Brd4-Nutm1 fusion gene initiates NUT carcinoma *in vivo*

Dejin Zheng[1,2,*], Ahmed Abdou Mahmoud El Negiry[1,2,5,*], Mohammed Amine Bendahou[3], Liangqi Xie[3], Mayra F. Tsoi[4] and Bin Gu[1,2 #]

Affiliations:

[1] Institute for Quantitative Health Science and Engineering, Michigan State University East Lansing, Michigan, United States

[2] Department of Obstetrics, Gynecology and Reproductive Biology, College of Human Medicine, Michigan State University East Lansing, Michigan, United States

[3] Infection Biology and Cancer Biology Program, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States.

[4] Department of Pathobiology and Diagnostic Investigation, College of Veterinary Medicine, Michigan State University East Lansing, Michigan, United States

[5] Home Institution: Department of Cytology and Histology, Faculty of Veterinary Medicine, Aswan University, Aswan, Egypt

*Equal contribution.

#Correspondence to Bin Gu (gubin1@msu.edu)

Abstract: Nut carcinoma (NC) is an aggressive cancer with no effective treatment. The majority (70%) of NUT carcinoma is associated with chromosome translocation events that lead to the formation of a *BRD4-NUTM1* fusion gene. However, because the *BRD4-NUTM1* gene is unequivocally cytotoxic when ectopically expressed in cell lines, there is no experimental evidence of the oncogenic potential of this fusion gene to initiate NC. We report here the first genetically engineered mouse model (GEMM) for NUT carcinoma. By inducing a chromosome translocation event mirroring the human event in mouse epithelial progenitors, we demonstrated that the *Brd4-Nutm1* fusion gene could induce aggressive head and neck, and skin carcinomas in mice. The tumors present similar histopathological and molecular features to human NC. As the sole immunocompetent GEMM for NC, the model will have a far-reaching impact on understanding NC oncogenesis and developing NC treatment, including targeted therapies and immune therapies in a preclinical setting, which will benefit NC patients through translational research.

Introduction:

NUT carcinomas (NC), formally named NUT Midline Carcinomas (NMC), are poorly-differentiated squamous cell carcinomas that often arise within midline organs, most commonly in head and neck structures and thoracic organs[1-3]. Other primary sites, including skin[4], pancreas[5], and pelvic organs[6,7], have also been reported. They are aggressive tumors that typically present with metastatic disease at diagnosis[8]. The prognosis of NC is dismal, with an overall survival time of 6.5 months, even with intensive treatment, including surgery and chemotherapy[8]. NC is rare; a recent estimation suggests that NC accounts for approximately 0.21% of all cancers, with an estimated 3500 new cases potentially emerging in the U.S. each year[9]. However, because NC lacks a distinctive diagnostic histopathological feature and is not widely recognized by many clinicians, the case rates are likely underestimated. NC is more prevalent in young people, with a medium incidence age in the 20s; it causes a dramatic loss of life[10]. Thus, it is critical to understand the pathogenesis of NC, which will lead to better treatment for NC patients.

Clinical studies since the 1990s have established a strong association of NC with chromosome translocation events that produce fusion genes involving a testis-specific gene Nuclear protein in testis (NUTM1) with various epigenetic factors, including BRD4 (70% of cases), and BRD3, NSD3, ZNF532, etc. (the rest of cases)[1,4,11-15]. Targeted and whole exosome sequencing analysis on limited numbers of human NCs revealed overall low mutation loads and no significant recurrent oncogenic mutations other than the NUTM1 fusion[16]. Furthermore, NC has been reported in young children and infants[5,12,17]. All these clinical data suggest that the NUTM1 fusion gene is the key and likely the sole driver of NC oncogenesis.

At the molecular level, NUTM1 is expressed specifically during spermiogenesis after the stage of mid-round spermatids[18]. Its primary function is to recruit the P300 histone acetyltransferase (HAT), globally acetylate spermatid histones which will then be recognized by the testis-specific Bromodomain and Extraterminal (BET) protein BRDT and replaced by protamine to achieve chromatin condensation in sperm heads[18]. Because the fusion partners to NUTM1 in NC are either BET proteins that recognize acetylated Histone (AceH) or interactors of BET protein, an oncogenesis mechanism was proposed: the AceH reader part of the NUTM1 fusion protein recognizes and binds to an endogenous AceH in cancer originating cells, often on endogenous enhancers, the NUTM1 part will then recruit P300 to acetylate a neighboring histone[19,20]. This positive feedback circle can expand the histone acetylation landscape in the cell and globally activate gene expression[19]. Among the activated genes are oncogenes such as *MYC* family genes which unleash cell proliferation, and stem cell transcription factors such as *SOX2* and *TP63,* which block cell differentiation. Supporting this model, AceH mega-domains and the overexpression of *MYC, SOX2,* and *TP63* have been reported in cells that ectopically express the *BRD4-NUTM1* fusion and cell lines derived from human NC[19]. Functionally, knocking down the expression of *BRD4-NUTM1, MYC, and SOX2* has all been shown to inhibit the survival and proliferation of NC cell lines[21-23]. Furthermore, small molecular inhibitors that block BET binding to AceH (BET inhibitors) or P300 (HAT inhibitors) enzymatic activities have all shown anti-proliferation effects in NC cell lines and some in small-scale clinical trials on NC patients[24-26]. All this evidence

supports a central role of a genetic regulatory network (GRN) coordinated by *BRD4-NUTM1* that is critical for NC oncogenesis and progression.

However, when ectopically expressed in non-NC cell lines, *BRD4-NUTM1* is unequivocally cytotoxic[2,27], contradictory to the proposed oncogenic functions. This cytotoxic effect is not surprising. In addition to proliferation-promoting genes, global activation of chromatin accessibility can activate tumor suppressors, apoptosis-promoting genes, and transposon elements that can cause cell death. Large-scale gene activation only occurs transiently in limited conditions in the mammalian life cycle, mainly during early embryonic development for zygotic genome activation (ZGA) and in terminally differentiated medullary thymic epithelial cells that present tissue-restricted antigens for establishing central immune tolerance[28,29]. To reconcile these contradictory observations, we consider two possible explanations. First, many fusion oncogenes have been shown to require a specific cell lineage status to achieve cell transformation and are otherwise toxic to other cell types. This is coined as a concept of 'cellular pliancy' first in pediatric cancer and then in cancer in general[30-32]. Thus, it is plausible that the *BRD4-NUTM1* oncogene must act on specific cell types to initiate NC. Since NCs are epithelial in origin and have been shown to express marker genes of the basal stem/progenitor cells such as *TP63* and *SOX2*, these cells might be the originating cells of NC. Second, it could simply be that the over-expression system used on cell lines expresses the *BRD4-NUTM1* oncogene at a too-high level. Most overexpression systems use constitutive promoters such as CMV or CAG, which are expected to drive expression levels much higher than the endogenous BET protein promoters that drive the *NUTM1* fusion gene expression *in vivo*. It is not uncommon to observe gene dosage-dependent toxicity of oncogenes[33,34]. Thus, the discrepancy in expression levels could be a simple explanation for the cytotoxicity of the *BRD4-NUTM1* oncogene.

