 and T-47D).

To assess transcription factor binding at the putative p75 TSS, we included the Transcription Factor ChIP-Seq Clusters track from ENCODE, that shows the binding of 338 transcription factors in 130 cell types from 1264 ChIP-seq experiments performed between February 2011 and November 2018. Aa described above, each gray box indicates binding of a transcription factor, and the darkness of the box is proportional to the maximum signal observed in any cell type. ENCODE RNA-seq from 9 Tier 1 cell lines shows transcription levels measured

by sequencing of polyadenylated RNA. In addition, RNA-seq specifically from human breast tissue, MCF7 and T-47D cell lines generated from the Burge lab and mapped to the genome using GEM mapper by the Guigo lab was visualized aligned to the hg19 genome assembly [179].

For the CpG island track, CpG islands were predicted by searching the sequence one nucleotide at a time, and CpG islands were marked as such if they had a GC content of 50% or greater, length greater than 200 bp, and ratio of observed number of CG dinucleotides to expected number greater than 0.6.

The Eukaryotic promoter database (EPD) track uses gene transcript coordinates from multiple sources (HGNC, GENCODE, Ensembl and RefSeq) and validates this data using CAGE and RAMPAGE experimental studies using FANTOM5, UCSC and ENCODE. For gene assembly and gene prediction databases, we looked at all possible CUX1 isoforms described theoretically and experimentally on RefSeq, GENCODE, Ensembl, Augustus and CCDS. CAGE-seq data on the FANTOM browser was utilized to look at 5' capped mRNAs arising from the CUX1 gene.

CHAPTER 3

ORTHOGONAL APPROACHES CONTROVERT THE EXISTENCE OF A CUX1 P75 ISOFORM

3.1 INTRODUCTION

Protein isoforms and splice variants often have important and distinct biological functions. Since protein isoforms, by definition, have considerable sequence homology and may be expressed at different levels within the cell, it can be challenging to accurately differentiate between and functionally characterize such isoforms [180]. One such protein reported to have multiple isoforms is CUT-like homeobox 1 (CUX1), a HOX-family transcription factor with critical roles in development and tumorigenesis. In vertebrates, the *CUX1* locus contains two distinct genes that partially share exons: *CUX1*, which encodes a transcription factor localized to the nucleus, and *CASP*, which encodes a golgi-associated transmembrane protein involved in retrograde transport [44–46]. Altered levels and mutations of the *CUX1* transcription factor have been implicated in cancer across several tumor types and species [181,182]. *CASP*, on the other hand, has not been implicated in human disease [47,48]. The RefSeq database documents seven mRNA isoforms for the human *CUX1* locus; five of these are *CASP* transcripts and two are *CUX1* (Figure 3.1). Due to its relevance to human health, we focus our attention herein on *CUX1*. For the sake of simplicity, our subsequent references to the *CUX1* gene or mRNA allude to those isoforms that encode CUX1, unless stated otherwise.

CUX1 is highly conserved, ubiquitously expressed, and essential for survival in mice and *Drosophila* [183]. CUX1 controls many cellular processes including determination of cell identity, cell cycle progression, cell-cell communication, and cell motility [183]. In cancer, however, there are conflicting reports of *CUX1* acting alternately as an oncogene or tumor suppressor gene [182].

To resolve this discrepancy, we hypothesized that distinct CUX1 protein isoforms explain these disparate functions.

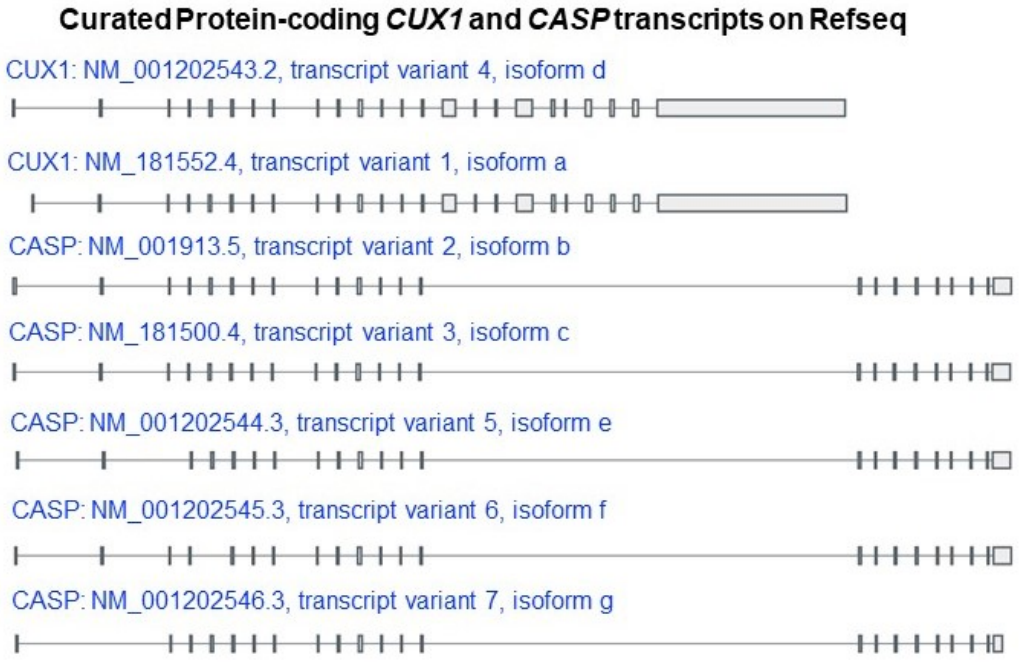


Figure 3.1: Schematic of seven predominant Refseq transcripts curated to be transcribed from the *CUX1* gene

The two RefSeq-annotated *CUX1* mRNA transcripts vary only by alternative first exons and encode a full-length protein of 1505 amino acids length, described in the literature as p200 (Figure 3.2). p200 CUX1 has four DNA-binding domains, comprised of three CUT-repeat domains and one homeodomain (Figure 3.2). A truncated p110 CUX1 isoform is generated by post-translational proteolytic processing of full-length p200 CUX1 by cathepsin L (Figure 3.2) [144]. This cleavage occurs during the S phase in normal cells, and can become constitutive in transformed cells [51,144]. p110 CUX1 lacks one CUT-repeat domain and the N-terminal region but retains the three C-terminal DNA-binding domains. A third isoform, p75 CUX1, is reported to

arise from an alternative TSS embedded within intron 20 and retains one CUT-repeat and the homeodomain (Figure 3.2) [141,182]. p75 has been identified in human breast cancer cell lines and mouse thymocytes [141]. Despite fewer DNA binding domains, p75 and p110 bind DNA more stably than p200 [51,146]. Rarer CUX1 isoforms have been described to be generated by post-translational proteolytic processing; p80, p90 and p150 CUX1 (*101-103*). However, these isoforms are less well characterized, and it is unclear if they bind DNA and exert transcriptional activity [52,146,154].

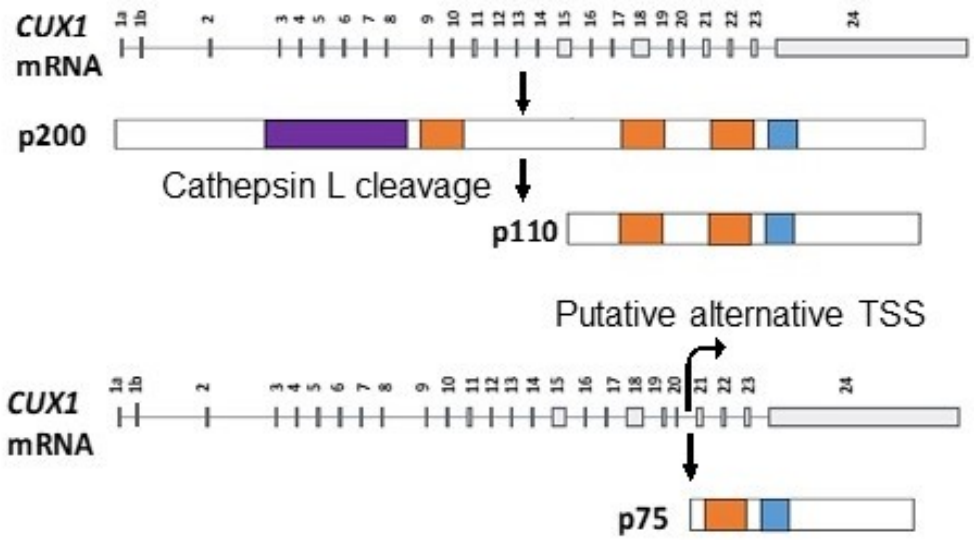


Figure 3.2: Schematic representation of the *CUX1* mRNA and main protein isoforms. There are two *CUX1* mRNA transcripts that vary only by the alternative first exons (1a and 1b). CUX1 encodes a full-length protein of 1505 amino acids which runs at 200 kDa (p200). A truncated p110 CUX1 protein is reported to be generated by proteolytic cleavage by cathepsin L. The p75 CUX1 isoform is reported to arise from an alternative transcription site embedded within intron 20.

By some reports, *CUX1* is thought to be oncogenic in cancer. Over-expression of short p75 or p110 CUX1 isoforms in fibroblasts and breast cancer cells causes increased proliferation, cell cycle progression, and tumor formation *in vivo* [51,52,71,90,141,142,151]. p75 CUX1 transgenic mice engendered a higher proportion of adenosquamous mammary carcinomas and lung metastases compared to the p110 or p200 transgenic mice [90]. On the other hand, large scale cancer genome resequencing efforts demonstrate *CUX1* inactivating mutations or deletions more characteristic of a tumor suppressor [132]. *CUX1* deletions and inactivating mutations are prevalent across cancer types [129,132,184]. In myeloid malignancies, *CUX1* falls within the commonly deleted region of chromosome 7q22 [112]. Consistent with a tumor suppressor role, CUX1 knockdown in mouse models leads to MDS/MPN that is reminiscent of human disease [63,136]. CUX1 knockdown in human hematopoietic stem cells provides an engraftment advantage in immunodeficient mice [112]. Even in *Drosophila* models, the *CUX1* orthologue, *cut*, exerts tumor suppressor activity [112,129].

Given the clinical significance, it is critical to parse out the putative oncogenic and tumor suppressive roles of *CUX1*. We reasoned that uncovering isoform-specific properties would reveal therapeutic strategies for inhibiting oncogenic CUX1 isoforms or promoting the expression of tumor suppressive isoforms to treat malignancies with *CUX1* alterations. To this end, we characterized the CUX1 isoforms in a panel of human cell types and leveraged publicly available functional genomic datasets across a spectrum of tissue types. To our surprise, we identified no evidence supporting the existence of the p75 mRNA isoform in any cell type examined and demonstrate that p75 is likely a western blotting artefact. Focusing on hematopoietic cells, we only identify the p200 isoform. Our data indicate that in hematopoietic cells, the tumor suppressive role of *CUX1* is attributable to the full-length protein. In addition,

the lack of evidence for a p75 transcript calls into question prior interpretations of studies using exogenously overexpressed short isoforms that ascribe an oncogenic role for CUX1.

3.2 RESULTS

3.2.1 Human hematopoietic cells only express the p200 CUX1 isoform

Given the relevance of CUX1 to myeloid malignancies, we first sought to identify the CUX1 isoforms expressed in human AML cells. Immunoblotting with a commercially available mouse monoclonal antibody that recognizes an epitope shared across CUX1 isoforms (clone B-10, Figure 3.3), reveals six of eight AML cell lines express a dominant p200 CUX1 band (Figure 3.4). We also observed a less prominent 75 kDa band in five cell lines (Figure 3.4). p200 was also the predominant isoform in primary human CD34+ hematopoietic stem and progenitor cells, the normal counterpart thought to give rise to myeloid malignancies (Figure 3.5). To discern if the p75 band we observed corresponds to that described previously, we assessed cell lines reported to express p75 or p110: murine NIH-3T3 fibroblasts (p110), and MCF7, T47D, and MDA-MB-231 human breast cancer cell lines (p75) [141]. In contrast to prior findings, we did not observe a short isoform protein band in any of these cell lines previously reported to express p75 (Figure 3.6).

We next blotted AML cell lines with other CUX1 antibodies to verify if we could detect the p75 band with other CUX1 antibodies. We used a rabbit polyclonal antibody (PUC) generated in-house that recognizes aa 1223-1242 of CUX1, and another commercially available rabbity polyclonal antibody, ABE217, that recognizes aa 861 of CUX1 (Figure 3.3) [140]. ABE217 is expected to detect only the full-length CUX1 protein, while PUC binds an epitops that is shared by all CUX1 isoforms, similar to B-10. There were several background bands observed with the PUC antibody, but no dominant p75 band (Figure 3.7). The faint band at 75kDa seen with the ABE217 antibody cannot be the p75 isoform, as the antibody binds an epitope of CUX1 upstream of the p75 protein sequence (Figures 3.3, 3.8).

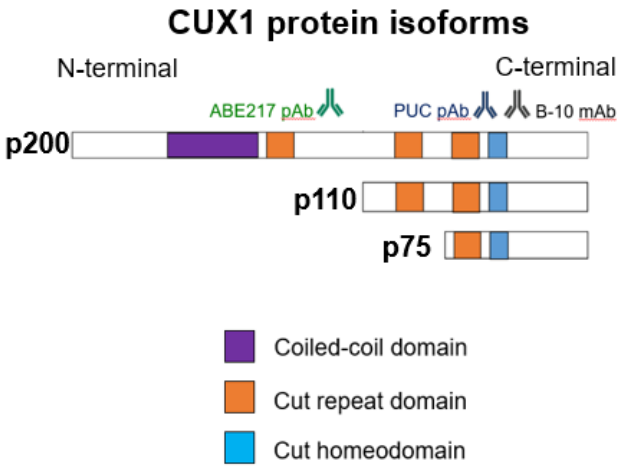


Figure 3.3: Schematic representation of the predominant CUX1 protein isoforms, with protein domains indicated, and the CUX1 antibodies used in this study. B-10 is a commercially available mouse monoclonal antibody derived against amino acids 1308-1332 of CUX1, and would thus be expected to recognize all CUX1 isoforms. PUC is an in-house generated rabbit polyclonal antibody recognizing amino acids 1223-1242 of CUX1, and would also be expected to recognize all CUX1 isoforms. ABE217 is a commercially available rabbit polyclonal antibody derived against amino acid 861 of CUX1, and only recognizes the full-length p200 protein. All of the antibodies used in this study are specific to CUX1, and do not detect CASP, the other protein encoded by the CUX1 gene.

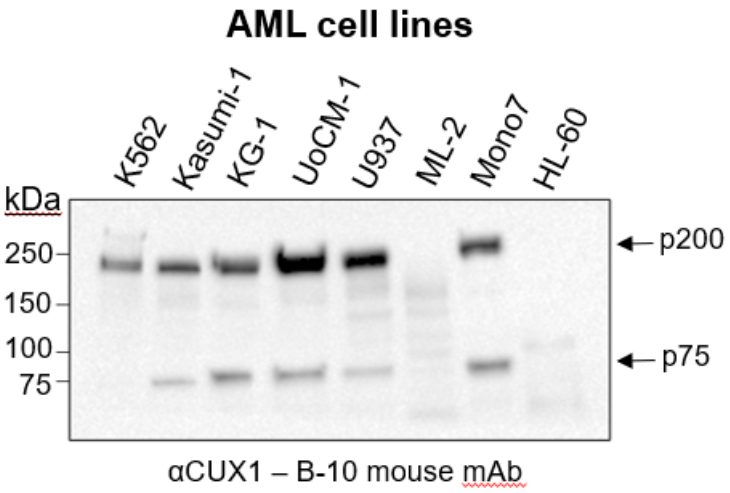


Figure 3.4: Immunoblot of CUX1 in the indicated human AML cell lines, using the B-10 antibody (n=3). 10µg of protein was loaded for the K562 and Kasumi-1 cell line, and 15 µg of protein was loaded for all other cell lines.

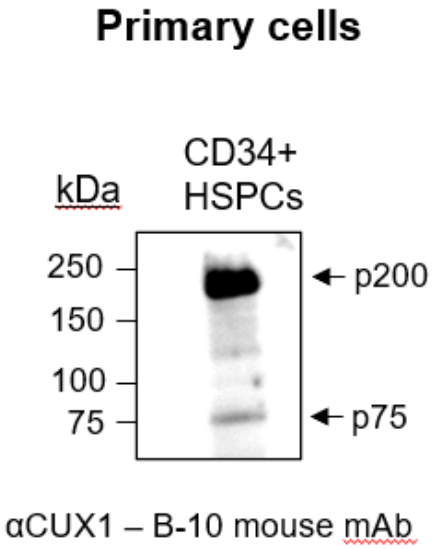


Figure 3.5: Immunoblot of CUX1 in human CD34+ HSPCs using the B-10-HRP antibody (n=3)

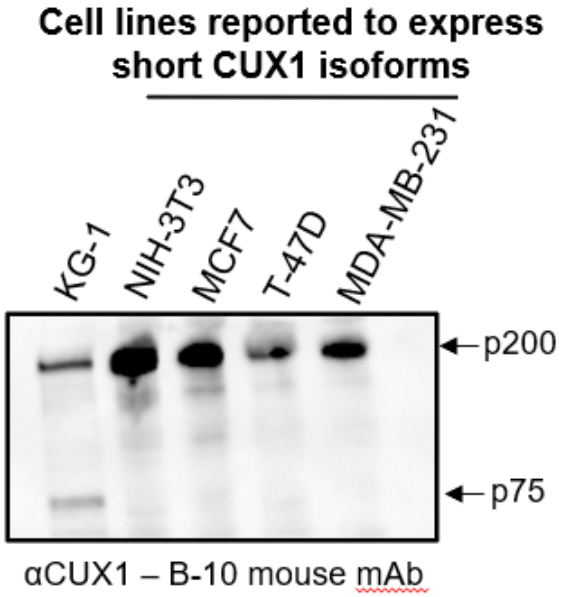


Figure 3.6: Immunoblot of CUX1 in the NIH-3T3 fibroblast line and several human breast cancer cell lines previously reported to express p75 CUX1 using the B-10-HRP antibody (n=3).

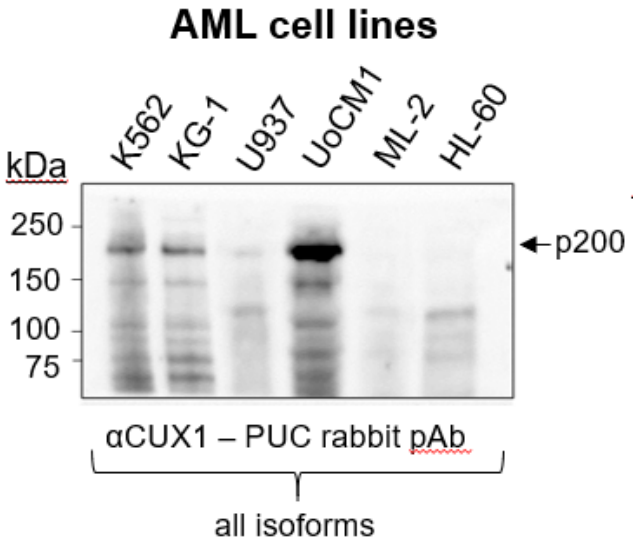


Figure 3.7: Immunoblot of CUX1 in indicated human AML cell lines, using the PUC antibody (n=3).