To test whether *BRD4-NUTM1* can drive NC-like cancer, we developed a genetically engineered mouse model (GEMM) for NUT carcinoma by controlling the two key parameters mentioned above. The mouse model recapitulates the endogenous chromosome translocation that forms a *Brd4-Nutm1* fusion gene, thus maintaining an

endogenous expression level of the oncogene. The chromosome translocation is driven by cre-loxp mediated recombination[35,36], thus allowing testing and selection of NC cell of origin in whole organisms using cell-type specific and whole body Cre lines. Using this model, we demonstrated that the *Brd4-Nutm1* fusion gene could drive NC oncogenesis in Keratin 14 expressing basal progenitor cells. This is the first functional evidence demonstrating the oncogenic activity of the *Brd4-Nutm1* fusion gene in NC initiation. The new mouse model can serve as a critical preclinical model for studying this devastating disease.

Results:

***Brd4-Nutm1* induces NC-like tumors from epithelial progenitor cells:**

To experimentally demonstrate the oncogenic activity of the *Brd4-Nutm1* fusion gene *in vivo*, we developed a GEMM model for NC. To induce a chromosome translocation mimicking the human event in NC, we introduced loxp sites into the appropriate introns of the mouse *Brd4* and *Nutm1* genes with the same orientation, similar to previously reported chromosome engineering technologies[36,37] (Figure 1A). To track the development of NC *in vivo*, we inserted a Luc2TdTomato reporter cassette downstream of the endogenous *Nutm1* coding sequence, separated by a T2A self-cleaving sequence[38]. Without cre activity, the mice carrying both loxp sites and the Nut-linked reporters express the Luc2TdTomato reporter specifically in post-meiotic spermatids as expected (Fig 1B and C). The expression of the cre-recombinase in a given cell population will induce chromosome translocation that results in the generation and expression of the *Brd4-Nutm1* fusion gene at a low frequency. Because the expression of the fusion gene will be under the endogenous gene expression control mechanisms of the *Brd4* gene, this design recapitulates the native expression control and avoid any ectopic overexpression. If any of the *Brd4-Nutm1* are transformed and grow to a critical mass that reaches the detection sensitivity of BLI, somatic BLI signals representing perspective tumors can be traced and monitored. For brevity, mice carrying both loxp sites and the reporter cassette linked to $Nutm1^{loxp}$ is designated NC translocator (NCT). Thus, from this point, unless otherwise explained, $NCT^{+/wt}$ refers to $Brd4^{loxp/wt}$; $Nutm1^{loxp/wt}$ whereas $NCT^{+/+}$ refers to $Brd4^{loxp/loxp}$;$Nutm1^{loxp/loxp}$.

Most NC in humans are squamous cell carcinomas; around 40% of them arise in epithelial tissues of the head and neck region, including oronasal mucosa, salivary gland, and pharyngeal-laryngeal mucosa[3]. Cutaneous NC has also been reported in patients[4]. Molecularly, NC expresses markers, including TP63, SOX2, and cMYC, reminiscent of a molecular phenotype of epithelial basal progenitor cells[39-41]. These clinical and molecular patterns suggest that a substantial proportion of NC might arise from basal progenitor cells of the embryonic ectoderm-derived epithelial tissues, particularly in the head and neck region. Thus, we crossed the NCT mouse line to a K14Cre mouse line that expresses the Cre recombinase in ectoderm-derived K14 positive cells[42], including a stem/progenitor cell population in these tissues. As shown in Figure 2A, there were apparent BLI signals in somatic tissues already at weaning (postnatal day 21) in K14Cre$^{+/-}$; NCT$^{+/wt}$ mice, whereas no signal was observed in control littermates. The oncogenic activity of Brd4-Nutm1 is reproducible among individuals, as all (n>50) K14Cre$^{+/-}$; NCT$^{+/wt}$ mice developed somatic BLI signals that were eventually demonstrated to be tumors (Supplementary Figure 1). As indicated by the BLI signals, all mice developed tumors in the head region. At the same time, some of them also developed tumors at other anatomical locations, such as the back skin (Supplementary Figure 1). These locations were then confirmed by autopsy. Using BLI signals, we observed continuous growth of tumors in K14Cre$^{+/-}$; NCT$^{+/wt}$ mice (Figure 2B). By quantifying the BLI signal, we measured the growth kinetics of the tumor growth in mouse heads. As shown in Figure 2C, all individual tumors grew exponentially, demonstrating their aggressive nature. All K14Cre$^{+/-}$; NCT$^{+/wt}$ were dead or reached an endpoint by the age of seven weeks (with a median survival time of 5.5 weeks) (Fig. 2D), likely due to obstruction of the airway or digestive tract by the large tumor burden in the oral cavity and extending into the nasal sinuses. Through autopsy and histopathologic analysis, we identified tumors in multiple anatomical locations, including the oral cavity, salivary glands, eyelids, external ear canal, and dorsal skin (Supplementary Table 1). This pattern was consistent with the expected K14Cre expression profile in mice and the reported presentation of NC in humans. Thus, the

induction of the *Brd4-Nutm1* fusion gene drives aggressive cancer in mice with similar tissue distribution as in human NC.

Mouse NC Presents Histopathological Features of Human NC.

Most human NCs are categorized as poorly differentiated squamous cell carcinomas, characterized by primitive round cells with variable areas of abrupt squamous differentiation and keratinization[9]. This histological presentation indicates a poorly differentiated status of NC cells. As shown in Figure 3, mouse NC consists of large areas of basaloid, polygonal, or spindle-shaped cells. Many tumors exhibit areas of squamous cell differentiation and keratinization that occupy variable proportions of the tumor. Thus, tumors developed in the NC GEMM present similar histopathological features as human NC. Henceforth, we will refer to them as mouse NC.

Mouse NC Presents Molecular Features of Human NC.