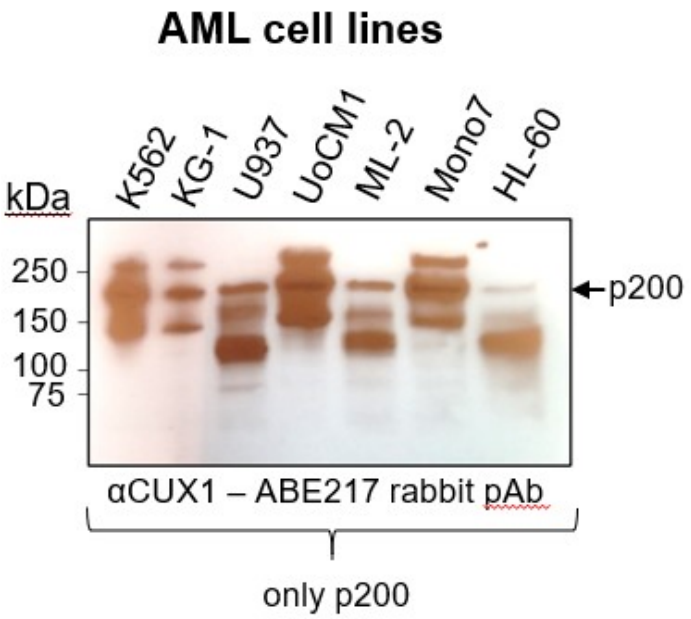


Figure 3.8: Immunoblot of CUX1 in indicated human AML cell lines, using the ABE217 antibody (n=3).

To circumvent potential antibody artefacts, we took an alternative approach to determine if the p75 band is in fact CUX1. KG-1 cells have a partial deletion of chromosome arm 7q that includes *CUX1*, thus they are mono-allelic for *CUX1*. We tagged the remaining endogenous CUX1 allele with an in-frame, C-terminal GFP tag by CRISPR/Cas9 homology-mediated repair, such that all described CUX1 isoforms would be tagged with GFP (Figure 3.9). Probing these cells with an anti-GFP antibody only identified a p200 CUX1 band unique in the tagged cell line (Figure 3.10). A p75 band tagged with GFP, which is 27 kDa in size, would be expected to run around 100 kDa. We do see such a band in the lane with GFP-tagged CUX1, but we also see the corresponding band in the lane with untagged, wild type CUX1, indicating that this band is non-specific. Probing CUX1 with the ABE217 antibody, only revealed one band supershifted by GFP tagging, corresponding to p200 CUX1. We also used a CUX1-mCherry tagged mouse model to

look at potential CUX1 isoforms that were tagged with mCherry. Recapitulating our observation in the GFP-tagged cell line, we only observe a tagged band indicative of p200 CUX1 (Figure 3.11). Together, this data indicates that human AML and hematopoietic stem and progenitor cells express p200 and not shorter CUX1 isoforms.

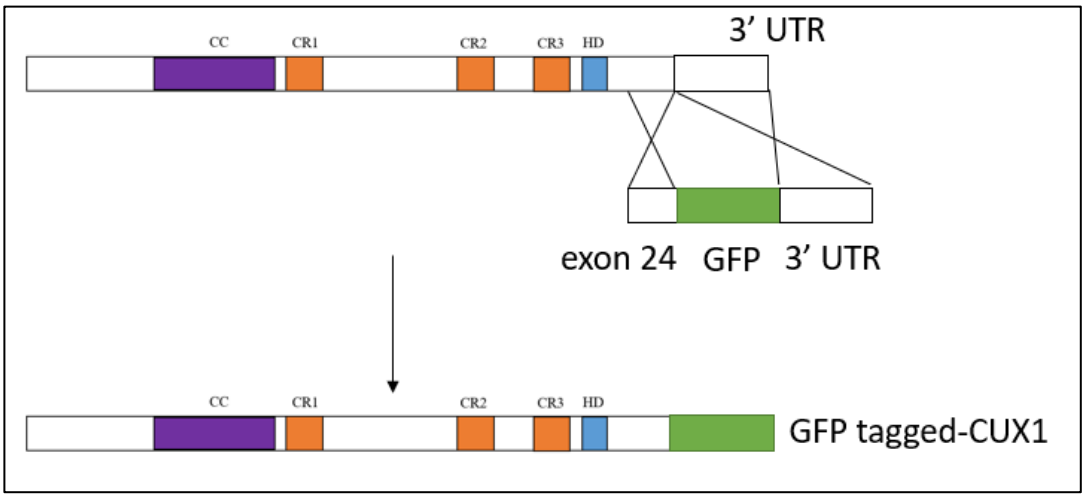


Figure 3.9: Approach to tag CUX1 C-terminally with GFP. A HDR template consisting of homology arms encompassing exon 24 of CUX1 right before the stop codon and the 3'UTR flanking the coding sequence of GFP was used along with a gRNA targeting exon 24 and Cas9 to mediate GFP insertion in frame with exon 24 of CUX1

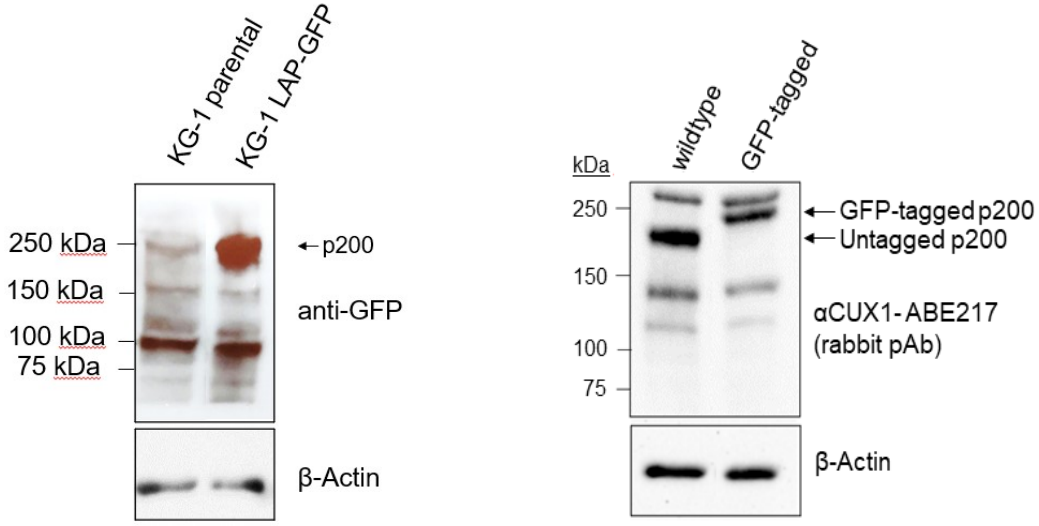


Figure 3.10: (Left) Immunoblot of GFP in a KG-1 cell line where CUX1 is C-terminally tagged with GFP. Protein from unedited KG-1 cells is also included (n=3). Blot is cropped from the

Figure 3.10 (Continued)
same gel to remove an irrelevant lane. (Right) Immunoblot of CUX1 in a KG-1 cell line where CUX1 is C-terminally tagged with GFP. Protein from wildtype KG-1 cells is also included (n=3). Blot is cropped from the gel to remove an irrelevant lane.

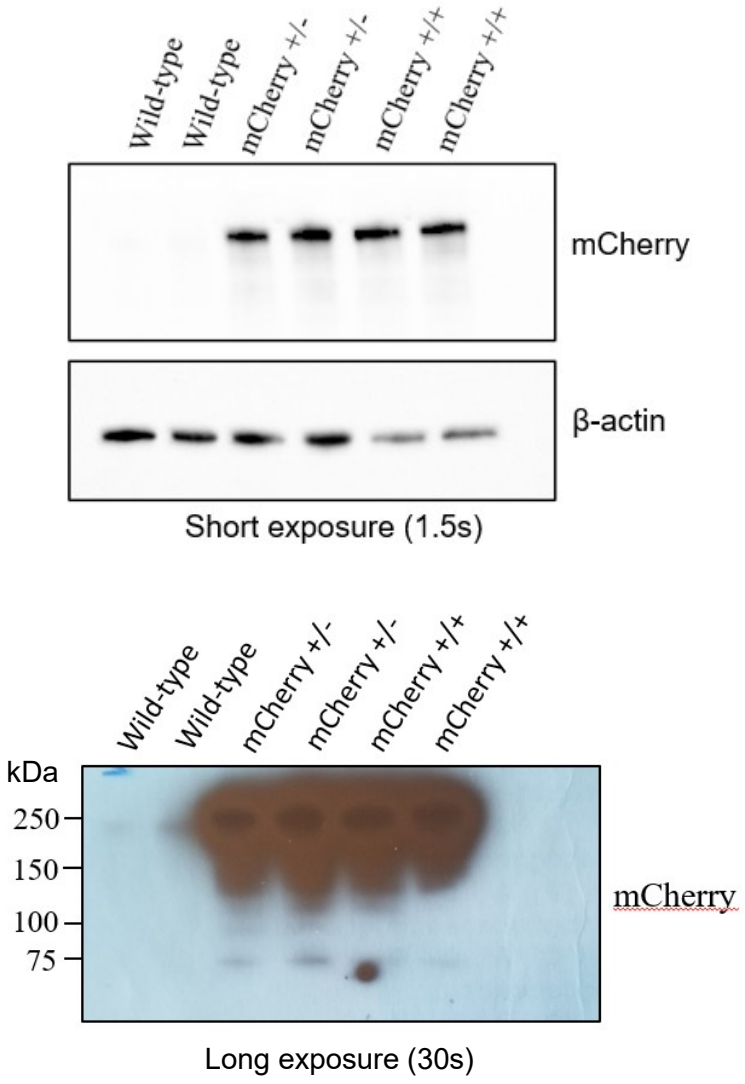


Figure 3.11: (Top) Short exposure (1.5 s) of a representative immunoblot for mCherry in thymus tissue of a CUX1-mCherry tagged reporter mouse. Protein from wild-type untagged mice and mice heterozygous for the CUX-tagged allele are also included. (Bottom) Longer exposure of the same blot from above indicates that there is no band representative of an mCherry-tagged p75 CUX1 isoform migrating around the expected size of ~101 kDa.

3.2.2 CRISPR/Cas9 genomic editing precludes the existence of a CUX1 p75 isoform

To further interrogate if the short isoforms observed are encoded by the *CUX1* locus or are a non-specific western blotting artefact, we devised several CRISPR/Cas9 strategies to selectively target genomic DNA encoding p75, p200, or both (Figure 3.12). We used the RNP-based CRISPR/Cas9 delivery system to ensure facile and highly efficient gene editing [185,186]. We used the KG-1 cell line as it has only one CUX1 allele, enabling facile monoallelic editing, and expresses both p75 and p200 bands. As illustrated in Figure 3.12, we designed a gRNA targeting exon 4 of CUX1 which is only expressed in the genomic region encoding the p200 isoform to selectively delete the p200 protein. We identified multiple single-cell CRISPR-edited clones that had a complete loss of p200 CUX1 at the genomic level with ICE analysis (Figure 3.13). This editing was confirmed at the protein level with western blotting, best appreciated with the ABE217 antibody (Figure 3.13). The B-10 antibody also shows a loss of p200, with a residual non-specific band migrating at a slightly higher molecular weight (Figure 3.14). The expression level of the p75 band was unchanged, as expected based on our targeting strategy (Figure 3.14).

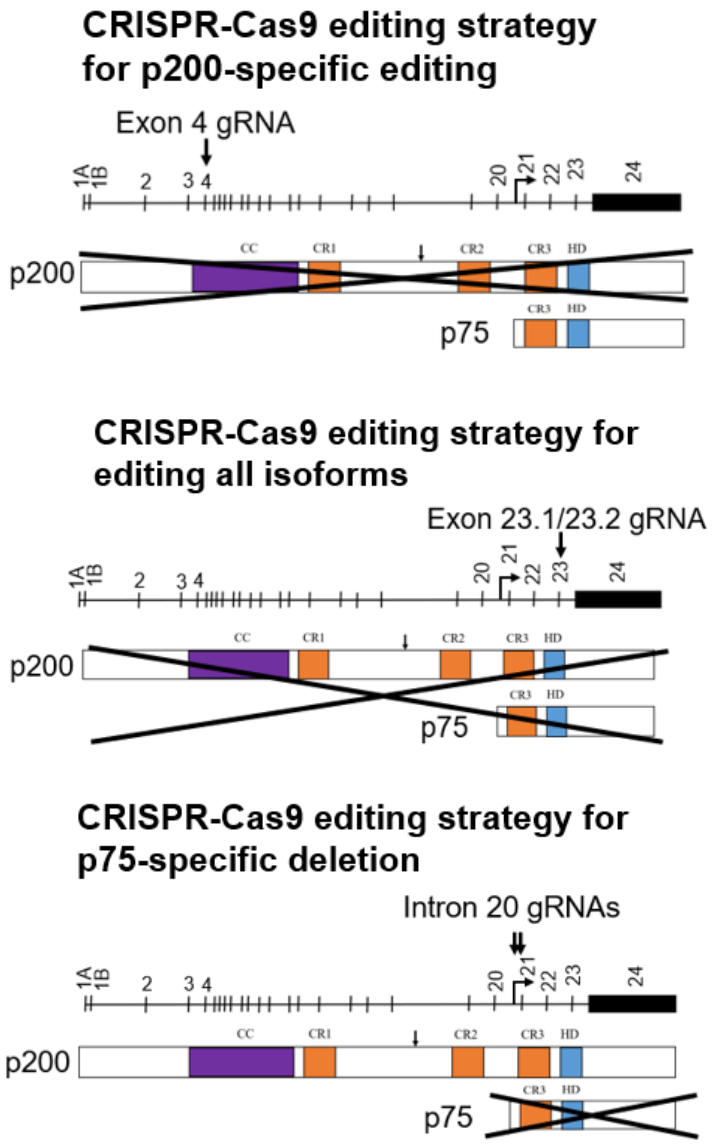


Figure 3.12: CRISPR/Cas9 editing approach targeting exon 4 to selectively edit p200 in the KG-1 cell line, which only has one copy of *CUX1* (top); CRISPR/Cas9 editing approach targeting exon 23 of CUX1 to delete all CUX1 isoforms in the KG-1 cell line (middle); CRISPR/Cas9 editing approach using two gRNAs flanking the predicted p75 TSS within intron 20 to selectively delete p75 in the KG-1 cell line (bottom).

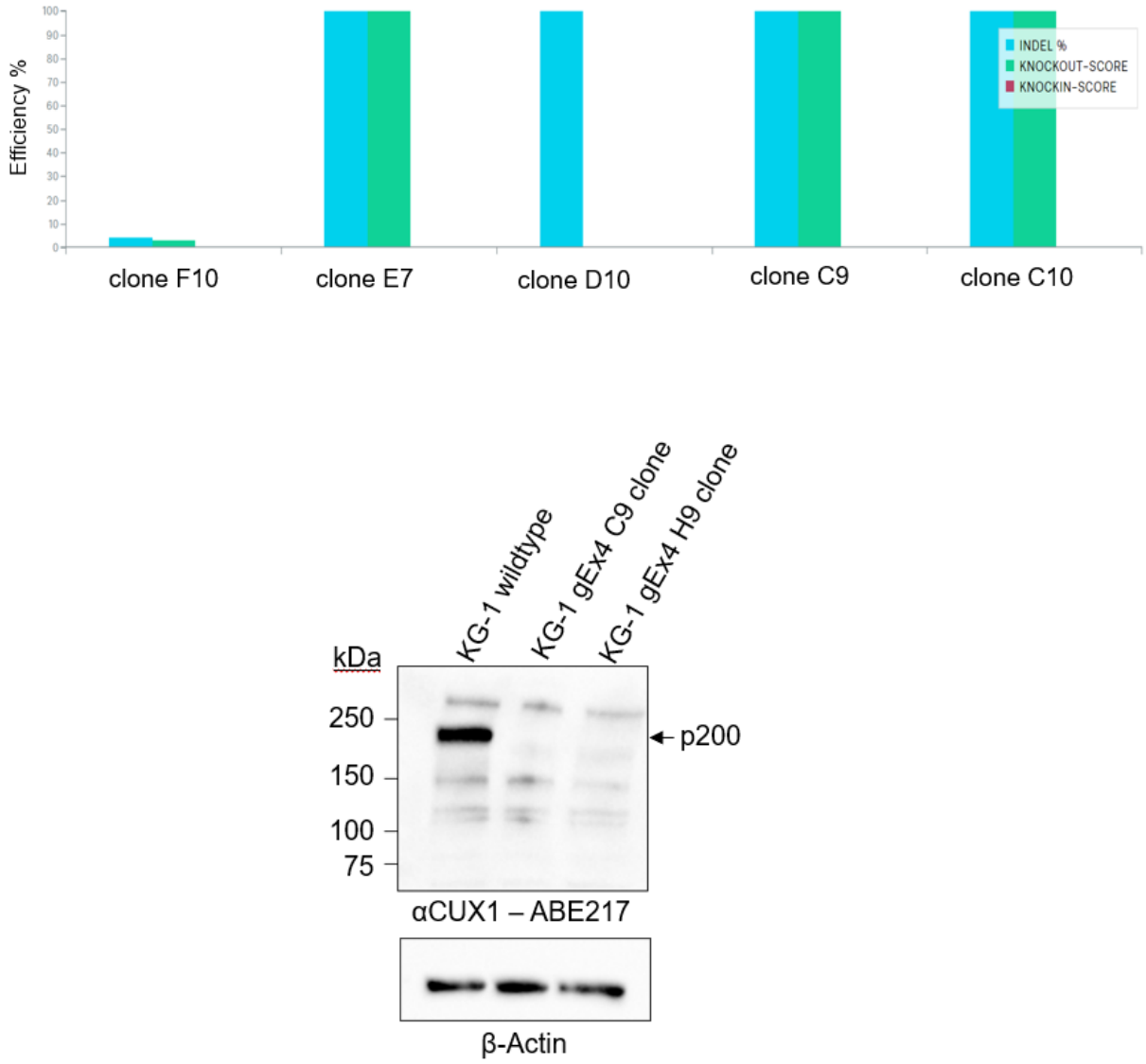


Figure 3.13: (Top) ICE editing efficiency for KG-1 clones edited with a gRNA targeting exon 4 of CUX1 shows three out of 5 clones are successfully edited. (Bottom) Immunoblot for CUX1 in KG-1 single cell clones edited with a gRNA targeting exon 4 of CUX1 using the ABE217 antibody.

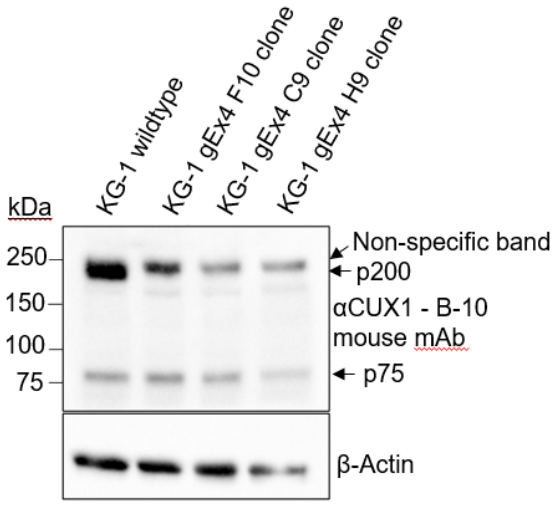


Figure 3.14: Immunoblot for CUX1 in KG-1 single cell clones edited with a gRNA targeting exon 4 of *CUX1* using the B-10-HRP antibody.