Because of the variable presentation, histopathological features alone are not sufficient to diagnose NC. In clinical practice, NC is diagnosed based on two criteria: the presence of the t(15;19) chromosome translocation detected by Fluorescence In Situ Hybridization (FISH)[12]; and the presence of the NUTM1 proteins detected by immunohistochemistry[43]. Accordingly, we confirmed the fusion gene formation in mouse NC using dual-color FISH with DNA probes recognizing the *Brd4* and *Nutm1* fragments of the fusion gene (Figure 4A). Further validating the fusion gene formation, we detected the expected cre-loxp-mediated fusion junction by Sanger sequencing (Figure 4B). Due to limited amino acid sequence conservation, the NUTM1 broadly used for human NC diagnosis (CST C52B1) does not recognize the mouse NUTM1 protein. We used a rabbit monoclonal antibody generated by immunization with a recombinant mouse NUTM1 protein fragment (Huabio HA721690). As shown in Figure 4C, immunofluorescence(IF) with this antibody showed a specific diffuse nuclear stain on round spermatids in mouse testis, the cell type known to express NUTM1 specifically. In mouse NC tissue, IF staining with the mouse NUTM1 antibody showed robust signals in NC cells (Figure 4C). NUTM1 signals present as nuclear foci, similar to the previously reported pattern of NUTM1 in human NC cells[20], representing the BRD4-

NUTM1 proteins concentrated on AcH megadomains[19]. Since immunohistochemistry (IHC) on formalin-fixed, paraffin-embedded tissue sections, rather than IF on cryosections, is the gold standard for diagnosis, we performed IHC of NUTM1 on whole head sections of NC-bearing mice. As shown in Figure 4D, NUTM1 labelling was limited to neoplastic cells , whereas they were absent from normal tissues. At higher magnification, NC cells showed strong nuclear labelling for NUTM1, consistent with the reported pattern in human NC. Importantly, there was no NUTM1 labelling in stromal cells surrounding mouse NC, demonstrating the specificity of the *Brd4-Nutm1* expression in NC cells.

In addition to the *NUTM1* fusion genes, human NC has been reported to express other hallmark genes, including the basal progenitor cell markers P63, Krt14, SOX2, and oncogene cMYC. The majority of mouse NC cells express P63, Krt14, and cMYC (Figure 5A). However, SOX2 is expressed in a small proportion of NC cells with a clustered pattern in all tested mouse NCs (Figure 5B). Previously, it was reported that SOX2 was expressed in the majority of cells in tissue samples of two NC cases[23]. It remains to be seen whether and how SOX2 is expressed in more human NC samples in addition to other stem cell markers.

As previously reported, the *Brd4-Nutm1* drives the formation of large histone acetylation domains (called mega-domains) when ectopically expressed and in NC cell lines through a positive feedback mechanism. Phenotypically, these mega-domains present as large nuclear foci recognized by immunofluorescence staining against Acetyl-histone marks such as Histone H3K27ace. IF analysis demonstrated that this prominent feature was recapitulated in mouse NC. In these experiments, all tissue sections were counterstained with a goat antibody against the TdTomato reporter protein linked to the NUTM1 via a self-cleaving T2A peptide, thus demarcating NC cells (TdTomato^{+}) from peri-NC stromal cells (TdTomato^{-}). As shown in Figure 5B, NC cells present clear H3K27ace nuclear foci, while stromal cells present a diffused and weaker H3K27ace nuclear stain.

In our model, cells contain one copy of the *Brd4-Nutm1* fusion gene while maintaining one normal copy of the *Brd4* and *Nutm1* gene. Theoretically, the wildtype BRD4 proteins from the normal copy can be recruited into the BRD4-NUTM1 nuclear foci by

two potential mechanisms: 1. By binding to the acetylated histones underlying the mega-domain; 2 by biological condensate-related mechanisms through protein interactions mediated by low-complexity domains presented on both the BRD4 and the NUTM1 proteins. To test if wildtype Brd4s are indeed enriched in the Brd4-Nutm1 foci, we stained mouse NC tissues with an antibody that recognizes the c-terminal end of the BRD4 protein, thus, will only recognize the wildtype proteins. As shown in Figure 5C, the wildtype BRD4 protein shows clear nuclear foci. Therefore, although the exact mechanism remains to be discovered, the BRD4-NUTM1 nuclear foci may serve as hubs for enriching transcription factors such as BRD4.

Mouse NC demonstrates human-like histopathological heterogeneity.

Human NC demonstrates remarkable heterogeneous histopathological characteristics. Although they generally consist of large areas of undifferentiated basaloid cells. Yet, they exhibit a spectrum of levels of differentiation, from focally abrupt squamous differentiation to a diffuse well-differentiated phenotype[2,9]. There is also no clear correlation between the anatomical locations of primary NC and their histopathological phenotype. However, because only a small number of NC patients has been reported worldwide, and the number of tumors reported in each anatomical location is even smaller, it is challenging to dissociate the individual genetic effects from the intrinsic heterogeneity conferred by NC biology. We compared multiple tumors grown in the same individual to investigate the histopathological presentation under a homogeneous genetic background. Illustrated in Figure 6A, multiple tumors developed within the head of this mouse. While the right maxillary tumor (b) and the two mandibular tumors (c and d) are poorly differentiated carcinomas consisting predominantly of basaloid cells,, the left maxillary tumor (a) is well differentiated with prominent keratinization. They also have variable degrees of fibrous stromal components. Similar histopathological heterogeneity was observed in multiple NC mice, indicating that heterogeneity can arise even under the same individual germline genetic composition.

To further characterize the differentiation status of NCs, we performed IHC staining of the Involucrin (IVL) protein (Figure 6B), a well-established marker of differentiated

keratinocytes in squamous epithelium. Interestingly, the expression of IVL does not always correlate with the degree of differentiation, as defined by a board-certified veterinarian pathologist. This suggests that NC cells may not follow a similar differentiation process to normal keratinocytes, and molecular characterization is vital for NC research and diagnosis.

Furthermore, we compared the histopathological presentation of NC arising in similar anatomical locations from different mice. As shown in Figure 6C, high heterogeneity between mice was observed among tumors derived from the periodontal region. Thus, although new association patterns between the histopathological presentation of NC and individual genetic composition or anatomical location of tumors could emerge along with investigations into more human NC patient samples, our data support that NC is a heterogeneous disease, and that this heterogeneity is likely defined by non-genetic factors.