As a control, we designed a gRNA targeting exon 23 of CUX1 (gEx23.1) to delete all CUX1 isoforms (Figure 3.12). As exon 23 is shared by all *CUX1* isoforms, including p75, it would be expected to disrupt all bona fide CUX1 proteins. Transfection with gEx23.1 followed by single cell cloning identified a clone with a single base pair insertion generating a frameshift mutation (Figure 3.15). This abolished expression of the p200 CUX1 band, but was ineffective at deleting the shorter p75 band when probed with the B-10 antibody (Figure 3.15). The B-10 CUX1 antibody binds an epitope after the exon 23 gRNA cut site, and the fact that we could still observe the p75 band indicates that the p75 band observed is not encoded by the CUX1 locus.

Since the B-10 antibody cannot bind to CUX1 when edited with the gRNA targeting exon 23, we blotted for CUX1 in the gEx23.1-edited KG1 clone with our other CUX1 antibody, ABE217, which can still bind its epitope when exon 23 of CUX1 is disrupted. We observed that

the band for p200 CUX1 seemed to be shifted downward with the ABE217 antibody, indicating generation of a C-terminal truncated protein with this gRNA (Figure 3.16). The expected

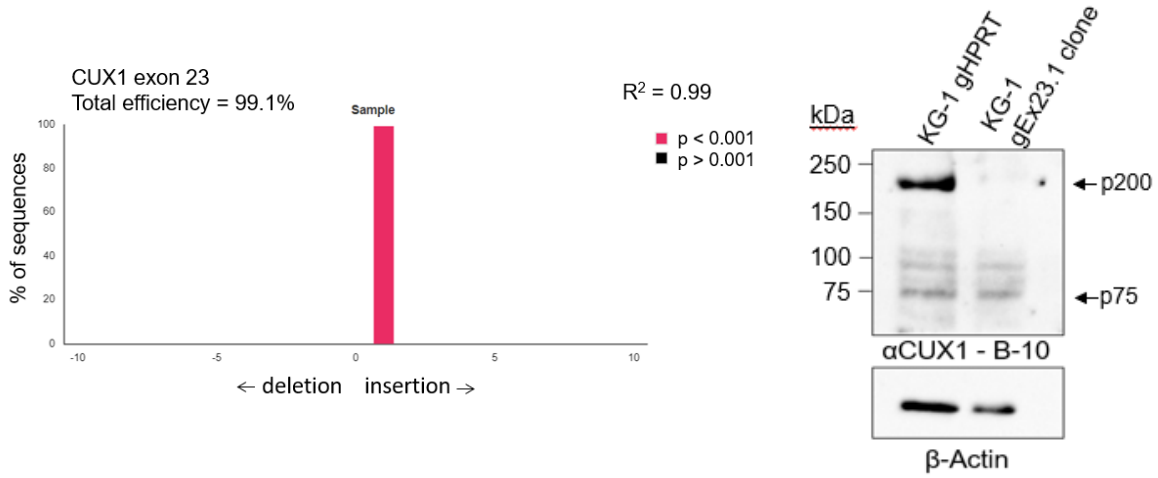


Figure 3.15: TIDE editing efficiency of KG-1 gEx23.1 clone indicates that the clone has a 1bp insertion that will result in a frameshift mutation and disruption of protein expression. Immunoblot for CUX1 on the right in the KG-1 clone edited with a gRNA targeting exon 23 of *CUX1* using the B-10 antibody (n=3).

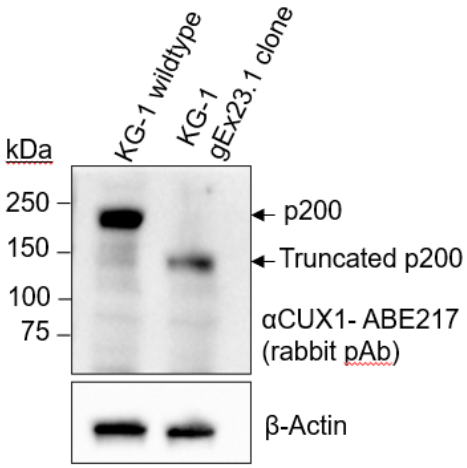


Figure 3.16: Immunoblot for CUX1 in a KG-1 clone edited with a gRNA targeting exon 23 of CUX1 using the ABE217 antibody (n=3).

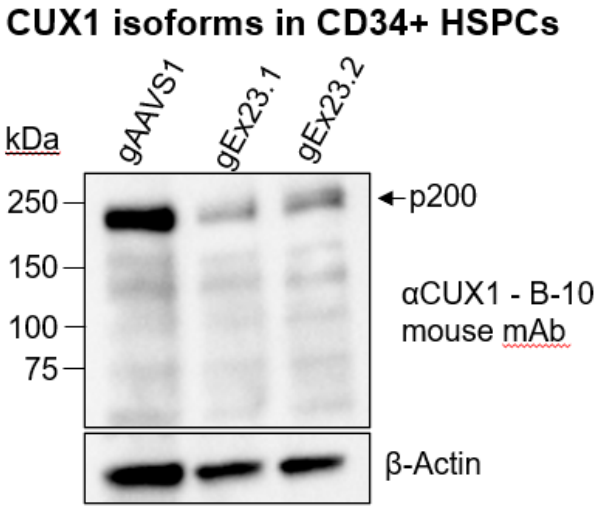


Figure 3.17: Immunoblot for CUX1 using the B-10 antibody in human CD34+ HSPCs. Bulk populations were edited with a control AAVS1 gRNA, or with one of two different gRNAs targeting exon 23 of CUX1 (n=3).

downward size shift for the gEx23.1-edited KG1 clone is approximately 28 kDa, leading to a truncated p200 CUX1 band. The predicted size of CUX1 is ~167 kDa but runs at 200 kDa presumably due to post-translational modifications; the truncated protein ran at ~140 kDa, consistent with a 28kDa deletion. (Figure 3.16). We were able to detect this truncated protein only with the ABE217 antibody, as the antibody binds an epitope before the CRISPR cut site with gEx23 (Figure 3.3). We also deleted CUX1 using the exon 23 gRNA in CD34+ human HSPCs, and saw no change in expression of any of the other bands other than p200 CUX1 (Figure 3.17). We used two different gRNAs targeting exon 23 in the CD34+ cells, and observed similar levels of knockdown using both gRNAs. The residual p200 band in the deletion lanes is likely due to cells that were not edited, as these are pooled HSPCs and not single cell clones.

We considered the possibility that p75 does not contain exon 23, perhaps due to alternative splicing. As a different approach, we designed a pair of gRNAs flanking the predicted intronic p75 transcriptional start site mapped to be ~2.5 kb upstream of exon 21 of CUX1 (Figure 3.12) [141]. We reasoned that eliminating the putative intronic transcription start site would remove the p75 isoform while leaving expression of the p200 isoform intact. We deleted approximately 2.56 kb of intronic DNA, leaving 79 base pairs intact proximal to exon 21. We employed a PCR screening strategy using primers spanning both gRNAs to screen for a deletion band and observed deletion bands in 1 out of 6 clones scored (Figure 3.18). Immunoblotting of this successfully deleted single-cell clone (KG-1 Δp75 #21) indicates no change in p75 (Figure 3.19). This suggests that this 2.56 kb segment of DNA does not harbor the putative p75 TSS, and is incongruent with the previous report [141]. In summary, these experiments show that the presumptive p75 isoform contains neither exon 23 nor a TSS within 2.56 kb upstream of exon 21 as originally described, further implicating the p75 band as a western blotting artefact.

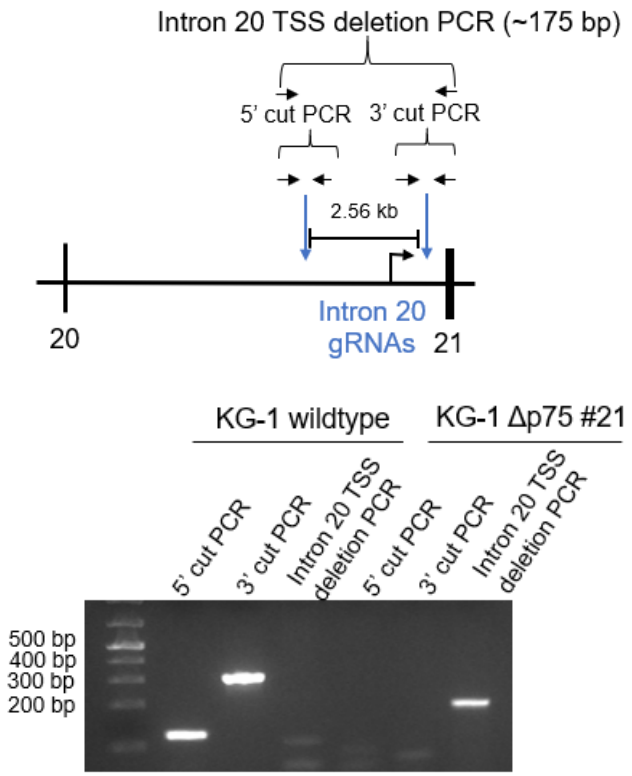


Figure 3.18: PCR screening strategy for identifying single cell clones with deletion of the p75 putative TSS. PCR products from primer pairs spanning the first cut site or the second cut site in intron 20 indicate lack of successful deletion. A PCR product from primers spanning the deletion site indicate a successful deletion.

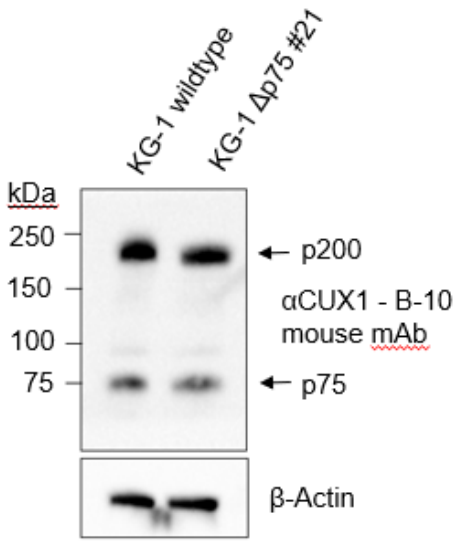


Figure 3.19: Immunoblot for CUX1 using the B-10 antibody in a KG-1 single cell clone with successful deletion of the p75 TSS (n=3).

3.2.3 Proteomics approaches do not support the existence of CUX1 p75

We considered that the p75 artefact results from the denaturing conditions of western blotting. To test this, we performed a CUX1 immunoprecipitation (IP) in the KG-1 cell line to look at the bands that were pulled down by our CUX1 antibody. In the IP, the B-10 anti-CUX1 antibody binds to CUX1 protein in its native state, and thus will pull down the protein by its exposed epitope. Boiling proteins for immunoblotting denatures proteins and reveals epitopes that are normally hidden in the native protein structure, and might thus reveal epitopes that are not specific to CUX1 present in the nuclear extract that are inadvertently bound by the antibody. We observe the p75 protein band in our input controls, but do not observe the p75 band when we immunoprecipitate CUX1 and blot for CUX1 expression (Figure 3.20).

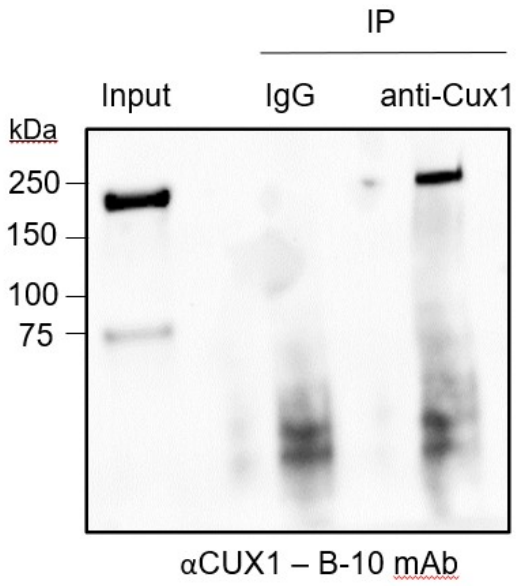


Figure 3.20: B-10 immunoblot after B-10 immunoprecipitation of CUX1 in the KG-1 cell line. KG-1 cells were also immunoprecipitated with a mouse anti-IgG antibody as a negative control. Input control is also included (n=3). Blot is cropped on the edge to remove a redundant lane.

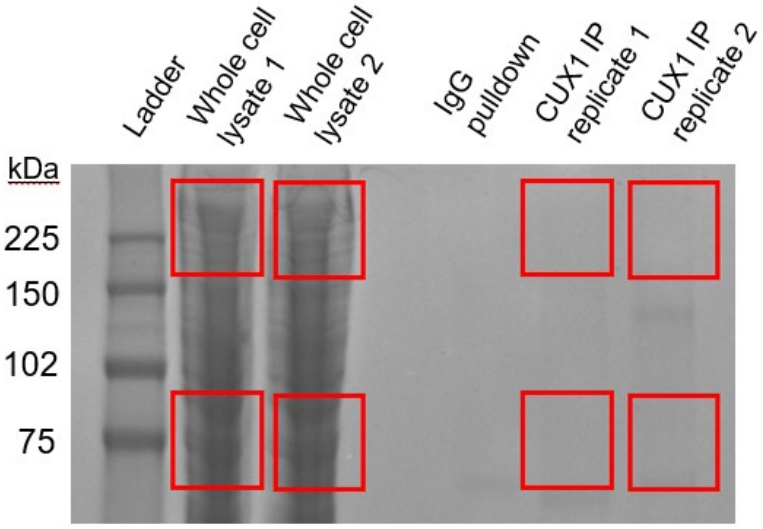


Figure 3.21: p200 and p75 regions (red rectangles) were excised from Imperial stained SDS/PAGE of two whole cell extract replicates in the KG-1 cell line, and two B-10 immunoprecipitated CUX1 replicates from the KG-1 cell line.

Next, we looked at other proteomic approaches to characterize the legitimacy of the 75 kDa band that was routinely expressed in our AML cell line CUX1 western blots. To verify if any peptides from this band mapped to the putative p75 protein sequence, we utilized two approaches to study peptides mapping to CUX1. In the first approach, we immunoprecipitated CUX1 using the B-10 antibody or prepared whole cell lysate from KG-1 cells, and subject the samples to SDS-PAGE. LC-MS/MS was performed on regions corresponding to 200kDa and 75kDa excised from the Imperial stained protein gel for both conditions. (Figure 3.21). In the 200 kDa band after CUX1 immunoprecipitation, we observed 51 peptides (21 unique, shown as blue rectangles) across both replicates that mapped to the CUX1 protein sequence (Figure 3.22). In the p75 band of the immunoprecipitated samples, we observed one peptide that mapped to the N-terminal region of CUX1, but no peptides mapped to the C-terminal region described to

consist p75 [141] (Figure 3.22). Our proteomic analysis of whole cell lysate (Figure 4B, 1st two lanes) revealed 8 peptides (5 unique, shown as blue rectangles) from the p200 band mapping to the full-length Cux1 protein sequence (Figure 3.23 and Supplementary Table 1). In the p75 band isolated from lysates, we observed seven peptides in total and three unique peptides that mapped to the CUX1 sequence. One unique peptide mapped to the N-terminal CUX1 region while the other two unique peptides mapped to the CASP protein sequence (exons 25-33), which is also transcribed from the same gene as CUX1 (Figure 3.1). Since CASP shares sequence homology with CUX1 and is 77kDa in size, this explains why our proteomic analysis detected peptides mapping to CASP in our whole cell lysate sample. We did not detect any peptides from the p75 band in the lysates that mapped to the p75 sequence described previously. Taking this data

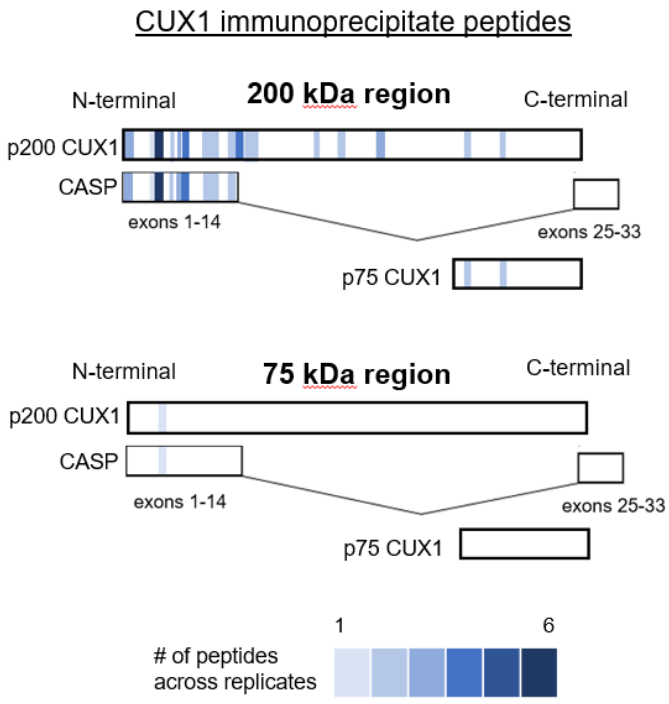


Figure 3.22: Schematic showing the number and distribution of peptides mapping to CUX1 and CASP protein sequences from the p200 and p75 bands excised from immunoprecipitated CUX1

Figure 3.22 (Continued)
run on a gel (n=2). Heat map at the bottom depicts the number of each peptide detected across replicates. Peptides with ambiguous assignments are shown in both potential candidate isoforms.

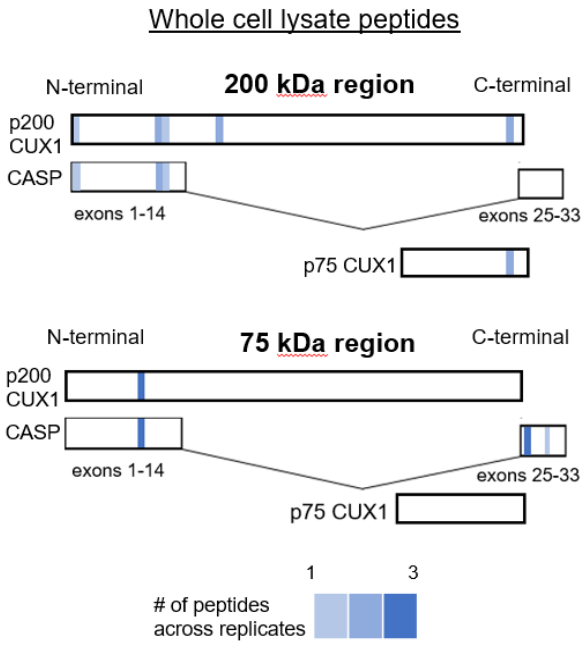


Figure 3.23: Schematic showing the number and distribution of peptides mapping to CUX1 and CASP protein sequences from the p200 and p75 bands excised from whole cell lysate samples (n=4). Heat map at the bottom depicts the number of each peptide detected across replicates. Peptides with ambiguous assignments are shown in both potential candidate isoforms.

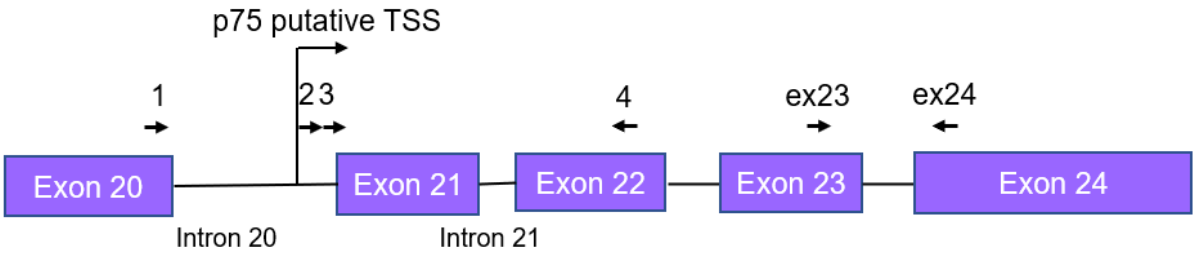
together, we conclude that there is no proteomic evidence of the p75 isoform in a representative AML cell line that possessed the p75 protein band.