Limited immune cell infiltration in mouse NC:

NC is not expected to have robust immune cell infiltration as a cancer type driven by a single fusion oncogene with a low mutation burden. Previously, a few clinical reports suggested neutrophil infiltration in human NC[44,45]. However, there is no consensus on the immune infiltration phenotype of NC. By examining the H&E-stained tissue sections, we rarely observed robust immune cell infiltration in mouse NC. To corroborate this observation, we performed immunofluorescence targeting different immune cell markers, including CD3ε, CD4 and CD8(T-cells); CD20 (B-cells), CD161c (NK cells), Gr-1 (Granulocytes) and F4/80 (Macrophage). As shown in Figure 7A, there is no marked infiltration by any of these immune cell types in NC tumors. As for immune checkpoint proteins (Figure 7B), PD-L1 was not consistently expressed on mouse NC cells, suggesting other mechanisms may be implicated in preventing lymphocyte infiltration and killing. The “don’t eat me” ligand CD47 was consistently and diffusely expressed by NC cells, possibly contributing to the suppression of anti-tumor activities of NC-associated macrophages. Additionally, the MHC class I molecule expression was not detected on most NC cells, though there is positive staining on peri-NC stromal cells. This suggests that the loss of expression of MHC class I might be another

immune-suppressive mechanism in NC. Thus, mouse NC presents an immune-desert-like phenotype for which little immune infiltration was observed. How this immune phenotype contributes to the disease progression and its interaction with NC therapies can be further investigated using this GEMM.

Discussion:

Nut carcinoma is an aggressive cancer with a dismal prognosis. Because NC is extremely rare in the human population and is often diagnosed at a late disease stage, relevant *in vivo* models are vital for understanding its pathogenesis and for therapy development[32,46]. The new NC GEMM addressed key issues that are challenging with current human cell line and xenograft models and created new opportunities for understanding NC.

First, a long-standing dilemma in NUT carcinoma research is that although the disease is strongly associated with chromosome translocations events that induce the expression of NUTM1 fusion genes such as *BRD4-NUTM1*, the ectopic expression of the fusion genes in non-NC cells are unequivocally cytotoxic[2,27]. This dilemma raises a critical question of whether the *NUTM1* fusion genes are indeed oncogenic. The NC GEMM demonstrated that when expressed at the endogenous level in ectoderm-derived epithelial progenitor cells, *BRD4-NUTM1* can efficiently induce squamous cell carcinoma with variable degrees of differentiation. Furthermore, the tumor that BRD4-NUTM1 induced presented similar histopathological and molecular features of human NCs. This is the first functional demonstration that the *BRD4-NUTM1* fusion gene is sufficient to induce NC-like tumors *in vivo* from a cell population that includes epithelial progenitor cells. However, given the fact that human NC can arise in many tissues, some of which, such as the lung and pancreas, have different stem cell hierarchies and molecular programs[47,48], further research is required to investigate whether the K14-positive ectoderm-derived epithelial progenitor cellular context is necessary for NC initiation. The NCT mouse strain has great potential for this investigation since it can simply be crossed to any cell type-specific Cre driver for testing the oncogenic potential of *BRD4-NUTM1*. Combining this powerful experimental model with the ever-advancing single cell/spatial genomics/transcriptomics technologies, the GEMM can open up new routes for pinpointing NC's cell of origin.

Second, human NCs present remarkable phenotypical heterogeneity. However, due to the small number of cases reported today, whether these heterogeneities reflect human population genetic variability or intrinsic heterogeneity of molecular actions of BRD4-NUTM1 fusion protein in different tumors needs to be clarified. The NC GEMM offered a genetically homogeneous system to investigate this problem. As we described, mouse NC exhibits substantial histopathological heterogeneity even within a single individual. These results suggest that non-genetic factors contribute significantly to the phenotypic heterogeneity of NC. Uncovering its mechanism in NC is critical because phenotypic and molecular heterogeneity in cancer favors the development of therapy resistance[49,50]. Previous research has suggested that BRD4-NUTM1 builds AceH megadomains around endogenous enhancers of the affected cell type[19]. Recently, using human pluripotent stem cell differentiation models, a report indicates that the chromatin binding profile of *BRD4-NUTM1* is dependent on cell type[51]. Although these descriptions can explain phenotypic regulatory mechanisms to an extent, these cells (293 cell lines and human pluripotent stem cells and their mesoderm derivatives) are not the appropriate NC-originating cells. They will eventually die after the expression of BRD4-NUTM1. In theory, unlimited numbers of mouse NCs, each with phenotypic characterizations, can be analyzed using comprehensive genomic methods using the NC GEMM. This can provide a clear understanding of the genomic and regulatory mechanisms underlying NC's phenotypic heterogeneity.

Third, as immune therapy has transformed the therapeutic landscape of aggressive cancers such as melanoma and lymphoma, it is critical to investigate the immune system's interaction with targeted or immune therapies for NC. Due to the rare prevalence of NC, it is difficult to investigate the immune effects of NC therapy by studying patient cohorts. Additionally, all xenograft mouse models of NC are established on immunocompromised mice, preventing their use in immunology studies. Accordingly, although there is a report in one patient of the expression of PD-L1 and a clinical response to immune-checkpoint inhibition therapy[52], we still lack an accurate understanding of NC's immune phenotype and the underlying mechanism. The GEMM reported here is the first immune-competent experimental model for NC. Consistent with the previous knowledge that NC is a single oncogene-driven cancer with low mutation

burden and, therefore, likely limited neo-antigens, we found that the mouse NC presented an immune desert phenotype with limited immune cell infiltration. However, even though NCs are likely to have a low mutation burden and limited neo-antigen, they might possess a few potential immune-activating mechanisms. For example, NUTM1 is a testis-specific protein that might naturally activate immune response as a cancer-testis antigen or a vaccine. Also, the presumably global activation of chromatin by the BRD4-NUTM1 protein might activate immune-activating elements such as transposons. Thus the new GEMM can provide a valuable platform for studying the immune therapies for NC.

In summary, we have reported a GEMM for NC that can enhance our understanding of this disease and open up many opportunities for developing targeted therapies.

Materials and Methods:

Ethical statement:

All animal work was carried out under PROTO202000143, approved by the Michigan State University (MSU) Campus Animal Resources (CAR) and Institutional Animal Care and Use Committee (IACUC) in AAALAC credited facilities.