3.2.4 No p75 CUX1 transcript is detected at the RNA level in human AML and breast cancer cell lines

It remained possible that the p75 protein is below the level of detection in the proteomics approach. As a more sensitive test, we looked for the presence of p75 CUX1 at the mRNA level,

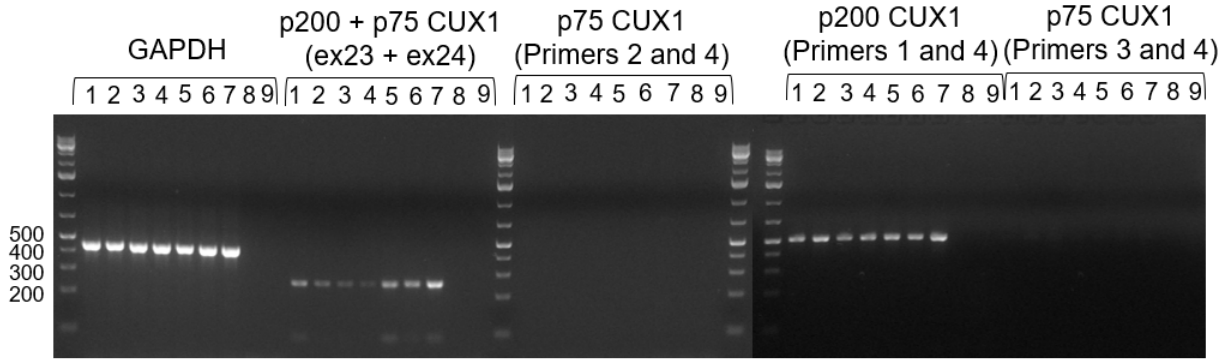
using two primer sets (primers 2+4 and 3+4) spanning the intron 20-exon 21 boundary (Figure 3.24). The qPCR primers (3+4) were previously used to provide proof for the p75 transcript at the mRNA level [141]. We assessed cDNA from three AML cell lines (K562, Kasumi-1, KG-1) and three breast cancer cell lines reported to express p75 (T47D, MDA-MB-231 and MCF-7) (Figure 3.25). Primers for GAPDH and total CUX1 (primers 1+4, ex23+24) served as positive controls. We did not detect any bands with the previously reported p75 primers (3+4) or with our upstream designed primers (2+4), even after 30 cycles of PCR amplification (Figure 3.25). The 3' primer from our primer set used to detect p75 expression worked independently with a different primer binding exon 20 and amplified p200 (primers 1+4), indicating that the lack of a transcript was not due to primer design issues. This suggests that the p75 transcript band reported previously could be due to the 5' primer being too close to the intron-exon border and amplifying pre-spliced intronic sequence. According to the previous description of the p75 TSS originating at least 2.5 kb upstream of the exon 21 border, our upstream primers do not seem to capture a transcript indicative of p75 in either human AML cell lines or the breast cancer cell lines previously reported to express p75 [141].

We also performed qPCR using SYBR Green with the same primers listed above. We did not see any amplification with the p75 primers until cycle 35, and the melt curves for this amplicon did not overlap across technical replicates in each cell type. The level of expression of this transcript was also $10^3 – 10^4$ lower than that of p200 CUX1, indicating that expression of p75 is much lower than what was described previously (Figure 3.26). Since we did not observe amplification under cycle 35, comparable to the water and no RT controls, we concluded that there was no evidence for a p75 transcript either.



Primers	Size (bp)	Unique to p75	Previously used to detect p75
ex23 + ex24	256 bp	No (all isoforms)	No
2+4	719 bp	Yes	No
1+4	510 bp	No (p200)	No
3+4	497 bp	Yes	Yes (PMID: 12438259)

Figure 3.24: (Top) Schematic of primers used to detect p75 in qPCR. (Bottom) Table of anticipated PCR products and whether it is unique to the p75 isoform



cDNA Sample Order:

1) K562
2) Kasumi-1
3) KG-1
4) KG-1 Δp75 #21
5) T47D
6) MDA-MB-231
7) MCF-7
8) No template
9) No RT control

Figure 3.25: Reverse-transcriptase PCR products after 30 cycles of PCR spanning the intron 20 region of CUX1 using cDNA reverse-transcribed from mRNA in 6 different cell lines (K562, Kasumi-1, KG-1, T47D, MDA-MB-231 and MCF-7) (n=3). cDNA from the KG-1 Δp75 #21 cell line was used as a negative control for p75 mRNA expression. GAPDH, p200-specific primers (1

Figure 3.25 (Continued)
and 4) and CUX1 primers spanning all isoforms (ex23 and 24) were used as positive controls. Gel is cropped for better presentation.

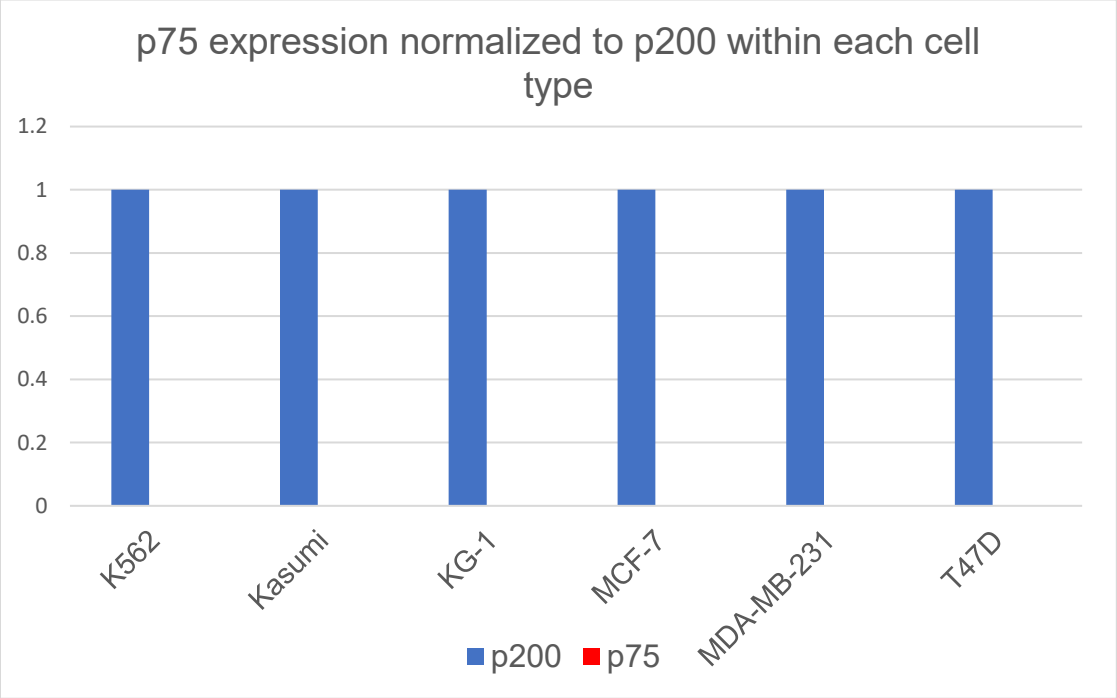


Figure 3.26: qPCR results showing mRNA expression for p75 CUX1 normalized to that of p200 CUX1 within each of the cell lines displayed above.

3.2.5 Functional genomics consortia datasets lack epigenetic or transcriptional evidence for a CUX1 p75 intronic transcriptional start site

Heretofore, our analysis has encompassed thirteen cell types. We sought to extend our analysis to comprehensively assess additional normal and malignant tissue types. As such, we leveraged consortia-generated functional genomics datasets for evidence of p75 CUX1 across a variety of cell types. Genome-wide studies of promoters using the H3K4me mark, an epigenetic mark of active promoters, or CAGE-tag data, which maps TSSs, have shown that TSSs are

differentially used in cancer [187]. We first assessed epigenetic marks that canonically decorate promoters. Promoters have high levels of H3K4me3 and H3K27ac deposition, and low H3K4me1 deposition [187]. As expected, we observed robust promoter marks at the p200 TSS across seven Tier 1 ENCODE cell lines (Figure 3.27). However, we did not observe any H3K4me3 peaks in the intron 20 region of CUX1, including in MCF7 (Figure 3.27). We observed some H3K27ac peaks in intron 20, but these peaks were not conserved across the ENCODE Tier 1 cell lines (Figure 3.27). Additionally, we looked at the ChromHMM track for MCF-7. ChromHMM is a software that can integrate multiple chromatin datasets, like ChIP-seq data of various histone modifications, to characterize chromatin states by recurring combinatorial and spatial patterns of these marks [188,189]. ChromHMM is generated using a multivariate hidden Markov model that considers the presence or absence of each chromatin mark, to generate a biological characterization of each section of the genome. The MCF7 ChromHMM track shows an active TSS (red mark) at the exon 1a and 1b region, but only weak transcription (dark green) and no promoter regions in the intron 20 region (MCF7 ChromHMM track, Figure 3.27). Thus, epigenetic features of promoters are not present at the putative p75 TSS, even in a cell type documented to express p75.

Chromatin accessibility is functionally indicative of promoters and enhancers, as these regions are open to be bound by transcription factors compared to closed chromatin [190,191]. To assess chromatin accessibility at putative transcriptional start sites, we looked at DNase hypersensitivity tracks and transcription factor ChIP-seq tracks at exon 1 and intron 20 of *CUX1*. We observe increased DNase accessibility and transcription factor binding at exon 1. However,

there does not seem to be equivalent chromatin accessibility at intron 20. There is some transcription factor binding signal observed at intron 20, although not as strong as at exon 1.

CpG islands are relatively rare and tend to be associated with promoters in vertebrates [33]. Evolutionarily, a cytosine base (C) tends to get methylated and converted to a thymidine (T) because of spontaneous deamination, so the conservation of CpG islands tends to be indicative of a selective pressure to retain that region, perhaps having to do with regulation of gene expression [33]. We observe a strong CPG island at the p200 TSS, but none within the reported p75 TSS (Figure 3.27).

We also looked for signs of a promoter within the intron 20 region by using the EPDnew promoters track. This track was assembled by using gene transcript coordinates from multiple sources (HGNC, GENCODE, Ensembl, RefSeq) and validated using data from CAGE and RAMPAGE experimental studies from FANTOM5, UCSC and ENCODE, and is a rich resource for predicting putative TSSs [192]. Again, we observed no evidence for an alternative TSS within intron 20 of *CUX1*.

Given that the epigenetic marks do not rule out the possibility of a cis-regulatory element in intron 20, we next sought to identify evidence of a transcript arising from intron 20. We assessed publicly available next-generation sequencing-based cap-analysis gene expression (CAGE-seq) data that captures 5' capped mRNA transcripts and maps transcription start sites (TSSs) in promoter regions [9,34,193]. CAGE-seq data shows a conserved 5' capped mRNA in 5 different ENCODE cell lines (GM78, H1ES, K562, HepG2 and MCF7) at exon 1 of CUX1. However, we did not observe any conserved CAGE tags in intron 20 of CUX1 that would be

indicative of a novel transcript arising from this region, including from MCF7, which is reported to express p75 (Figure 3.27).

Another approach to identify TSSs is by using POL2RA ChIP-seq. While POL2RA peaks are detected at the p200 TSS site, neither K562 nor MCF7 cells harbor POL2RA peaks anywhere within intron 20. These data are inconsistent with an intron 20 TSS.

We also utilized data from the FANTOM5 project to look for mapped transcription start sites on CUX1 using CAGE data across 975 human samples [9]. 5' CAGE TSS collation was performed on primary mapping BAM files, and peaks were observed in regions that had 3 or more CAGE tags. We were able to observe a prominent CAGE peak for exon 1 of CUX1, but no significant peaks within the intron 20 region of CUX1 indicative of a TSS (Figure 3.28).

Further evidence for the lack of a transcriptional start site in intron 20 of CUX1 comes from the lack of RNA-seq reads mapping to the putative transcriptional start site region in intron 20. We leveraged RNA-seq data available on the UCSC Genome Browser collected from human breast tissue and processed by the Burge lab [179]. We did not observe any sequencing reads mapping to the intron 20 region in this tissue type, that would indicate the existence of an alternative transcript arising from this region (Figure 3.29) In the MCF7 and T-47D cell lines previously reported to express p75, we observed some sequencing reads within the intron 20 region, but these were not conserved across three replicates, and were not contiguous with the exon 21 reads (Figure 3.29).

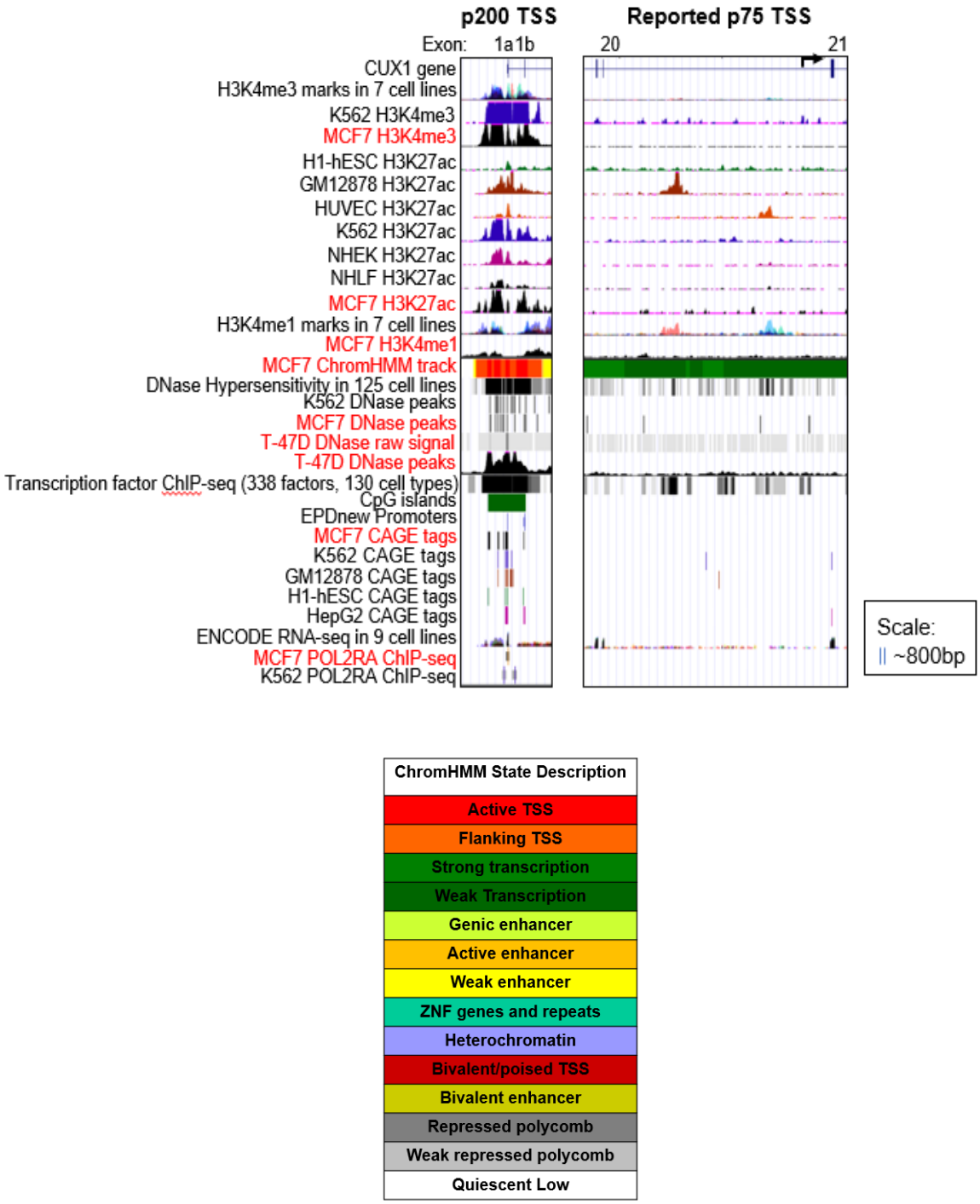


Figure 3.27: (Top) CUX1 Refseq gene model is aligned with tracks of indicated transcriptional and epigenetic marks (including H3K4me1, H3K4me3 and H3K27ac), CAGE-seq data in 5 ENCODE cell lines, DNase hypersensitivity tracks, POL2RA ChIP-seq tracks and transcription

Figure 3.27 (Continued)
factor ChIP-seq tracks). Tracks highlighted in red are derived from a cell type previously reported to express p75 CUX1. (Bottom) Chromatin states assigned by the ChromHMM model, and color-coded to signify different chromatin states

For a completely different approach, we turned to CUX1 isoform expression using exon-level expression calculated based on a collapsed gene model. Only two CUX1 isoforms are calculated to be expressed in most tissue types (ENST00000360264.7 and ENST00000292535.11), and these are both full-length transcripts of CUX1 (Figure 3.30). There is no evidence for short CUX1 isoforms using this browser data (Figure 3.30). Finally, we note that no gene assembly or gene prediction database for coding or non-coding RNA annotate a p75 isoform. NCBI Refseq, CCDS, GENCODE, Ensembl gene predictions, AUGUSTUS, or ORFeome all lack a p75 transcript (Figure 3.30) [194–198]. Collectively, these extensive datasets spanning thousands of samples provide no evidence of a p75 CUX1 transcript.

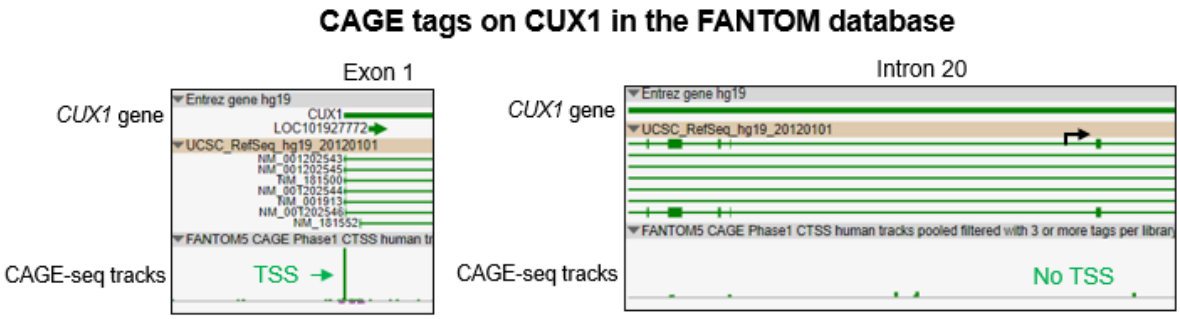


Figure 3.28: CAGE-seq data from the FANTOM browser at exon1 and intron 20 of CUX1 in 975 human primary cells and cancer cell lines.

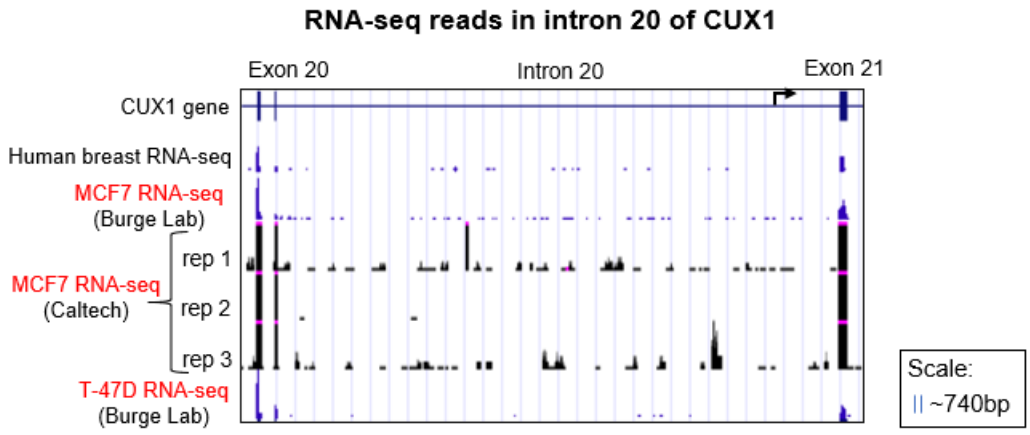


Figure 3.29: RNA-seq peaks for CUX1 expression in human breast tissue, MCF7 and T-47D cell lines from the Burge lab sequencing data [179], visualized on the UCSC Genome Browser. No intronic reads contiguous with exon 21 are observed.

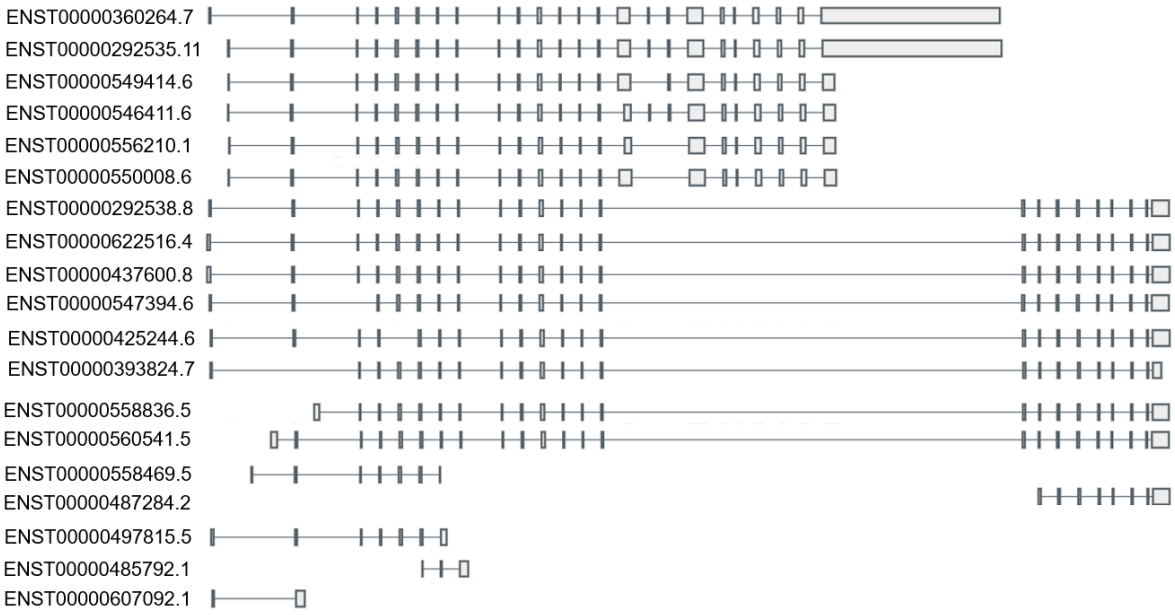


Figure 3.30: CUX1 isoforms on the GTEx browser

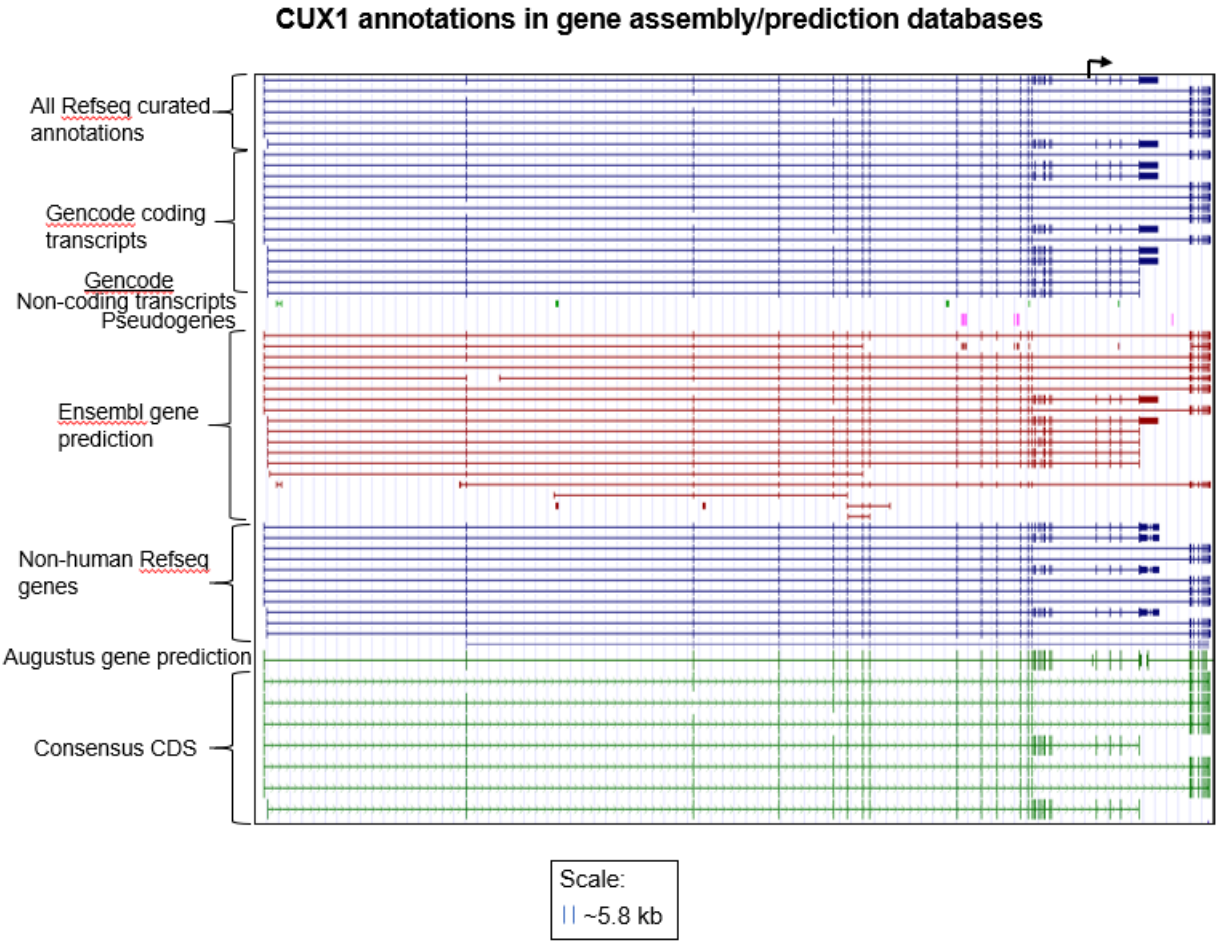


Figure 3.31: All CUX1 isoforms observed across several gene assembly and gene prediction databases (NCBI Refseq, Gencode, Ensembl, Augustus, CCDS) reveal no transcript resembling the p75 isoform.

Collectively, this data shows that there is no evidence of promoter or enhancer marks indicative of a separate p75 CUX1 transcript, no CAGE tracks mapping to an mRNA arising from the intron 20 region, and no evidence of a transcriptional start site mapping to this region based on ENCODE and GTEx data.

3.3 DISCUSSION

Next-generation sequencing and CRISPR/Cas9 genome editing has revolutionized biomedical research in many ways. Perhaps less appreciated, however, is the role of these technologies in establishing the legitimacy of research findings. CRISPR/Cas9 editing, for instance, has invalidated cancer dependencies, drug targets, and viral receptors, as some illustrative examples [199–201]. In this report, we employ the power of functional genomics and CRISPR editing to demonstrate that the *CUX1* gene does not encode a p75 isoform as described [141]. This conclusion is buttressed via multiple orthogonal approaches including biochemical studies with several antibodies, proteomics and extensive mining of functional genomic datasets across a plethora of cell types. Taken together, our data is inconsistent with a p75 CUX1 isoform arising from an intronic TSS and suggests that prior reports were based on a western blotting artefact [141].

Other studies support our conclusion. p75 was first identified in HeLa cells, HEK293 cells, breast cancer cell lines and mouse thymus [141]. The original manuscript identifying human CUX1 generated antiserum against the entire CUX1 protein, yet western blotting of HeLa cells identified only p200 [202]. Probing HEK293 cells with an antibody against the C-terminus region of CUX1 shows no p75 in another study [142]. In a report that p75 CUX1 causes polycystic kidney disease, the endogenous p75 expression was never documented at the protein level; all subsequent experiments were performed by over-expressing p75 cDNA [203]. Probing a western blot of entire *Drosophila* embryos for the highly conserved CUX1 ortholog, Cut, only reveals the full-length protein [41]. While p200 CUX1 protein increases after TGF-β treatment in normal lung fibroblasts, p75 does not [204]. Finally, in a study that reported that androgen-resistant prostate

cancer cell lines upregulate p200, p75 was unchanged [89]. Collectively, these studies either fail to document an endogenous p75 protein, or uncouple the biology of p200 from p75.

It is unclear what the p75 cDNA product previously reported represents. Perhaps it is a result of recursive splicing, where long introns are spliced in a sequential manner in tissue-specific contexts leading to a lag in splicing of intron 20 material [205]. Alternatively, studies of nascent transcription indicate that splicing does not always occur in the order of transcription, and introns that are spliced later temporally tend to be longer and have higher RNA-binding protein occupancy [206]. In keeping with this notion, intron 20 is the sixth longest intron in *CUX1*, and has an elevated RBP occupancy compared to sixteen out of twenty-three introns in the gene, and thus may be spliced later than other introns in *CUX1* (Figure 3.32) [206]. It is conceivable that the p75 cDNA product previously observed represents an intermediate, incompletely spliced p200 mRNA.

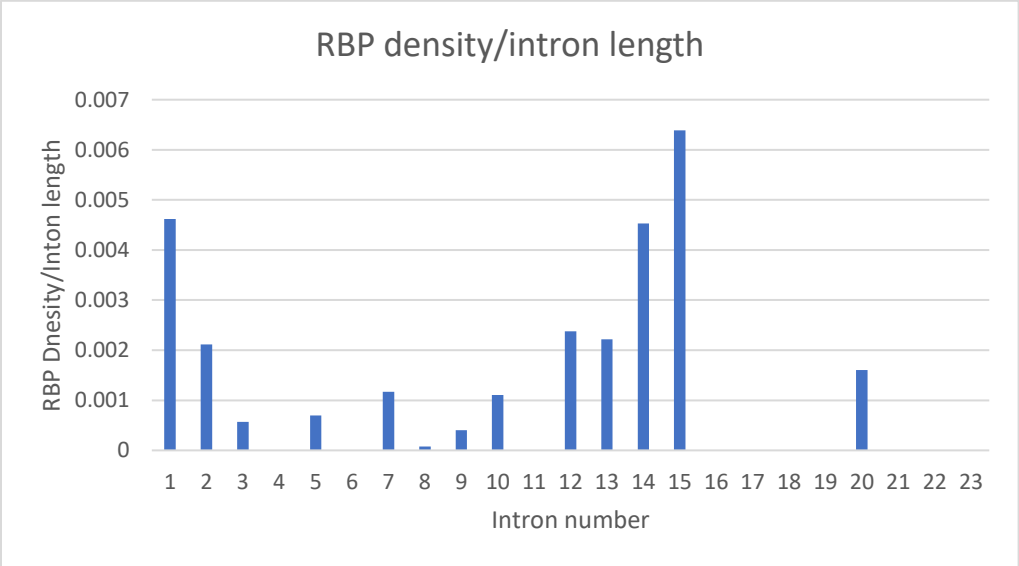


Figure 3.32: RNA-binding protein occupancy calculated for each intron in the CUX1 gene, using data from a recent paper studying splicing kinetics [206].

In our analysis of hematopoietic cells, we only document the expression of the p200 CUX1 isoform. We cannot comment on the validity of p110 or other short isoforms (p80, p90 and p150). As these isoforms are generated post-translationally, functional genomics datasets are ineffective in determining their legitimacy. Future studies of these putative isoforms in other tissue types should employ stringent techniques such as CRISPR/Cas9 mutagenesis to ensure against being misled by western blotting artefacts.

Our finding calls into question studies that ascribed oncogenic functions to p75 CUX1. Many of these publications did not study endogenous p75, but instead employed overexpression models, which can confound results [90,141–143,207]. We speculate that overexpression of CUX1 has dominant negative effects. For instance, in mice, both overexpression of p75 and knockout or knockdown of CUX1 leads to myeloproliferative disease [63,67,142]. One interpretation of these seemingly incongruent findings is that artificial overexpression of p75 either interferes with the stoichiometry of endogenous CUX1 protein complexes or blocks full-length CUX1 from binding to its target genes. Indeed, p75 CUX1 has increased DNA-binding affinity compared to endogenous p200 CUX1 [49]. The net effect of CUX1 overexpression may be the disruption of endogenous CUX1 tumor suppressor activity. In support of this model, the p150 CUX1 isoform was found to exert a dominant negative phenotype upon p200 CUX1 [154]. In this light, the use of overexpression systems to characterize p110 may also misattribute this isoform with oncogenic properties that are not valid [90,143,147,148,151].

There are relatively fewer reports that p200 CUX1 is oncogenic. p200 CUX1 has been shown to causes promote cell line migration, invasion, and evasion of apoptosis, and p200 transgenic mice develop organ hyperplasia [71,79,165]. 7q copy number gains and CUX1

overexpression has been documented in primary cancers [182,208,209]. However, these findings should be interpreted with caution. Chromosome 7 also encodes oncogenes, including *EGFR* (on 7p), and *BRAF*, *CDK6,* and *EZH2* (on 7q) [210–212]. Thus, in cancers with chromosome 7 copy number gains, the driver may be a true oncogene, while *CUX1* is a passenger. Indeed, rigorous pan-cancer gene-level analysis of copy number alterations and mutation patterns in primary patient samples reveal *CUX1* genetic changes significantly characteristic of a tumor suppressor gene [129,132,184]. There is now a growing body of work that CUX1 is tumor suppressive [63,112,129,130,136,140].

Given the importance of CUX1 in development and disease across a wide variety of tissue types, it is critical to carefully dissect and understand the genomic structure of the *CUX1* locus and encoded protein. The complexity of the gene has led to confusion in the field resulting in serious inaccuracies, most recently referenced [213,214]. We expect that our current study will help rectify these obstacles going forward.

Table 3.1: List of references that show no evidence for p75 CUX1

Citation	Evidence against p75 CUX1
Neufeld, E. J., Skalnik, D. G., Lievens, P. M., Orkin, S. H., & Riley, W. (1992), Nature Genetics	No short CUX1 isoforms identified in HeLa cells when nuclear extract was affinity-purified using a CCAAT site-containing oligonucleotide; anti-sera raised in guinea pig against purified CUX1 (CDP) reveals only one band, and a shorter non-specific band also observed in the absence of the primary antibody. This contradicts the original p75 report that identified a conserved p75 band in RNase mapping experiments in HeLa cells.
Cadieux, C., Fournier, S., Peterson, A. C., Bédard, C., Bedell, B. J., & Nepveu, A. (2006), Cancer Research	No short CUX1 isoforms were observed in the parental HEK293 cell line used to overexpress short CUX1 isoforms. This contradicts the original p75 report that identified a p75 band in RNase mapping experiments in HEK293 cells.
Cadieux, C., Harada, R., Paquet, M., Côté, O., Trudel, M., Nepveu, A., & Bouchard, M. (2008), Journal of Biological Chemistry	No verification of p75 CUX1 expression at the protein level
Ikeda, T., Fragiadaki, M., Shi-wen, X., Ponticos, M., Khan, K., Denton, C., … Abraham, D. (2016). Biochemistry and Biophysics Reports	Short Cux1 isoforms are not induced upon TGF-β treatment in normal lung fibroblasts, and the only Cux1 isoform whose expression is increased is p200
Li, H., Yang, F., Hu, A., Wang, X., Fang, E., Chen, Y., … Zheng, L. (2019). EMBO Molecular Medicine	Hardly any p75 CUX1 detected using rt-PCR with primers from Goulet et al, 2002, in fetal adrenal medulla and neuroblastoma cell lines
Dorris E. R., O'Neill A., Treacy A.,…Watson R. William. (2020). Oncotarget	No significant difference in gene expression levels of CUX1 using a primer set that detects only p200 CUX1, and a primer set that detects both p200 and p75 CUX1 in prostate cancer cell lines

CHAPTER 4

DISCUSSION AND FUTURE DIRECTIONS

-7/del(7q) cytogenetic abnormalities are associated with several hematopoietic diseases, such as MDS, MPN, AML and t-MN, and often co-occur with -5/del(5q) and del(17p) abnormalities [215]. Candidate tumor suppressor genes have been identified on 7q, one of which is CUX1. There is a lot of controversy surrounding the role of CUX1 in tumor progression though, with several groups reporting the involvement of CUX1 in tumor progression, and other groups firmly cementing the role of CUX1 as a tumor suppressor gene. One possible explanation to reconcile the seemingly opposing functions of CUX1 has been that short CUX1 isoforms are more tumorigenic, while the full-length p200 CUX1 protein is more tumor suppressive. In this study, we decided to interrogate the presence of short CUX1 isoforms in hematopoietic cell lines, and to further assess the role of the short p75 CUX1 isoform to determine if it was indeed oncogenic and acted in opposition of the p200 CUX1 protein.

In this study, we examined the isoforms of CUX1 expressed in human AML cell lines initially, and eventually expanded to looking at CUX1 isoform expression in human breast cancer cell lines previously reported to express p75. We also leveraged publicly available functional genomic datasets to look at a wide range of cell lines for epigenetic or transcriptional evidence of an alternative transcript isoform. Through several such orthogonal approaches, we found no evidence for a p75 CUX1 isoform. This is surprising as p75 is widely referenced in the literature on CUX1. In this study, we show using several orthogonal approaches that p75 CUX1 is a western blotting artefact and is not an isoform transcribed from an alternative TSS within the CUX1 gene.

Combing through the CUX1 literature, we found several studies that indirectly support our claim of p75 not being real. The first report describing p75 originally identified it using RNase mapping in HeLa cells, HEK293 cells, several human breast cancer cell lines and in the mouse thymus [141]. It was also described to be expressed in double positive and single positive mouse thymocytes [141]. Interestingly, the original manuscript identifying human CUX1 generated CUX1 antiserum, and used this to blot for CUX1 protein in HeLa cells. They did not observe any p75 using this approach, and the only CUX1 band observed was p200 [202]. Moreover, the paper that demonstrated p75 transgenic mice develop a myeloid neoplasm-like disease did not observe any short isoforms in the parental HEK293 cells used to overexpress short CUX1 isoforms, using an antibody against the C-terminal region of CUX1 [142]. In a report implicating p75 CUX1 in polycystic kidney disease, the expression of endogenous p75 was not observed at the protein level, and all downstream experiments were performed by over-expressing p75 CUX1 cDNA in the kidneys of transgenic mice [203]. Antiserum generated against the entire CUX1 protein only detects p200 CUX1 in the mouse thymus in another study [67]. Blotting for CUX1 in Drosophila embryo only reveals the full-length protein, and since CUX1 is highly conserved across different organisms, this might be another piece of evidence against the existence of p75 [41]. Short CUX1 isoforms are also not induced upon TGF-β treatment in normal lung fibroblasts, and the only CUX1 isoform whose expression is induced upon treatment is p200 [204] Finally, in a recent study by Dorris et al in androgen-resistant prostate cancer cell lines, no significant difference in gene expression levels of CUX1 was observed in qRT-PCR using a primer set that only detects p200, and a primer set that detects both p200 and p75 CUX1, supporting the argument that this isoform is not detectable by other groups as well [89]. This further bolsters our claim that p75 CUX1 has not been systematically validated at an endogenous level.