Mouse line generation:

All mouse generation work follows our published 2-cell (2C)- Homologous Recombination (HR)- CRISPR (2C-HR-CRISPR) method and protocol and is described briefly as follows[53-55]. First, 2-cell stage embryos of the CD1 strain were microinjected with Cas9-mSA mRNA (75ng/ul), sgRNA targeting around the stop codon of the Nutm1 gene (GGUGGCCCUCUGCUUCCUAC, chemically modified sgRNA from Synthego Inc. 50ng/ul) and a biotinylated PCR donor consisted of homology arms (5’arm 1000bp and 3’arm 600bp) spanning the coding sequence of T2A-Luc2TdTomato cassette (5ng/ul). The Luc2TdTomato cassette was cloned from the pCDNA3.1(+)/Luc2-TdT plasmid (Addgene 32904, kind gift from Christopher Contag)[38]. Injected embryos were implanted into pseudopregnant females the same day to generate founder pups. Positive founders were screened by long-range PCR spanning homology arms using the following primer sets. 5’ arm gtF: GAGACTGTCATAGACAGCATCCAAGAT and 5’arm gtR:

tcatggctttgtgcagctgc; 3'arm gtF: ggactacaccatcgtggaacagta and 3'arm gtR: CATATTTAACAGTCCCACGGAGAG. A positive founder mouse was out-crossed to CD1 mice to generate N1 mice. The N1 mice were genotyped by PCR. Additionally, genomic regions spanning the targeting cassette and 3' and 5' homology arms were Sanger-sequenced to validate correct targeting. Insertion copy number was evaluated by digital droplet PCR ddqPCR. Heterozygous N1 mice have only one insertion copy, demonstrating single-copy insertion. An N1 that passed all quality controls was further outcrossed with CD1 for two generations to remove any potential off-target mutations introduced by CRISPR-Cas9 editing to generate N3. Heterozygous N3s were bred together to establish a homozygous colony of the Nutm1-T2A-Luc2TdTomato mousseline.

Then two additional mouse lines were generated in parallel. First, a loxp site was inserted into the intron 10 of the Brd4 gene (Brd4loxp) by electroporating CD1 zygotes with Cas9 proteins, a chemically modified sgRNA (GAAUAGAGUGCUGCUACACU, Synthego Inc.), and a single strand oligo DNA donor (ssODN), with 30 base homology arms on each side spanning the loxp site (TCTCCCTGCAGAGTGGCAAGATGCCCAGTGataacttcgtatagcatacattatacgaagttatTAGCAGCACTCTATTCTCTCCATCTGGGTA). Electroporated zygotes were implanted into pseudopregnant females the same day to generate founder pups. Positive founder pups were screened by PCR using primers: BRD4 loxp gt F: TGCCTGGGCTCTTCTGTCCAC, and BRD4 loxp gt R: CAACCCAAGGTTGATGGGACCC. A positive founder mouse was out-crossed to CD1 mice to generate N1 mice. The N1 mice were genotyped by PCR. Additionally, genomic regions spanning the targeting cassette and 3' and 5' homology arms were Sanger-sequenced to validate correct targeting. An N1 that passed all quality controls was further outcrossed with CD1 for two generations to remove any potential off-target mutations introduced by CRISPR-Cas9 editing to generate N3. Second, a loxp site was inserted into the intron 1 of the Nutm1-T2A-Luc2TdTomato gene (Nutm1-loxp) by microinjecting Nutm1-T2A-Luc2TdTomato homozygous 2-cell embryos. (75ng/ul), sgRNA targeting around the intron (ACCGGAUCCCAACCGCUCUA, chemically modified sgRNA from Synthego Inc. 50ng/ul) and a biotinylated PCR donor consisting of

homology arms (5'arm 500bp and 3'arm 500bp) spanning a loxp site and artificial genotyping primer sequences (5ng/ul). Injected embryos were implanted into pseudopregnant females the same day to generate founder pups. Positive founders were screened by long-range PCR spanning homology arms using the following primer sets. 5' arm gtF: GCATGAAATCAGACCAAGTGGGC and 5'arm gtR: AAGCTGACCCTGAAGTTCATCTG; 3'arm gtF: ggactacaccatcgtggaacagta and 3'arm gtR: CATATTTAACAGTCCCACGGAGAG. A positive founder mouse was out-crossed to CD1 mice to generate N1 mice. The N1 mice were genotyped by PCR. Additionally, genomic regions spanning the targeting cassette and 3' and 5' homology arms were Sanger-sequenced to validate correct targeting. An N1 that passed all quality controls was further outcrossed with CD1 for two generations to remove any potential off-target mutations introduced by CRISPR-Cas9 editing to generate N3. Starting from N3s, the Brd4loxp and Nutm1loxp strains were bred to double homozygous to generate the NC translocator (NCT) mouse line.

For inducing the Brd4-Nutm1 fusion in basal progenitor cells of stratified epithelium, NCT+/+ mice were crossed with heterozygous KRT14-Cre mice (Jackson Lab 018964). The KRT14-Cre mouse line from Jackson Lab was on the B6 background and had black fur. To facilitate sensitive bioluminescent imaging, the strain was outcrossed for two generations with CD1 mice to generate albino mice. Progenies from the NCTXKrt14-Cre cross were genotyped; the cre-positive ones were used as experimental groups, while the cre-negative ones were used as controls.

Bioluminescence imaging:

NC mice were subjected to regular bioluminescence imaging to monitor tumor signals every week. The imaging frequency was increased to every other day to generate the tumor growth curve. During the imaging procedure, mice received an intraperitoneal injection of 150 mg/kg D-Luciferin Potassium (Goldbio). After 20 minutes, images of both dorsal and ventral positions were captured using IVIS Spectrum In Vivo Imaging System (Perkin Elmer) under anesthesia. To quantify the tumor signal intensity, regions of interest (ROIs) were delineated around the head area. The total flux (photons/sec) within the ROIs was measured using the Living Image® Software. The average value of the total flux from both dorsal and ventral images was used for generating tumor growth curves.

Necropsy and tissue collection

Routine necropsies were performed by ventral midline approach. Tumors, draining lymph nodes, heads, lungs, and liver were harvested and fixed in 10% neutral buffered formalin for 24 h or flash-frozen in liquid nitrogen for further analyses. Tissue samples were either fixed in 4% paraformaldehyde at 4° C for 24 h or 10% neutral buffered formalin at room temperature for 24 h. Heads were decalcified in 10% EDTA for 12-14 days at 4° C by replacing the decalcification solution every two days. Tissues were either routinely processed, and embedded in paraffin wax or used for cryo-sections.

Histopathology

Serial, 5-µm thick sections were cut and stained with hematoxylin-eosin (H&E). Neoplasms were characterized based on anatomic location, necrosis, inflammation (polymorphonuclear and mononuclear), and histologic features (focal dysplasia, undifferentiated carcinoma, poorly differentiated carcinoma with regional squamous differentiation, and well-differentiated squamous cell carcinoma as described[56]. Histopathologic examination was performed by a board-certified veterinary pathologist (MFT).

Immunohistochemistry

Antigen retrieval was done using sodium citrate pH 6 for 20 min at 97 ^{o}C. Immunohistochemistry was performed using antibodies targeting NUTM1 (1:100, Huabio HA721690), TdTomato (1:500, Origene, ABOO40-500), cMYC (1:500, Proteintech, 10828-1-AP), SOX2 (1:1000, R&D Systems, AF2018), TP63 (1:1000, Sigma Aldrich, AMAB91224), CK14 (1:500, Novus Biologicals, NBP2-67585), and involucrin (1:1000, BioLegend, 924401) using Agilent Dako Autostainer Link 48 and EnVision FLEX kit (Dako). Immunoreactivity for each antibody was assessed by light microscopy.

Cryosection:

Tissues for cryosections were dehydrated using 30% sucrose at 4° C overnight. Samples were embedded in OCT and snap-frozen on a steel bench block immersed in liquid nitrogen. Samples were sectioned at 4-µm thickness using a cryostat (Leica).

Immunofluorescence:

Slides were dried at room temperature for 30 minutes and permeabilized using 0.1% Tween 20 for 30 minutes. Samples were blocked using PBST containing 5% normal donkey serum for 1 hour at room temperature. The primary antibodies (same as described in the IHC section) diluted in blocking solution were applied on tissue sections and incubated overnight at 4 ^{o}C. Slides were rinsed three times for 5 minutes each with PBS. Secondary antibodies diluted in blocking solution were applied on tissue sections and incubated in a dark humid chamber for 1 hour at room temperature. Slides were rinsed three times for 5 minutes each with PBS. Nuclei were stained with Hoechst 33258 and then mounted with Vectashield® mounting medium. Slides were imaged with a Leica Thunder microscope imaging system.

Duo color FISH:

The Oligopaint FISH probe libraries were constructed as described previously (Xie et al., Nature Methods, 2020). A ssDNA oligo pool was ordered and synthesized as custom arrays from GenScript. Each oligo consists of a 30-37 nucleotide (nt) homology to the 30kb region, flanking the mBrd4-mNut1m fusion point from the mm10 genome assembly. Each Oligo was run through the OligoArray2.0 with the parameters -n 22 -D 1000 -l 32 -L 32 -g 52 -t 75 -T 85 -s 60 -x 60 -p 35 -P 80-m "GGGGG;CCCCC; TTTTT;AAAAA" according to algorithms developed from the Wu laboratory (https://oligopaints.hms.harvard.edu/). The mBrd4 and mNut1m oligo subpool consists of a unique set of primer pairs for PCR amplification, a 20 nt T7 promoter sequence for in vitro transcription, and a 20 nt region for reverse transcription. The mBrd4 oligo pool was amplified by GCGGGACGTAAGGGCAACCG and GCGTTGCGGTGCGATCTTCT. The mNut1m was amplified by TGATAACCCACGCACGGCTG and GACCCGGGCCACTAACACGA. Following PCR amplification, each Oligopaint probes were generated by in vitro transcription and reverse transcription, in which ssDNA oligos conjugated with ATTO565 and ATTO647 fluorophores were introduced during the reverse transcription step. The Oligopaint covered genomic regions (mm10) used in this study are listed below: mBrd4, chr17:32,210,998-32,244,827. mNut1m,

chr2:112,227,883-112,257,004. The oligo sequences for conjugating fluorophores during reverse transcription are: /5ATTO565N/AATTCGGCAGACCCGAATGC (mNut1m) and /5ATTO647N/CCTATCCCCTGTGTGCCTTG (mBrd4).

DNA FISH is performed on cryosections prepared as described above. Tissue slides were washed 2X with 1XPBS and permeabilized in 0.5% TritonX100 in 1XPBS for 60min in a Coplin Jar. After 2X wash in 1XPBS, slides were treated with 0.1M HCl for 5min, followed by 3X washes with 2XSSC and 30 min incubation in 2X SSC + 0.1% Tween20 (2XSSCT) + 50% (v/v) formamide (EMD Millipore, cat#S4117). For each sample, we prepare a 25ul hybridization mixture containing 2XSSCT+ 50% formamide +10% Dextran sulfate (EMD Millipore, cat#S4030) supplemented with 0.5µl 10mg/mL RNaseA (Thermo Fisher Scientific, cat# 12091021) +0.5µl 10mg/mL salmon sperm DNA (Thermo Fisher Scientific, cat# 15632011) and 20pmol probes with distinct fluorophores. The probe mixture was thoroughly mixed by vortexing and briefly microcentrifuged. The hybridization mix was transferred directly onto the slide covered with a clean coverslip. The coverslip was sealed onto the slide by adding a layer of rubber cement (Fixo gum, Marabu) around the edges. Each slide was denatured at 80°C for 5 min, then transferred to a humidified hybridization chamber and incubated at 42°C for 16 hours in a heated incubator. After hybridization, samples were washed 2X for 15 minutes in prewarmed 2XSSCT at 60 °C and then were further incubated at 2XSSCT for 10min at RT, at 0.2XSSC for 10min at RT, at 1XPBS for 2X5min with DNA counterstaining with DAPI. Then coverslips were mounted on slides with Prolong Diamond Antifade Mountant (Thermo Fisher Scientific Cat#P36961) for imaging acquisition.

DNA FISH images were acquired using a structured illumination microscopy on a Zeiss AxioObserver microscope (Zeiss Elyra7). Images were taken with a Plan-Apochromat 63x/1.40 oil DIC objective in a lens immersion medium having a refractive index of 1.515 and two pco.edge 4.2 CL HS sCMOS cameras. We used 405nm (Excitation wavelength) and 460nm (Emission wavelength) for the DAPI channel, 561nm (Excitation wavelength) and 579nm (Emission wavelength) for the ATTO565 channel,

and 633nm (Excitation wavelength) and 654nm (Emission wavelength) for the ATTO647 channel. We acquired 13 phase images for each focal plane. Image post-processing was performed with the ZEN 3.0 SR (black) software.

Software and statistics:

The tumor growth and survival data were analyzed and visualized using Prism 9 software (GraphPad, USA). Animal sacrifice was considered the endpoint for both analyses. The median survival times were calculated utilizing the Kaplan-Meier analysis, and significance values were determined by conducting the log-rank test. A p-value of less than 0.05 defined statistical significance.

Author contribution:

D.Z: formal analysis, investigation, methodology, validation, visualization, writing review, and editing;

A.E: formal analysis, investigation, methodology, validation, visualization, writing review and editing;

M.B: methodology and visualization for FISH

L.X: methodology and visualization for FISH

M.T: histopathological and IHC analysis, writing review and editing;

B.G: conceptualization, design and generation of NC GEMM, formal analysis, investigation, methodology, validation, visualization, writing original draft, writing review and editing; providing funding support;

Funding: This research is supported by an R37 grant (R37CA269076) from the National Cancer Institute (NCI) and startup funding both awarded to B.G.