There have been a few other studies published after the original paper describing p75 that have been able to show evidence for the alternative transcript using rt-PCR, using the same primers from the original paper [89,129,216]. Two of these studies report p75 as being very lowly expressed in comparison to p200, so it is doubtful how functional it is [89,216]. This correlates with our qPCR data, where p75 was amplified around cycle number 34, comparable to the no template control. We also devised our own primers further upstream of the intron 20-exon 21 boundary to measure the expression level of p75, and we were unable to see any amplification with this new primer set. Since the original p75 primer set had the 5' primer located just 40 nucleotides upstream of the intron 20-exon 21 boundary, it is possible that previous studies that looked at endogenous expression of this alternative transcript could have been capturing an incompletely spliced product. Indeed, there is evidence that intron 20 may be spliced later than other introns of the gene, from studies that correlate RNA-binding protein (RBP) occupancy with splicing kinetics [206]. Intron 20 possesses an elevated RBP occupancy normalized to length of intron compared to almost 75% of all introns in CUX1, and thus may be spliced later temporally during mRNA processing. Another explanation could be that the intron 20 has an alternative 3' splice site, which would explain why part of the intron is retained with exon 21. This supports the fact that we see a faint cDNA product with Nepveu's previously described primers, but no product using primers we designed located further upstream of the intron 20-exon 21 boundary. Alternatively, we could be seeing a hereby undescribed non-coding RNA or enhancer RNA being transcribed from this region, although the likelihood of this is low, given the extremely low expression of this transcript.

We do not find any evidence for a protein indicative of p75 though, using a CUX1-tagged cell line and mouse model, selective CRISPR/Cas9 editing of the p75 TSS and editing of total CUX1, CUX1 immunoprecipitation, or proteomics on CUX1. We do not find any peptides mapping to the CUX1 sequence when CUX1 is immunoprecipitated, and in the whole cell extract sample, the peptides that map to CUX1 also map to the common sequence shared with CASP, and we also detect peptides that are unique to CASP and not shared with CUX1. Since CASP is 77kDa, and p75 CUX1 is 75 kDa, part of the confusion in the literature could be attributed to CASP being mistaken for a CUX1 isoform. Indeed, several studies looking at p75 expression use the wrong antibody to detect it to this day, and mistakenly attribute CASP as p75 CUX1 [145,213,214]. Our study will be crucial to set the record straight about p75, and avoid such egregious mistakes from being disseminated.

There is evidence for activation of oncogenes through alternative transcript initiation in the literature. For example, ALK has been documented to encode a transcript from an alternative TSS embedded in intron 19, named ALK^{ATI} [217]. This alternative TSS has H3K4me3 deposition and RNA polymerase II binding, and has contiguous RNA-seq reads arising within intron 19 and spanning exon 20 [217]. Looking at all these metrics for CUX1 in publicly available functional genomics datasets, we were unable to find any evidence of an alternative TSS in intron 20 indicative of p75. The ALK transcript was also only identified in 11% of melanoma samples, and did not seem to be expressed in other cancer types or normal tissues, indicating that such transcripts from alternative TSSs might be extremely tissue specific.

There is also evidence for different protein isoforms functioning in opposing ways in tumor progression. For example, RUNX1 isoforms have been shown to be differentially

expressed in samples from patients with MDS. RUNX1a, the shorter isoform lacking exon 7b and exon 8, is overexpressed in CD34+ HSPCs, compared to RUNX1b/c, and is known to exert a dominant negative effect on the full-length RUNX1b protein [218,219]. This differential expression is mediated by a mutation in the splicing factor SRSF2, leading to aberrant splicing of RUNX1 to generate a disproportionate ratio of RUNX1a to RUNX1b/c [219]. Similarly, a different study on SON isoforms reported that alternatively spliced short SON isoforms are strongly upregulated in AML and disrupt the function of the full-length SON protein by competing for occupancy on chromatin and thereby repressing the transcriptional repressor activity of full-length SON [220]. Based on all this evidence, we thought it plausible that short CUX1 isoforms could have a very distinct cellular and molecular role compared to full-length CUX1. But through validation using several orthogonal approaches, we found no evidence for a p75 CUX1 isoform widely referenced in the literature.

There are a few studies that describe oncogenic effects upon knockdown of p200 CUX1. In NIH3T3 fibroblasts, knockdown of p200 CUX1 using RNA interference (RNAi) led to decreased migration through a membrane, and decreased invasion through a three-dimensional matrix [71]. This result was also recapitulated in cancer cell lines of different tissue origins, such as fibrosarcoma, glioblastoma, and breast cancer [71]. CUX1 protein expression was associated with higher grade breast cancers [71]. Multiple studies have shown that p200 knockdown enhances TRAIL- and drug-induced apoptosis in pancreatic cancer cell lines [79]. All these studies were performed in solid tumor types, such as breast cancer, pancreatic cancer, and melanoma. We believe that this could be due to CUX1 having specific effects in different cell types. There are

several studies now that firmly cement p200 CUX1 as a tumor suppressor in hematopoietic systems [63,112,129,130,136,140].

These cell-type specific discrepancies of CUX1 function can also be extended more broadly to symbolize the functional consequences of copy number alterations of chromosome 7 in different cancer types. -7/del(7q) is commonly seen in patients with myeloid malignancies or dysplasia, whereas chromosome 7 tends to be amplified in several other cancer types such as colorectal cancer, non-small cell lung cancer, breast cancer and bladder cancer [221]. Chromosome 7 encodes several other well-described oncogenes, such as BRAF, CDK6 and EGFR, and is thus commonly amplified in many solid cancer types [210–212]. This could also be an explanation for why CUX1 over-expression is correlated with poor prognosis in several solid cancers, as this tends to be a passenger gene amplification with copy number gain in these cancer types.

It is rather intriguing that transgenic p75 mice develop a myeloproliferative-like myeloid neoplasm, similar to the MDS/MPN phenotype we observe in Cux1-low mice generated in our lab [63,67,142]. One conceivable explanation for this could be that short isoforms like p75 disrupt binding of full-length CUX1 to its genomic targets, and thereby disrupt its transcriptional activity and tumor suppressor activity in a dominant negative fashion. There is also evidence for short isoforms exerting a dominant negative effect on p200 CUX1, such as a study of p150 CUX1 [154]. Another line of evidence from this theory comes from the study of the binding strength and kinetics of different CUX1 constructs that revealed that p75 is capable of binding more stably to DNA than p200 CUX1, suggesting that p75 may compete for binding sites with p200 [49]. It is also interesting that cellular and phenotypic effects of overexpressing p200 CUX1 are modest, and are often exacerbated by concurrent overexpression of p110 or p75 [151].

Further work is needed to characterize the several other short CUX1 isoforms that are described in the literature. There have been reports of a p80, p90 and p150 CUX1 [52,146,154,155]. Unlike the p75 isoform, all these other short isoforms are generated by post-transcriptional proteolytic processing of the full-length protein by various caspases. It is unclear how these isoforms function in relation to p200. Careful characterization of these isoforms using CRISPR/CAS9 editing of the proteolytic processing site, similar to what we described for selectively editing p75, or caspase knockout models, is needed to study the validity and function of these isoforms better. In addition, p110 can be more rigorously queried by using cathespin inhibitors, such as E-64d, to verify if there are any differences in bands seen, although we did not observe any band around 110 kDa when blotting the AML and breast cancer cell lines, or in the CUX1-tagged cell line or mouse model that we generated.

Further work is also needed to understand if the p75 isoform can indeed exert a dominant negative effect on p200 CUX1, to better understand the discrepant reports on its role in tumor progression. This can be queried by transfecting cells expressing p200 CUX1 with exogenous p75, and looking at CUX1 binding sites on the genome, using ChIP-seq or CUT and RUN. We would have to use a ChIP-grade CUX1 antibody that binds the N-terminal region of CUX1, in order to only detect p200 binding sites. We would expect to see fewer CUX1 binding sites in the condition where exogenous p75 is transfected, compared to the empty vector control.

CRISPR/Cas9 is such a powerful tool in this age to answer several of these questions, and can be utilized to understand the complex nature of the CUX1 gene better. For example, it is still unclear what exact role the auto-inhibitory domain or the repressive domains of CUX1 play in its transcriptional activity. Precise editing of these domains using CRISPR/Cas9, followed by the

generation and integration of ChIP-seq and RNA-seq datasets will allow us to parse out the effects of these regions on transcriptional function. In parallel, a domain screen targeting the different domains of CUX1 will also elucidate which domains are essential to the main function of the protein [222].

Now that our work has firmly established p200 CUX1 is the dominant isoform in hematopoietic cells, and previous work has firmly established p200 as a tumor suppressor gene, it would be interesting for future work to focus on the mechanism by which p200 CUX1 functions as a tumor suppressor. RNA-seq from CUX1-knockdown human CD34+ HSPCs revealed that *CUX1* transcriptionally activates *PIK3IP1*, a negative regulator of the PI3K/Akt pathway [63,129]. This could be one mechanism by which it prevents oncogenesis. Additionally, recent work from our lab has implicated CUX1 as playing a role in DNA damage recognition, and potentiating DNA repair by recruiting factors that allow double strand break (DSB) recognition [140]. Knockdown of CUX1 disrupts the global histone post-translational modification (PTM) landscape, and renders cells unable to response to stressors such as DNA damage [140]. CUX1 loss seems to reduce the co-localization of EHMT2, a histone methyltransferase that stabilizes phospho-ATM and is necessary for downstream DNA repair factor recruitment, with γH2AX foci. CUX1 has been previously shown to recruit EHMT2 to DNA to mediate transcriptional repression, but this interaction has not been rigorously studied in hematopoietic tissue [223]. Future work could characterize whether the interaction between CUX1 and EHMT2 is transcriptional, epigenetic or some type of protein-protein interaction. CUX1 may be recruiting EHMT2 directly to the chromatin, but could also be turning on a transcriptional program that somehow affects EHMT2 expression, and subsequent chromatin recruitment.

There is evidence that CUX1 is regulated by the cell cycle, and loss of CUX1 seems to potentiate increased cell cycle progression. It is possible that CUX1 may exert its tumor suppressor activity by regulating important cell cycle checkpoints. Future work would interrogate more comprehensively which cell cycle proteins CUX1 interacts with in a coIP-MS. Some of these candidate interactions could be disrupted by mutating the site of protein-protein interactions on either protein, then the effect of this perturbation could be studied to look at the effect it has on cell cycle progression.

Future directions

1) More rigorous work into all short CUX1 isoforms reported, and whether they have tumor-promoting or tumor-suppressive effects in hematopoiesis is needed. This could be achieved by either editing the cathepsin cleavage sites on CUX1 or using a cathepsin inhibitor to stop expression of the proteolytically processed CUX1 protein isoforms, and the subsequent functional consequence of the loss of these isoforms could be studied.
2) Which domains of CUX1 are crucial for its tumor suppressor function? A domain screen could be used to understand the functions of all the domains of CUX1 better, including the poorly understood auto-inhibitory domain and the repressive domains.
3) Does the p80 isoform have dramatically different transcriptional activity compared to the full-length protein, given that it is missing the two active repressive domains in the C-terminal region?
4) The functional mechanism by which p200 is tumor-suppressive could also be better elucidated. This could be transcriptional, such as by negatively regulating oncogenic signaling pathways (such as PI3K/Akt signaling), or epigenetic, such as by recruiting

proteins that in turn play a tumor suppressive function, or by modifying the histone modification landscape to mediate activation of tumor suppressor genes. What functions of CUX1 (cell cycle control, proliferation control, DNA damage recognition) are most pertinent for its tumor suppressive activity?

5) How does CUX1 regulate cell cycle progression? Does CUX1 have different transcriptional programs through different stages of the cell cycle? Or does CUX1 bind different proteins at different stages of the cell cycle to mediate cell cycle progression? For the first hypothesis, this could be tested by sorting out cells in each stage of the cell cycle using a cell-permeable DNA dye to delineate cell cycle phase, and performing ChIP-seq or CUT and RUN to determine CUX1 binding sites in each phase of the cell cycle. The second hypothesis can be tested by performing a co-IP-MS as mentioned above to determine cell cycle proteins that CUX1 interacts with selectively in each stage of the cell cycle.
6) Conversely, the cell cycle has also shown to exert a regulatory role in CUX1 expression and DNA-binding ability. Does this also affect function of the protein? How is this regulation mediated? There is evidence that the landscape of PTMs is modulated on CUX1 through the cell cycle. This could be more comprehensively characterized using proteomics, and identifying the enzymes responsible for these modifications.
7) Can the conflicting roles of CUX1 in different tissue types be reconciled by CUX1 binding to different proteins in different tissues? This could be answered by performing a co-IP-mass spectrometry in a tissue type where it has been described to be oncogenic (eg: breast cancer) and a tissue type where it has been described to be a tumor suppressor (eg: AML), and comparing the protein interactome of CUX1 in both tissues.

APPENDIX

Table A.1: List of peptides from the p200 band mapping to CUX1/CASP sequence in the CUX1 immunoprecipitated sample

Sequence	Probability	P-Score	Observed Mass	Actual Mass	Retention Time (minutes)	Start	Stop
(R)ELDATATVLANR(Q)	81%	69.276	637.341	1,272.67	44.73	12	23
(R)ELDATATVLANR(Q)	61%	54.608	637.341	1,272.67	44.57	12	23
(R)LIDVPDPVPALDLGQQLQLK(V)	92%	83.047	1,086.62	2,171.22	72.99	85	104
(R)LHDIETENQK(L)	95%	95.099	613.8042	1,225.59	23.41	108	117
(R)LHDIETENQK(L)	88%	91.087	409.5386	1,225.59	23.38	108	117
(R)LHDIETENQK(L)	43%	54.066	409.5386	1,225.59	23.39	108	117
(R)LHDIETENQKLR(E)	61%	63.408	499.267	1,494.78	28.32	108	119
(K)IREYEQTLK(N)	82%	79.986	393.8838	1,178.63	29.47	146	154
(K)NQAETIALEK(E)	96%	96.331	558.7984	1,115.58	36.82	155	164
(K)NQAETIALEKEQK(L)	79%	76.073	501.2667	1,500.78	31.47	155	167
(K)LQNDFAEK(E)	97%	109.86	482.7404	963.4662	31.07	168	175
(K)LQNDFAEKER(K)	80%	77.379	417.2106	1,248.61	28.4	168	177
(K)VQSLQTALEK(T)	91%	80.688	558.8166	1,115.62	41.01	197	206
(R)EQLSSANHSLQLASQIQK(A)	90%	94.01	661.3482	1,981.02	42.76	255	272
(R)SSLEVELAAK(E)	91%	80.455	523.79	1,045.57	44.74	286	295
(R)EIAQLVEDVQR(L)	84%	70.1	650.3488	1,298.68	50.81	298	308
(R)EIAQLVEDVQR(L)	75%	64.297	650.3488	1,298.68	50.92	298	308
(R)EIAQLVEDVQR(L)	74%	61.65	650.3488	1,298.68	50.95	298	308
(K)LKGQADYEEVKK(E)	75%	72.958	469.9208	1,406.74	23.98	345	356
(R)SLQSENAALR(I)	98%	114.89	544.7884	1,087.56	32.27	391	400
(R)SLQSENAALR(I)	97%	101.64	544.7884	1,087.56	32.14	391	400
(R)ISNSDLSGSAR(R)	95%	91.584	553.7755	1,105.54	26.66	401	411
(R)ISNSDLSGSAR(R)	69%	60.398	553.7755	1,105.54	26.78	401	411
(R)QRENPGQSLNR(L)	61%	63.408	433.5568	1,297.65	22.03	630	640
(R)SILQQAR(R)	90%	99.511	408.2403	814.4661	32.57	708	714
(R)NAASSEEAKAEETGGGK(E)	54%	59.862	545.9201	1,634.74	23.96	834	850
(R)VKEVLTDNNLGQR(L)	67%	66.708	495.9388	1,484.79	34.56	1135	1147
(R)VVLAPEEK(E)	93%	86.794	442.758	883.5015	33.23	1249	1256

Table A.2: List of peptides from the p75 band mapping to CUX1/CASP sequence in the CUX1 immunoprecipitated sample

Sequence	Probability	P-Score	Observed Mass	Actual Mass	Retention Time (minutes)	Start	Stop
(R)LHDIETENQK(L)	44%	53.683	409.5386	1,225.59	23.62	108	117

Table A.3: List of peptides from the p200 band mapping to CUX1/CASP sequence in the whole cell lysate sample

Sequence	Probability	P-Score	Observed Mass	Actual Mass	Retention Time (minutes)	Start	Stop
(R)ELDATATVLANR(Q)	51%	52.705	637.341	1,272.67	35.24	12	23
(K)APDVEQAIEVLTR(S)	77%	65.395	720.8883	1,439.76	60.43	273	285
(R)SSLEVELAAK(E)	67%	59.368	523.79	1045.57	35.27	286	295
(K)FALNSLLQR(Q)	94%	89.296	531.3087	1060.60	52.12	478	486
(R)RPSSLQSLFGLPEAAGAR(D)	61%	61.102	619.6708	1855.99	53.83	1452	1469

Table A.4: List of peptides from the p75 band mapping to CUX1/CASP sequence in the whole cell lysate sample

Sequence	Probability	P-Score	Observed Mass	Actual Mass	Retention Time (minutes)	Start	Stop
(K)APDVEQAIEVLTR(S)	96%	90.653	720.8883	1,439.76	67.15	238	250
(R)ITEAVATATEQR(E)	99%	138.65	645.3384	1,288.66	30.75	382	393
(R)ITEAVATATEQR(E)	83%	70.363	645.3384	1,288.66	21.59	382	393
(K)FLQSYPGR(G)	90%	79.116	484.2534	966.4923	35.85	507	514

REFERENCES

1. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).

2. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: The next generation. *Cell* **144**, 646–674 (2011).

3. Lander, E. S. *et al.* Erratum: Initial sequencing and analysis of the human genome: International Human Genome Sequencing Consortium (Nature (2001) 409 (860-921)). *Nature* **412**, 565–566 (2001).

4. Ayoubi, T. A. Y. & Van De Yen, W. J. M. Regulation of gene expression by alternative promoters. *FASEB J.* **10**, 453–460 (1996).

5. Wang, X., Hou, J., Quedenau, C. & Chen, W. Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol. Syst. Biol.* **12**, 875 (2016).

6. Cooper, S. J., Trinklein, N. D., Anton, E. D., Nguyen, L. & Myers, R. M. Comprehensive analysis of transcriptional promoter structure and function in 1% of the human genome. *Genome Res.* **16**, 1–10 (2006).

7. Kimura, K. *et al.* Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. *Genome Res.* **16**, 55–65 (2006).

8. Baek, D., Davis, C., Ewing, B., Gordon, D. & Green, P. Characterization and predictive discovery of evolutionarily conserved mammalian alternative promoters. *Genome Res.* **17**, 145–155 (2007).

9. Forrest, A. R. R. *et al.* A promoter-level mammalian expression atlas. *Nature* **507**, 462–470 (2014).

10. Xu, C., Park, J. K. & Zhang, J. Evidence that alternative transcriptional initiation is largely nonadaptive. *PLoS Biol.* **17**, 1–22 (2019).