Acknowledgment:

We want to thank MSU Transgenic and Genome Editing Facility (TGEF) for their help in generating the NC GEMM and the Histology laboratory at Veterinary Diagnostic Laboratory (Taylor Vaughn) for their help in histopathology and IHC analysis. We would

also like to thank MSU Precision Health Program Tissue Analysis Core for assistance in slide scanning.

References:

1 Kubonishi, I. *et al.* Novel t(15;19)(q15;p13) chromosome abnormality in a thymic carcinoma. *Cancer Res* **51**, 3327-3328 (1991).

2 French, C. A. Pathogenesis of NUT midline carcinoma. *Annu Rev Pathol* **7**, 247-265, doi:10.1146/annurev-pathol-011811-132438 (2012).

3 Chau, N. G. *et al.* An Anatomical Site and Genetic-Based Prognostic Model for Patients With Nuclear Protein in Testis (NUT) Midline Carcinoma: Analysis of 124 Patients. *JNCI Cancer Spectr* **4**, pkz094, doi:10.1093/jncics/pkz094 (2020).

4 Nishimura, Y., Ryo, E., Yamazaki, N., Yatabe, Y. & Mori, T. Cutaneous Primary NUT Carcinoma With BRD3-NUTM1 Fusion. *Am J Surg Pathol* **45**, 1582-1584, doi:10.1097/PAS.0000000000001801 (2021).

5 Shehata, B. M. *et al.* NUT midline carcinoma in a newborn with multiorgan disseminated tumor and a 2-year-old with a pancreatic/hepatic primary. *Pediatr Dev Pathol* **13**, 481-485, doi:10.2350/09-10-0727-CR.1 (2010).

6 Ball, A. *et al.* A rare case of NUT midline carcinoma. *Gynecol Oncol Case Rep* **3**, 1-3, doi:10.1016/j.gynor.2012.09.004 (2012).

7 Jung, M., Kim, S. I., Kim, J. W., Jeon, Y. K. & Lee, C. NUT Carcinoma in the Pelvic Cavity With Unusual Pathologic Features. *Int J Gynecol Pathol* **41**, 292-297, doi:10.1097/PGP.0000000000000801 (2022).

8 Bauer, D. E. *et al.* Clinicopathologic features and long-term outcomes of NUT midline carcinoma. *Clin Cancer Res* **18**, 5773-5779, doi:10.1158/1078-0432.CCR-12-1153 (2012).

9 French, C. A. NUT Carcinoma: Clinicopathologic features, pathogenesis, and treatment. *Pathol Int* **68**, 583-595, doi:10.1111/pin.12727 (2018).

10 Lauer, U. M., Hinterleitner, M., Horger, M., Ohnesorge, P. V. & Zender, L. NUT Carcinoma-An Underdiagnosed Malignancy. *Front Oncol* **12**, 914031, doi:10.3389/fonc.2022.914031 (2022).

11 Kees, U. R., Mulcahy, M. T. & Willoughby, M. L. Intrathoracic carcinoma in an 11-year-old girl showing a translocation t(15;19). *Am J Pediatr Hematol Oncol* **13**, 459-464, doi:10.1097/00043426-199124000-00011 (1991).

12 French, C. A. *et al.* Midline carcinoma of children and young adults with NUT rearrangement. *J Clin Oncol* **22**, 4135-4139, doi:10.1200/JCO.2004.02.107 (2004).

13 French, C. A. *et al.* BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. *Cancer Res* **63**, 304-307 (2003).

14 Chien, Y. W. *et al.* Primary malignant epithelioid and rhabdoid tumor of bone harboring ZNF532-NUTM1 fusion: the expanding NUT cancer family. *Genes Chromosomes Cancer* **58**, 809-814, doi:10.1002/gcc.22785 (2019).

15 French, C. A. *et al.* NSD3-NUT fusion oncoprotein in NUT midline carcinoma: implications for a novel oncogenic mechanism. *Cancer Discov* **4**, 928-941, doi:10.1158/2159-8290.CD-14-0014 (2014).

16 He, M. *et al.* Tumor Mutation Burden and Checkpoint Immunotherapy Markers in NUT Midline Carcinoma. *Appl Immunohistochem Mol Morphol* **28**, 495-500, doi:10.1097/PAI.0000000000000781 (2020).
17 Lemelle, L. *et al.* NUT carcinoma in children and adults: A multicenter retrospective study. *Pediatr Blood Cancer* **64**, doi:10.1002/pbc.26693 (2017).
18 Shiota, H. *et al.* Nut Directs p300-Dependent, Genome-Wide H4 Hyperacetylation in Male Germ Cells. *Cell Rep* **24**, 3477-3487 e3476, doi:10.1016/j.celrep.2018.08.069 (2018).
19 Alekseyenko, A. A. *et al.* The oncogenic BRD4-NUT chromatin regulator drives aberrant transcription within large topological domains. *Genes Dev* **29**, 1507-1523, doi:10.1101/gad.267583.115 (2015).
20 Reynoird, N. *et al.* Oncogenesis by sequestration of CBP/p300 in transcriptionally inactive hyperacetylated chromatin domains. *EMBO J* **29**, 2943-2952, doi:10.1038/emboj.2010.176 (2010).
21 French, C. A. *et al.* BRD-NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* **27**, 2237-2242, doi:10.1038/sj.onc.1210852 (2008).
22 Grayson, A. R. *et al.* MYC, a downstream target of BRD-NUT, is necessary and sufficient for the blockade of differentiation in NUT midline carcinoma. *Oncogene* **33**, 1736-1742, doi:10.1038/onc.2013.126 (2014).
23 Wang, R. *et al.* Activation of SOX2 expression by BRD4-NUT oncogenic fusion drives neoplastic transformation in NUT midline carcinoma. *Cancer Res* **74**, 3332-3343, doi:10.1158/0008-5472.CAN-13-2658 (2014).
24 Ameratunga, M. *et al.* First-in-human Phase 1 open label study of the BET inhibitor ODM-207 in patients with selected solid tumours. *Br J Cancer* **123**, 1730-1736, doi:10.1038/s41416-020-01077-z (2020).
25 Zhang, X. *et al.* Therapeutic targeting of p300/CBP HAT domain for the treatment of NUT midline carcinoma. *Oncogene* **39**, 4770-4779, doi:10.1038/s41388-020-1301-9 (2020).
26 Morrison-Smith, C. D. *et al.* Combined Targeting of the BRD4-NUT-p300 Axis in NUT Midline Carcinoma by Dual Selective Bromodomain Inhibitor, NEO2734. *Mol Cancer Ther* **19**, 1406-1414, doi:10.1158/1535-7163.MCT-20-0087 (2020).
27 Haruki, N. *et al.* Cloned fusion product from a rare t(15;19)(q13.2;p13.1) inhibit S phase in vitro. *J Med Genet* **42**, 558-564, doi:10.1136/jmg.2004.029686 (2005).
28 Svoboda, P. Mammalian zygotic genome activation. *Semin Cell Dev Biol* **84**, 118-126, doi:10.1016/j.semcdb.2017.12.006 (2018).
29 Anderson, M. S. *et al.* Projection of an immunological self shadow within the thymus by the aire protein. *Science* **298**, 1395-1401, doi:10.1126/science.1075958 (2002).
30 Chen, X., Pappo, A. & Dyer, M. A. Pediatric solid tumor genomics and developmental pliancy. *Oncogene* **34**, 5207-5215, doi:10.1038/onc.2014.474 (2015).
31 Puisieux, A., Pommier, R. M., Morel, A. P. & Lavial, F. Cellular Pliancy and the Multistep Process of Tumorigenesis. *Cancer Cell* **33**, 164-172, doi:10.1016/j.ccell.2018.01.007 (2018).
32 Gu, B. & Hakun, M. C. Challenges and Opportunities in NUT Carcinoma Research. *Genes* **12**, 235 (2021).