11. de Klerk, E. & 't Hoen, P. A. C. Alternative mRNA transcription, processing, and translation: Insights from RNA sequencing. *Trends Genet.* **31**, 128–139 (2015).

12. Schibler, U., Hagenbüchle, O., Wellauer, P. K. & Pittet, A. C. Two promoters of different strengths control the transcription of the mouse alpha-amylase gene Amy-1a in the parotid gland and the liver. *Cell* **33**, 501–508 (1983).

13. Sobczak, K. & Krzyzosiak, W. J. Structural determinants of BRCA1 translational regulation. *J. Biol. Chem.* **277**, 17349–17358 (2002).

14. Arce, L., Yokoyama, N. N. & Waterman, M. L. Diversity of LEF/TCF action in development and disease. *Oncogene* **25**, 7492–7504 (2006).

15. Iynedjian, P. B. Mammalian glucokinase and its gene. *Biochem. J.* **293**, 1–13 (1993).

16. Struhl, K. Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* **14**, 103–105 (2007).

17. Pellagatti, A. *et al.* Impact of spliceosome mutations on RNA splicing in myelodysplasia: Dysregulated genes/pathways and clinical associations. *Blood* **132**, 1225–1240 (2018).

18. Lee, S. C. W. & Abdel-Wahab, O. Therapeutic targeting of splicing in cancer. *Nat. Med.* **22**, 976–986 (2016).

19. Schmitz, R., Ceribelli, M., Pittaluga, S., Wright, G. & Staudt, L. M. Oncogenic mechanisms in Burkitt lymphoma. *Cold Spring Harb. Perspect. Med.* **4**, 1–13 (2014).

20. Bardales, J. A., Wieser, E., Kawaji, H., Murakawa, Y. & Darzacq, X. Selective activation of alternative MYC core promoters by Wnt-responsive enhancers. *Genes (Basel).* **9**, (2018).

21. Hnilicová, J. *et al.* Histone deacetylase activity modulates alternative splicing. *PLoS One* **6**, (2011).

22. Zhou, H. L. *et al.* Hu proteins regulate alternative splicing by inducing localized histone hyperacetylation in an RNA-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* **108**, (2011).

23. Alló, M. *et al.* Control of alternative splicing through siRNA-mediated transcriptional gene silencing. *Nat. Struct. Mol. Biol.* **16**, 717–724 (2009).

24. Ameyar-Zazoua, M. *et al.* Argonaute proteins couple chromatin silencing to alternative splicing. *Nat. Struct. Mol. Biol.* **19**, 998–1005 (2012).

25. Pradhan, S. K. *et al.* EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. *Mol. Cell* **61**, 27–38 (2016).

26. Ahmad, K. & Henikoff, S. The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol. Cell* **9**, 1191–1200 (2002).

27. Raisner, R. M. *et al.* Histone variant H2A.Z Marks the 5′ ends of both active and inactive genes in euchromatin. *Cell* **123**, 233–248 (2005).

28. Barski, A. *et al.* High-Resolution Profiling of Histone Methylations in the Human Genome. *Cell* **129**, 823–837 (2007).

29. Ng, H. H., Robert, F., Young, R. A. & Struhl, K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell* **11**, 709–719 (2003).

30. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Mol. Cell* **25**, 15–30 (2007).

31. Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65–95 (2012).

32. Lauberth, S. M. *et al.* H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* **152**, 1021–1036 (2013).

33. Gardiner-Garden, M. & Frommer, M. CpG Islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).

34. Shiraki, T. *et al.* Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15776–15781 (2003).

35. Gu, W. *et al.* CapSeq and CIP-TAP identify pol ii start sites and reveal capped small RNAs as C. elegans piRNA precursors. *Cell* **151**, 1488–1500 (2012).

36. Ni, T. *et al.* A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods* **7**, 521–527 (2010).

37. Park-Sarge, O.-K. & Curry, T. E. *Molecular endocrinology : methods and protocols*. *Springer protocols* (2009).

38. Batut, P., Dobin, A., Plessy, C., Carninci, P. & Gingeras, T. R. High-fidelity promoter profiling reveals widespread alternative promoter usage and transposon-driven developmental gene expression. *Genome Res.* **23**, 169–180 (2013).

39. Core, L. J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.* **46**, 1311–1320 (2014).

40. Harada, R., Dufort, D., Denis-Larose, C. & Nepveu, A. Conserved Cut Repeats in the human cut homeodomain protein function as DNA binding domains. *J. Biol. Chem.* **269**, 2062–2067 (1994).

41. Blochlinger, K., Bodmer, R., Jan, L. Y. & Jan, Y. N. Patterns of expression of Cut, a protein required for external sensory organ development in wild-type and cut mutant Drosophila embryos. *Genes Dev.* **4**, 1322–1331 (1990).

42. Truscott, M. *et al.* The N-terminal Region of the CCAAT Displacement Protein (CDP)/ Cux Transcription Factor Functions as an Autoinhibitory Domain that Modulates DNA Binding *. **279**, 49787–49794 (2004).

43. Li, S. De *et al.* Transcriptional repression of the cystic fibrosis transmembrane conductance regulator gene, mediated by CCAAT displacement protein/cut homolog, is associated with histone deacetylation. *J. Biol. Chem.* **274**, 7803–7815 (1999).

44. Lievens, P. M. J., Tufarelli, C., Donady, J. J., Stagg, A. & Neufeld, E. J. CASP, a novel, highly conserved alternative-splicing product of the CDP/cut/cux gene, lacks cut-repeat and homeo DNA-binding domains, and interacts with full-length CDP in vitro. *Gene* **197**, 73–81 (1997).

45. Bürglin, T. R. & Cassata, G. Loss and gain of domains during evolution of cut superclass homeobox genes. *Int. J. Dev. Biol.* **46**, 115–123 (2002).

46. Malsam, J., Satoh, A., Pelletier, L. & Warren, G. Golgin tethers define subpopulations of COPI vesicles. *Science (80-.).* **307**, 1095–1098 (2005).

47. Gillingham, A. K., Pfeifer, A. C. & Munro, S. CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a golgi membrane protein related to giantin. *Mol. Biol. Cell* **13**, 4100–4109 (2002).

48. Lowe, M. The physiological functions of the Golgin vesicle tethering proteins. *Front. Cell Dev. Biol.* **7**, 1–13 (2019).

49. Moon, N. S., Bérubé, G. & Nepveu, A. CCAAT displacement activity involves CUT repeats 1 and 2, not the CUT homeodomain. *J. Biol. Chem.* **275**, 31325–31334 (2000).

50. Truscott, M. *et al.* CDP/Cux Stimulates Transcription from the DNA Polymerase α Gene Promoter. *Mol. Cell. Biol.* **23**, 3013–3028 (2003).

51. Moon, N. S. *et al.* S phase-specific proteolytic cleavage is required to activate stable DNA binding by the CDP/Cut homeodomain protein. *Mol. Cell. Biol.* **21**, 6332–6345 (2001).

52. Goulet, B., Truscott, M. & Nepveu, A. A novel proteolytically processed CDP/Cux isoform of 90 kDa is generated by cathepsin L. *Biol. Chem.* **387**, 1285–1293 (2006).

53. Liu, S., Mcleod, E. & Jack, J. Four distinct regulatory regions of the cut locus and their effect on cell type specification in Drosophila. *Genetics* (1991).

54. Tavares, A. T., Tsukui, T. & Belmonte, J. C. I. Evidence that members of the Cut/Cux/CDP family may be involved in AER positioning and polarizing activity during chick limb development. *Development* **127**, 5133–5144 (2000).

55. Sharma, M., Fopma, A., Brantley, J. G. & Vanden Heuvel, G. B. Coexpression of Cux-1 and Notch signaling pathway components during kidney development. *Dev. Dyn.* **231**, 828–838 (2004).

56. Topka, S., Glassmann, A., Weisheit, G., Schüller, U. & Schilling, K. The Transcription Factor Cux1 in Cerebellar Granule Cell Development and Medulloblastoma Pathogenesis. *Cerebellum* **13**, 698–712 (2014).

57. Li, N., Zhao, C. T., Wang, Y. & Yuan, X. B. The transcription factor Cux1 regulates dendritic morphology of cortical pyramidal neurons. *PLoS One* **5**, (2010).

58. Nieto, M. *et al.* Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. *J. Comp. Neurol.* **479**, 168–180 (2004).

59. Vanden Heuvel, G. B., Bodmer, R., Mcconnell, K. R., Nagami, G. T. & Igarashi, P. Expression of a cut-related homeobox gene in developing and polycystic mouse kidney. *Kidney Int.* **50**, 453–461 (1996).

60. Pattison, S., Skalnik, D. G. & Roman, A. CCAAT displacement protein, a regulator of differentiation-specific gene expression, binds a negative regulatory element within the 5' end of the human papillomavirus type 6 long control region. *J. Virol.* **71**, 2013–2022

(1997).

61. Skalnik, D. G., Strauss, E. C. & Orkin, S. H. CCAAT displacement protein as a repressor of the myelomonocytic-specific gp91-phox gene promoter. *J. Biol. Chem.* **266**, 16736–16744 (1991).

62. Superti-Furga, G., Barberis, A., Schreiber, E. & Busslinger, M. The protein CDP, but not CP1, footprints on the CCAAT region of the γ-globin gene in unfractionated B-cell extracts. *BBA - Gene Struct. Expr.* **1007**, 237–242 (1989).

63. An, N. *et al.* Gene dosage effect of CUX1 in amurinemodel disruptsHSC homeostasis and controls the severity and mortality of MDS. *Blood* **131**, 2682–2697 (2018).

64. Mailly, F., Be, G., Phillips, S. & Nepveu, A. The Human Cut Homeodomain Protein Can Repress Gene Expression by Two Distinct Mechanisms : Active Repression and Competition for Binding Site Occupancy. **16**, 5346–5357 (1996).

65. Xu, H. *et al.* A group of tissue-specific microRNAs contribute to the silencing of CUX1 in different cell lineages during development. *J. Cell. Biochem.* **119**, 6238–6248 (2018).

66. Ellis, T. *et al.* The transcriptional repressor CDP (Cutl1) is essential for epithelial cell differentiation of the lung and the hair follicle. *Genes Dev.* **15**, 2307–2319 (2001).

67. Sinclair, A. M. *et al.* Lymphoid apoptosis and myeloid hyperplasia in CCAAT displacement protein mutant mice. *Blood* **98**, 3658–3667 (2001).

68. Wang, Y. *et al.* Modulation of CCAAT/enhancer-binding protein-α gene expression by metabolic signals in rodent adipocytes. *Endocrinology* **140**, 2938–2947 (1999).

69. Lawson, N. D., Khanna-Gupta, A. & Berliner, N. Isolation and characterization of the cDNA for mouse neutrophil collagenase: Demonstration of shared negative regulatory pathways for neutrophil secondary granule protein gene expression. *Blood* **91**, 2517–2524 (1998).

70. Khanna-Gupta, A., Zibello, T., Kolla, S., Neufeld, E. J. & Berliner, N. CCAAT displacement protein (CDP/cut) recognizes a silencer element within the lactoferrin gene promoter. *Blood* **90**, 2784–2795 (1997).

71. Michl, P. *et al.* CUTL1 is a target of TGFβ signaling that enhances cancer cell motility and invasiveness. *Cancer Cell* **7**, 521–532 (2005).

72. Coqueret, O., Bérubé, G. & Nepveu, A. The mammalian Cut homeodomain protein functions as a cell-cycle-dependent transcriptional repressor which downmodulates p21(WAF1/CIP1/SDI1) in S phase. *EMBO J.* **17**, 4680–4694 (1998).

73. Reardon, D. A. *et al.* Extensive genomic abnormalities in childhood medulloblastoma by comparative genomic hybridization. *Cancer Res.* **57**, 4042–4047 (1997).

74. Ramdzan, Z. M. *et al.* RAS Transformation Requires CUX1-Dependent Repair of Oxidative DNA Damage. *PLoS Biol.* **12**, (2014).

75. Ramdzan, Z. M. *et al.* The DNA repair function of CUX1 contributes to radioresistance. *Oncotarget* **8**, 19021–19038 (2017).

76. Peters J., G. *et al.* Effects of gemcitabine on cis-platinum-DNA adduct formation and repair in a panel of gemcitabine and cisplatin-sensitive or -resistant human ovarian cancer cell lines. *Int J Oncol* **28**, 237–244 (2006).

77. Carson, J. P. *et al.* Pharmacogenomic Identification of Targets for Adjuvant Therapy with the Topoisomerase Poison Camptothecin. *Cancer Res.* **64**, 2096–2104 (2004).

78. Krug, S. *et al.* CUX1: A modulator of tumour aggressiveness in pancreatic neuroendocrine neoplasms. *Endocr. Relat. Cancer* **21**, 879–890 (2014).

79. Ripka, S. *et al.* CUX1: Target of Akt signalling and mediator of resistance to apoptosis in pancreatic cancer. *Gut* **59**, 1101–1110 (2010).

80. Kühnemuth, B. *et al.* CUX1 modulates polarization of tumor-associated macrophages by antagonizing NF-κB signaling. *Oncogene* **34**, 177–187 (2015).

81. Slowikowski, K. *et al.* CUX1 and IκBζ mediate the synergistic inflammatory response to TNF and IL-17A in stromal fibroblasts. *bioRxiv* **1000**, 571315 (2019).

82. Ueda, Y., Su, Y. & Richmond, A. CCAAT displacement protein regulates nuclear factor-kappa beta-mediated chemokine transcription in melanoma cells. *Melanoma Res.* **17**, 91–103 (2007).

83. Fan, X. *et al.* The transcription factor CUTL1 is associated with proliferation and prognosis in malignant melanoma. *Melanoma Res.* **24**, 198–206 (2014).

84. Darsigny, M., St-Jean, S. & Boudreau, F. Cux1 transcription factor is induced in inflammatory bowel disease and protects against experimental colitis. *Inflamm. Bowel Dis.* **16**, 1739–1750 (2010).

85. Jing, H., Yen, J. H. & Ganea, D. A novel signaling pathway mediates the inhibition of CCL3/4 expression by prostaglandin E2. *J. Biol. Chem.* **279**, 55176–55186 (2004).

86. Nirodi, C. *et al.* The Role of CDP in the Negative Regulation of CXCL1 Gene Expression. *J. Biol. Chem.* **276**, 26122–26131 (2001).

87. Burton, L. J. *et al.* CCAAT-displacement protein/cut homeobox transcription factor (CUX1) represses estrogen receptor-alpha (ER-α) in triple-negative breast cancer cells and can be antagonized by muscadine grape skin extract (MSKE). *PLoS One* **14**, 1–17 (2019).

88. Wang, Z. *et al.* PIK3CA Is Regulated by CUX1, Promotes Cell Growth and Metastasis in Bladder Cancer via Activating Epithelial-Mesenchymal Transition. *Front. Oncol.* **10**, 1–14 (2020).

89. Dorris, E. R. *et al.* The transcription factor CUX1 negatively regulates invasion in castrate resistant prostate cancer. *Oncotarget* **11**, 846–857 (2020).

90. Cadieux, C., Yao, L., Vadnais, C., Drossos, M. & Nepveu, A. Mouse Mammary Tumor

Virus p75 and p110 CUX1 Transgenic Mice Develop Mammary Tumors of Various Histologic Types. 7188–7198 (2009). doi:10.1158/0008-5472.CAN-08-4899

91. Latreille, R. *et al.* Transcription factor CUX1 is required for intestinal epithelial wound healing and targets the VAV2-RAC1 Signalling complex. *Biochim. Biophys. Acta - Mol. Cell Res.* **1864**, 2347–2355 (2017).

92. Grimwade, D. *et al.* Refinement of cytogenetic classification in acute myeloid leukemia: Determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* **116**, 354–365 (2010).

93. Smith, S. M. *et al.* Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: The University of Chicago series. *Blood* **102**, 43–52 (2003).

94. Qian, Z. *et al.* Cytogenetic and genetic pathways in therapy-related acute myeloid leukemia. *Chem. Biol. Interact.* **184**, 50–57 (2010).

95. Kere, J., Ruutu, T. & de la Chapelle, A. Monosomy 7 in granulocytes and monocytes in MDS. (1985).

96. Abrahamson, G. *et al.* Clonality of cell populations in refractory anaemia using combined approach of gene loss and X-linked restriction fragment length polymorphism-methylation analyses. *Br. J. Haematol.* **79**, 550–555 (1991).

97. Will, B. *et al.* Stem and progenitor cells in myelodysplastic syndromes show aberrant stage-specific expansion and harbor genetic and epigenetic alterations. *Blood* **120**, 2076–2086 (2012).

98. Elias, H. K. *et al.* Stem cell origin of myelodysplastic syndromes. *Oncogene* **33**, 5139–5150 (2013).

99. Wong, J. C. *et al.* Functional evidence implicating chromosome 7q22 haploinsufficiency in myelodysplastic syndrome pathogenesis. *Elife* **4**, 1–16 (2015).

100. Pang, W. W. *et al.* Hematopoietic stem cell and progenitor cell mechanisms in myelodysplastic syndromes. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3011–3016 (2013).

101. Schiffer, C. A., Lee, E. J., Tomiyasu, T., Wiernik, P. H. & Testa, J. R. Prognostic impact of cytogenetic abnormalities in patients with de novo acute nonlymphocytic leukemia. *Blood* **73**, 263–270 (1989).

102. Godley, L. A. & Larson, R. A. Therapy-Related Myeloid Leukemia. *Semin. Oncol.* **35**, 418–429 (2008).

103. McNerney, M. E., Godley, L. A. & Le Beau, M. M. Therapy-related myeloid neoplasms: When genetics and environment collide. *Nat. Rev. Cancer* **17**, 513–527 (2017).

104. Le Beau, M. M. *et al.* Cytogenetic and molecular delineation of a region of chromosome 7

commonly deleted in malignant myeloid diseases. *Blood* **88**, 1930–1935 (1996).

105. Itzhar, N. *et al.* Chromosomal minimal critical regions in therapy-related leukemia appear different from those of de novo leukemia by high-resolution aCGH. *PLoS One* **6**, (2011).

106. Rücker, F. G. *et al.* Disclosure of candidate genes in acute myeloid leukemia with complex karyotypes using microarray-based molecular characterization. *J. Clin. Oncol.* **24**, 3887–3894 (2006).

107. Fischer, K. *et al.* Molecular cytogenetic delineation of deletions and translocations involving chromosome band 7q22 in myeloid leukemias. *Blood* **89**, 2036–2041 (1997).

108. Liang, H. *et al.* Molecular anatomy of chromosome 7q deletions in myeloid neoplasms: Evidence for multiple critical loci. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3781–3785 (1998).

109. Bejar, R., Levine, R. & Ebert, B. L. Unraveling the molecular pathophysiology of myelodysplastic syndromes. *J. Clin. Oncol.* **29**, 504–515 (2011).

110. Jerez, A. *et al.* Loss of heterozygosity in 7q myeloid disorders: Clinical associations and genomic pathogenesis. *Blood* **119**, 6109–6117 (2012).

111. Hosono, N. *et al.* Recurrent genetic defects on chromosome 7q in myeloid neoplasms. *Leukemia* **28**, 1348–1351 (2014).

112. McNerney, M. E. *et al.* CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. *Blood* **121**, 975–983 (2013).

113. Chen, C. *et al.* MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **25**, 652–665 (2014).

114. Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat. Genet.* **43**, 875–878 (2011).

115. Ong, C. K. *et al.* Exome sequencing of liver flukeg-associated cholangiocarcinoma. *Nat. Genet.* **44**, 690–693 (2012).

116. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science (80-.).* **331**, 435–439 (2011).

117. Zang, Z. J. *et al.* Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44**, 570–574 (2012).

118. Ernst, T. *et al.* Inactivating mutations of the histone methyltransferase gene EZH2 in myeloid disorders. *Nat. Genet.* **42**, 722–726 (2010).

119. Muto, T. *et al.* Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J. Exp. Med.* **210**, 2627–2639 (2013).

120. Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell

acute lymphoblastic leukemia. *Nat. Med.* **18**, 296–301 (2012).

121. Bracken, A. P. & Helin, K. Polycomb group proteins: Navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* **9**, 773–784 (2009).

122. Karanikolas, B. D. W., Figueiredo, M. L. & Wu, L. Polycomb group protein enhancer of zeste 2 is an oncogene that promotes the neoplastic transformation of a benign prostatic epithelial cell line. *Mol. Cancer Res.* **7**, 1456–1465 (2009).

123. Zhang, Q. *et al.* Mutations in EZH2 are associated with poor prognosis for patients with myeloid neoplasms. *Genes Dis.* **6**, 276–281 (2019).

124. Emerling, B. M. *et al.* MLL5, a homolog of Drosophila trithorax located within a segment of chromosome band 7q22 implicated in myeloid leukemia. *Oncogene* **21**, 4849–4854 (2002).

125. Madan, V. *et al.* Impaired function of primitive hematopoietic cells in mice lacking the mixed-lineage-leukemia homolog MII5. *Blood* **113**, 1444–1454 (2009).

126. Heuser, M. *et al.* Loss of MII5 results in pleiotropic hematopoietic defects, reduced neutrophil immune function, and extreme sensitivity to DNA demethylation. *Blood* **113**, 1432–1443 (2009).

127. Singh, H. *et al.* Putative RNA-splicing gene LUC7L2 on 7q34 represents a candidate gene in pathogenesis of myeloid malignancies. *Blood Cancer J.* **3**, 2–4 (2013).

128. Jourdain, A. A. *et al.* Loss of LUC7L2 and U1 snRNP subunits shifts energy metabolism from glycolysis to OXPHOS. *Mol. Cell* **81**, 1905-1919.e12 (2021).

129. Wong, C. C. *et al.* Inactivating CUX1 mutations promote tumorigenesis. *Nat. Genet.* **46**, 33–38 (2014).

130. Aly, M. *et al.* Distinct clinical and biological implications of CUX1 in myeloid neoplasms. *Blood Adv.* **3**, 2164–2178 (2019).

131. Teodora Kuzmanovic, 1 Bhumika J. Patel, 1 Srinivasa R. Sanikommu, 1 Yasunobu Nagata, 1 Hassan Awada, 1 Cassandra M. Kerr, 1 Bartlomiej P. Przychodzen, 1 Babal K. Jha, 1 Devendra Hiwase, 2 Deepak Singhal, 2 Anjali S. Advani, 3 Aziz Nazha, 3 Aaron T. Gerds, 4. Genomics of therapy-related myeloid neoplasms. **46**, 98–101 (2020).

132. Davoli, T. *et al.* Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell* **155**, 948 (2013).

133. Ebert, B. L. *et al.* Identification of RPS14 as a 5q - syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).

134. Stein, B. L. *et al.* Disruption of the ASXL1 gene is frequent in primary, post-essential thrombocytosis and post-polycythemia vera myelofibrosis, but not essential thrombocytosis or polycythemia vera: Analysis of molecular genetics and clinical phenotypes. *Haematologica* **96**, 1462–1469 (2011).

135. Pellagatti, A. *et al.* Targeted resequencing analysis of 31 genes commonly mutated in myeloid disorders in serial samples from myelodysplastic syndrome patients showing disease progression. *Leukemia* **30**, 247–250 (2016).

136. Supper, E. *et al.* Cut-like homeobox 1 (CUX1) tumor suppressor gene haploinsufficiency induces apoptosis evasion to sustain myeloid leukemia. *Nat. Commun.* **1**, 1–20 (2021).

137. Arthur, R. K., An, N., Khan, S. & McNerney, M. E. The haploinsufficient tumor suppressor, CUX1, acts as an analog transcriptional regulator that controls target genes through distal enhancers that loop to target promoters. *Nucleic Acids Res.* **45**, 6350–6361 (2017).

138. Mcnerney, M. E. *et al.* The spectrum of somatic mutations in high-risk acute myeloid leukaemia with -7/del(7q). *Br. J. Haematol.* **166**, 550–556 (2014).

139. Xu, J. *et al.* Silencing of MBD1 reverses pancreatic cancer therapy resistance through inhibition of DNA damage repair. *Int. J. Oncol.* **42**, 2046–2052 (2013).

140. Imgruet, M. K. *et al.* Loss of a 7q gene, CUX1 , disrupts epigenetically driven DNA repair and drives therapy-related myeloid neoplasms . *Blood* **138**, 790–805 (2021).

141. Goulet, B. *et al.* Characterization of a Tissue-specific CDP/Cux Isoform , p75 , Activated in Breast Tumor Cells. 6625–6633 (2002).

142. Cadieux, C. *et al.* Transgenic mice expressing the p75 CCAAT-displacement protein/cut homeobox isoform develop a myeloproliferative disease-like myeloid leukemia. *Cancer Res.* **66**, 9492–9501 (2006).

143. Vadnais, C. *et al.* Autocrine Activation of the Wnt/ -Catenin Pathway by CUX1 and GLIS1 in Breast Cancers. *Biol. Open* **3**, 937–946 (2014).

144. Goulet, B. *et al.* A cathepsin L isoform that is devoid of a signal peptide localizes to the nucleus in S phase and processes the CDP/Cux transcription factor. *Mol. Cell* **14**, 207–219 (2004).

145. Yu, L. *et al.* Neutrophil elastase-mediated proteolysis of the tumor suppressor p200 CUX1 promotes cell proliferation and inhibits cell differentiation in APL. *Life Sci.* **242**, (2020).

146. Truscott, M. *et al.* Carboxyl-terminal proteolytic processing of CUX1 by a caspase enables transcriptional activation in proliferating cells. *J. Biol. Chem.* **282**, 30216–30226 (2007).

147. Sansregret, L. *et al.* The p110 isoform of the CDP/Cux transcription factor accelerates entry into S phase. *Mol. Cell. Biol.* **26**, 2441–55 (2006).

148. Harada, R. *et al.* Genome-wide location analysis and expression studies reveal a role for p110 CUX1 in the activation of DNA replication genes. *Nucleic Acids Res.* **36**, 189–202 (2008).

149. Goulet, B. *et al.* Increased Expression and Activity of Nuclear Cathepsin L in Cancer

Cells Suggests a Novel Mechanism of Cell Transformation. **5**, 899–908 (2007).

150. Rafn, B. *et al.* ErbB2-Driven Breast Cancer Cell Invasion Depends on a Complex Signaling Network Activating Myeloid Zinc Finger-1-Dependent Cathepsin B Expression. *Mol. Cell* **45**, 764–776 (2012).

151. Kedinger, V. *et al.* p110 CUX1 homeodomain protein stimulates cell migration and invasion in part through a regulatory cascade culminating in the repression of E-cadherin and occludin. *J. Biol. Chem.* **284**, 27701–27711 (2009).

152. Griesmann, H. *et al.* CUX1 enhances pancreatic cancer formation by synergizing with KRAS and inducing MEK / ERK-Dependent Proliferation. *Cancers (Basel).* (2021).

153. Moon, N. S. *et al.* Expression of N-terminally truncated isoforms of CDP/CUX is increased in human uterine leiomyomas. *Int. J. Cancer* **100**, 429–432 (2002).

154. Maitra, U., Seo, J., Lozano, M. M. & Dudley, J. P. Differentiation-Induced Cleavage of Cutl1/CDP Generates a Novel Dominant-Negative Isoform That Regulates Mammary Gene Expression. *Mol. Cell. Biol.* **26**, 7466–7478 (2006).

155. Goulet, B., Markovic, Y., Leduy, L. & Nepveu, A. Proteolytic Processing of Cut Homeobox 1 by Neutrophil Elastase in the MV4;11 Myeloid Leukemia Cell Line. *Mol. Cancer Res.* **6**, 644–653 (2008).

156. Gocheva, V. *et al.* Distinct roles for cysteine cathepsin genes in multistage tumorigenesis. *Genes Dev.* **20**, 543–556 (2006).

157. Sevenich, L. *et al.* Synergistic antitumor effects of combined cathepsin B and cathepsin Z deficiencies on breast cancer progression and metastasis in mice. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2497–2502 (2010).

158. Vasiljeva, O. *et al.* Tumor cell-derived and macrophage-derived cathepsin B promotes progression and lung metastasis of mammary cancer. *Cancer Res.* **66**, 5242–5250 (2006).

159. Turk, V., Turk, B. & Turk, D. Lysosomal cysteine proteases: facts and opportunities. *EMBO J.* **20**, 4629–4633 (2001).

160. Hsing, L. C. & Rudensky, A. Y. The lysosomal cysteine proteases in MHC class II antigen presentation. *Immunol. Rev.* **207**, 229–241 (2005).

161. Bulynko, Y. A., Hsing, L. C., Mason, R. W., Tremethick, D. J. & Grigoryev, S. A. Cathepsin L Stabilizes the Histone Modification Landscape on the Y Chromosome and Pericentromeric Heterochromatin. *Mol. Cell. Biol.* **26**, 4172–4184 (2006).

162. Goulet, B. & Nepveu, A. Complete and limited proteolysis in cell cycle progression. *Cell Cycle* **3**, 984–987 (2004).

163. Qi, R., Singh, D. & Kao, C. C. Proteolytic processing regulates toll-like receptor 3 stability and endosomal localization. *J. Biol. Chem.* **287**, 32617–32629 (2012).

164. Burton, L. J. *et al.* Targeting the Nuclear Cathepsin L CCAAT Displacement Protein/Cut

Homeobox Transcription Factor-Epithelial Mesenchymal Transition Pathway in Prostate and Breast Cancer Cells with the Z-FY-CHO Inhibitor. *Mol. Cell. Biol.* **37**, e00297-16 (2017).

165. Ledford, A. W. *et al.* Deregulated expression of the homeobox gene Cux-1 in transgenic mice results in downregulation of p27kip1 expression during nephrogenesis, glomerular abnormalities, and multiorgan hyperplasia. *Dev. Biol.* **245**, 157–171 (2002).

166. De Celis, J. F., Garcia-Bellido, A. & Bray, S. J. Activation and function of Notch at the dorsal-ventral boundary of the wing imaginai disc. *Development* **122**, 359–369 (1996).

167. Neumann, C. J. & Cohen, S. M. A hierarchy of cross-regulation involving Notch, wingless, vestigial and cut organizes the dorsal/ventral axis of the Drosophila wing. *Development* **122**, 3477–3485 (1996).

168. Nepveu, A. Role of the multifunctional CDP/Cut/Cux homeodomain transcription factor in regulating differentiation, cell growth and development. *Gene* **270**, 1–15 (2001).

169. Sharma, M. *et al.* The homeodomain protein Cux1 interacts with Grg4 to repress p27kip1 expression during kidney development. *Gene* **439**, 87–94 (2009).

170. Wilson, B. J., Harada, R., Leduy, L., Hollenberg, M. D. & Nepveu, A. CUX1 transcription factor is a downstream effector of the proteinase-activated receptor 2 (PAR 2). *J. Biol. Chem.* **284**, 36–45 (2009).

171. Coqueret, O., Bérubé, G. & Nepveu, A. DNA binding by cut homeodomain proteins is down-modulated by protein kinase C. *J. Biol. Chem.* **271**, 24862–24868 (1996).

172. Michl, P., Knobel, B. & Downward, J. CUTL1 is phosphorylated by protein kinase A, modulating its effects on cell proliferation and motility. *J. Biol. Chem.* **281**, 15138–15144 (2006).

173. Santaguida, M. *et al.* Phosphorylation of the CCAAT Displacement Protein (CDP)/ Cux Transcription Factor by Cyclin A-Cdk1 Modulates Its DNA Binding Activity in G 2 *. **276**, 45780–45790 (2001).

174. Sansregret, L. *et al.* Hyperphosphorylation by Cyclin B / CDK1 in Mitosis Resets CUX1 DNA Binding Clock at Each Cell Cycle * □. **285**, 32834–32843 (2010).

175. Truscott, M., Harada, R., Vadnais, C., Robert, F. & Nepveu, A. p110 CUX1 Cooperates with E2F Transcription Factors in the Transcriptional Activation of Cell Cycle-Regulated Genes. *Mol. Cell. Biol.* **28**, 3127–3138 (2008).

176. Van Wijnen, A. J. *et al.* CDP/cut is the DNA-binding subunit of histone gene transcription factor HiNF-D: A mechanism for gene regulation at the Gi/S phase cell cycle transition point independent of transcription factor E2F. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 11516–11521 (1996).

177. Stein, G. S., Stein, J. L., Wijnen, A. J. Van & Lian, J. B. Histone gene transcription: A model for responsiveness to an integrated series of regulatory signals mediating cell cycle

control and proliferation/differentiation interrelationships. **54**, 393–404 (1994).

178. Nesvizhskii, A. I., Keller, A., Kolker, E. & Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **75**, 4646–4658 (2003).

179. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).

180. Stastna, M. & Van Eyk, J. E. Analysis of protein isoforms: Can we do it better? *Proteomics* **12**, 2937–2948 (2012).

181. Muzny, D. M. *et al.* Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).

182. Ramdzan, Z. M. & Nepveu, A. CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nat. Rev. Cancer* **14**, 673–682 (2014).

183. Sansregret, L. & Nepveu, A. The multiple roles of CUX1: Insights from mouse models and cell-based assays. *Gene* **412**, 84–94 (2008).

184. Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, 1–20 (2013).

185. Gundry, M. C. *et al.* Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9. *CellReports* **17**, 1453–1461 (2016).

186. Brunetti, L. *et al.* Mutant NPM1 Maintains the Leukemic State through HOX Expression. *Cancer Cell* **34**, 499-512.e9 (2018).

187. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).

188. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

189. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, 2478–2492 (2017).

190. Karabacak Calviello, A., Hirsekorn, A., Wurmus, R., Yusuf, D. & Ohler, U. Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets via protocol-specific bias modeling. *bioRxiv* 284364 (2018). doi:10.1101/284364

191. Weintraub, H. & Groudine, M. Chromosomal Subunits in Active Genes Have an Altered Conformation. *Science (80-.).* **193**, 848–856 (1976).

192. Dreos, R., Ambrosini, G., Périer, R. C. & Bucher, P. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. *Nucleic Acids Res.* **41**, 157–164 (2013).

193. Petryszak, R. *et al.* Expression Atlas update - An integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* **44**, D746–D752 (2016).

194. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**, 501–504 (2005).

195. Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes (Genome Research (2009) 19, (1316-1323)). *Genome Res.* **19**, 1506 (2009).

196. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res.* **22**, 1760–1774 (2012).

197. Hubbard, T. *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30**, 38–41 (2002).

198. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).

199. Lin, A., Giuliano, C. J., Sayles, N. M. & Sheltzer, J. M. CRISPR/Cas9 mutagenesis invalidates a putative cancer dependency targeted in on-going clinical trials. *Elife* **6**, 1–17 (2017).

200. Lin, A. *et al.* Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl. Med.* **11**, (2019).

201. Chen, J. *et al.* Ephrin receptor A2 is a functional entry receptor for Epstein-Barr virus. *Nat. Microbiol.* **3**, 172–180 (2018).

202. Neufeld, E. J., Skalnik, D. G., Lievens, P. M. & Orkin, S. H. Human CCAAT displacement proteinis homologous to the Drosophila homeoprotein, cut. **1**, (1992).

203. Cadieux, C. *et al.* Polycystic kidneys caused by sustained expression of Cux1 isoform p75. *J. Biol. Chem.* **283**, 13817–13824 (2008).

204. Ikeda, T. *et al.* Transforming growth factor-β-induced CUX1 isoforms are associated with fibrosis in systemic sclerosis lung fibroblasts. *Biochem. Biophys. Reports* **7**, 246–252 (2016).

205. Sibley, C. R. *et al.* Recursive splicing in long vertebrate genes. *Nature* **521**, 371–375 (2015).

206. Drexler, H. L., Choquet, K. & Churchman, L. S. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. *Mol. Cell* **77**, 985-998.e8 (2020).

207. Siam, R. *et al.* Transcriptional activation of the LatsI tumor suppressor gene in tumors of CUX1 transgenic mice. **10**, (2009).

208. Franch-Expósito, S. *et al.* CNApp, a tool for the quantification of copy number alterations and integrative analysis revealing clinical implications. *Elife* **9**, 1–22 (2020).

209. Wilson, M. A. *et al.* Copy Number Changes Are Associated with Response to Treatment with Carboplatin, Paclitaxel, and Sorafenib in Melanoma. *Clin. Cancer Res.* **22**, 374–382 (2016).

210. Bar, E. E., Lin, A., Tihan, T., Burger, P. C. & Eberhart, C. G. Frequent gains at chromosome 7q34 involving BRAF in pilocytic astrocytoma. *J. Neuropathol. Exp. Neurol.* **67**, 878–887 (2008).

211. Nagel, S. *et al.* Amplification at 7q22 targets cyclin-dependent kinase 6 in T-cell lymphoma. *Leukemia* **22**, 387–392 (2008).

212. Dordević, G. *et al.* EGFR protein overexpression correlates with chromosome 7 polysomy and poor prognostic parameters in clear cell renal cell carcinoma. *J. Biomed. Sci.* **19**, 1–8 (2012).

213. Xu, A. *et al.* Overexpressed P75CUX1 promotes EMT in glioma infiltration by activating β-catenin. *Cell Death Dis.* **12**, 1–15 (2021).

214. Feng, F. *et al.* CUX1 Facilitates the Development of Oncogenic Properties Via Activating Wnt/β-Catenin Signaling Pathway in Glioma. *Front. Mol. Biosci.* **8**, 1–16 (2021).

215. Mrózek, K. Acute Myeloid Leukemia with a Complex Karyotype. *Semin Oncol.* **16**, 387–393 (2013).

216. Li, H. *et al.* Therapeutic targeting of circ-CUX1/EWSR1/MAZ axis inhibits glycolysis and neuroblastoma progression. 1–17 (2019).

217. Wiesner, T. *et al.* Alternative transcription initiation leads to expression of a novel ALK isoform in cancer. *Nature* **526**, 453–457 (2015).

218. Tanaka, M., Suda, T., Takahashi, T. & Nagata, S. Expression of the functional soluble form of human Fas ligand in activated lymphocytes. *EMBO J.* **14**, 1129–1135 (1995).

219. Sakurai, H. *et al.* Overexpression of RUNX1 short isoform plays a pivotal role in the development of myelodysplastic syndromes/myeloproliferative neoplasms. *Exp. Hematol.* **43**, S93 (2015).

220. Kim, J. H. *et al.* SON and Its Alternatively Spliced Isoforms Control MLL Complex-Mediated H3K4me3 and Transcription of Leukemia-Associated Genes. *Mol. Cell* **61**, 859–873 (2016).

221. Grade, M., Difilippantonio, M. J. & Camps, J. Patterns of Chromosomal Aberrations in Solid Tumors. *Chromosom. Instab. Cancer Cells* 1–224 (2015). doi:10.1007/978-3-319-20291-4

222. Shi, J. *et al.* Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).

223. Nishio, H. & Walsh, M. J. CCAAT displacement protein/cut homolog recruits G9a histone lysine methyltransferase to repress transcription. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11257–11262 (2004).