33 Minas, T. Z. *et al.* Combined experience of six independent laboratories attempting to create an Ewing sarcoma mouse model. *Oncotarget* **8**, 34141-34163, doi:10.18632/oncotarget.9388 (2017).
34 Amin, A. D., Rajan, S. S., Groysman, M. J., Pongtornpipat, P. & Schatz, J. H. Oncogene Overdose: Too Much of a Bad Thing for Oncogene-Addicted Cancer Cells. *Biomark Cancer* **7**, 25-32, doi:10.4137/BIC.S29326 (2015).
35 Van Deursen, J., Fornerod, M., Van Rees, B. & Grosveld, G. Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc Natl Acad Sci U S A* **92**, 7376-7380, doi:10.1073/pnas.92.16.7376 (1995).
36 van der Weyden, L. & Bradley, A. Mouse chromosome engineering for modeling human disease. *Annu Rev Genomics Hum Genet* **7**, 247-276, doi:10.1146/annurev.genom.7.080505.115741 (2006).
37 Drynan, L. F. *et al.* Mll fusions generated by Cre-loxP-mediated de novo translocations can induce lineage reassignment in tumorigenesis. *EMBO J* **24**, 3136-3146, doi:10.1038/sj.emboj.7600760 (2005).
38 Patel, M. R. *et al.* Longitudinal, noninvasive imaging of T-cell effector function and proliferation in living subjects. *Cancer Res* **70**, 10141-10149, doi:10.1158/0008-5472.CAN-10-1843 (2010).
39 Arnold, K. *et al.* Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. *Cell Stem Cell* **9**, 317-329, doi:10.1016/j.stem.2011.09.001 (2011).
40 Senoo, M., Pinto, F., Crum, C. P. & McKeon, F. p63 Is essential for the proliferative potential of stem cells in stratified epithelia. *Cell* **129**, 523-536, doi:10.1016/j.cell.2007.02.045 (2007).
41 Murphy, M. J., Wilson, A. & Trumpp, A. More than just proliferation: Myc function in stem cells. *Trends Cell Biol* **15**, 128-137, doi:10.1016/j.tcb.2005.01.008 (2005).
42 Dassule, H. R., Lewis, P., Bei, M., Maas, R. & McMahon, A. P. Sonic hedgehog regulates growth and morphogenesis of the tooth. *Development* **127**, 4775-4785, doi:10.1242/dev.127.22.4775 (2000).
43 Haack, H. *et al.* Diagnosis of NUT midline carcinoma using a NUT-specific monoclonal antibody. *Am J Surg Pathol* **33**, 984-991, doi:10.1097/PAS.0b013e318198d666 (2009).
44 Bishop, J. A. & Westra, W. H. NUT midline carcinomas of the sinonasal tract. *Am J Surg Pathol* **36**, 1216-1221, doi:10.1097/PAS.0b013e318254ce54 (2012).
45 Carter, T. *et al.* Incidence of NUT carcinoma in Western Australia from 1989 to 2014: a review of pediatric and adolescent cases from Perth Children's Hospital. *BMC Cancer* **21**, 740, doi:10.1186/s12885-021-08432-0 (2021).
46 Eagen, K. P. & French, C. A. Supercharging BRD4 with NUT in carcinoma. *Oncogene* **40**, 1396-1408, doi:10.1038/s41388-020-01625-0 (2021).
47 Basil, M. C. *et al.* The Cellular and Physiological Basis for Lung Repair and Regeneration: Past, Present, and Future. *Cell Stem Cell* **26**, 482-502, doi:10.1016/j.stem.2020.03.009 (2020).
48 Shih, H. P., Wang, A. & Sander, M. Pancreas organogenesis: from lineage determination to morphogenesis. *Annu Rev Cell Dev Biol* **29**, 81-105, doi:10.1146/annurev-cellbio-101512-122405 (2013).

49 Marusyk, A., Janiszewska, M. & Polyak, K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell* **37**, 471-484, doi:10.1016/j.ccell.2020.03.007 (2020).
50 Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108**, 479-485, doi:10.1038/bjc.2012.581 (2013).
51 Alekseyenko, A. A. *et al.* Cell state-dependent chromatin targeting in NUT carcinoma. *Genetics* **224**, doi:10.1093/genetics/iyad083 (2023).
52 Zhou, J. *et al.* Primary Thyroid NUT Carcinoma With High PD-L1 Expression and Novel Massive IGKV Gene Fusions: A Case Report With Treatment Implications and Literature Review. *Front Oncol* **11**, 778296, doi:10.3389/fonc.2021.778296 (2021).
53 Gu, B., Posfai, E. & Rossant, J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat Biotechnol* **36**, 632-637, doi:10.1038/nbt.4166 (2018).
54 Gu, B., Posfai, E., Gertsenstein, M. & Rossant, J. Efficient Generation of Large-Fragment Knock-In Mouse Models Using 2-Cell (2C)-Homologous Recombination (HR)-CRISPR. *Curr Protoc Mouse Biol* **10**, e67, doi:10.1002/cpmo.67 (2020).
55 Gu, B., Gertsenstein, M. & Posfai, E. Generation of Large Fragment Knock-In Mouse Models by Microinjecting into 2-Cell Stage Embryos. *Methods Mol Biol* **2066**, 89-100, doi:10.1007/978-1-4939-9837-1_7 (2020).
56 Stelow, E. B. *et al.* NUT rearrangement in undifferentiated carcinomas of the upper aerodigestive tract. *Am J Surg Pathol* **32**, 828-834, doi:10.1097/PAS.0b013e31815a3900 (2008).

Figures and legends:

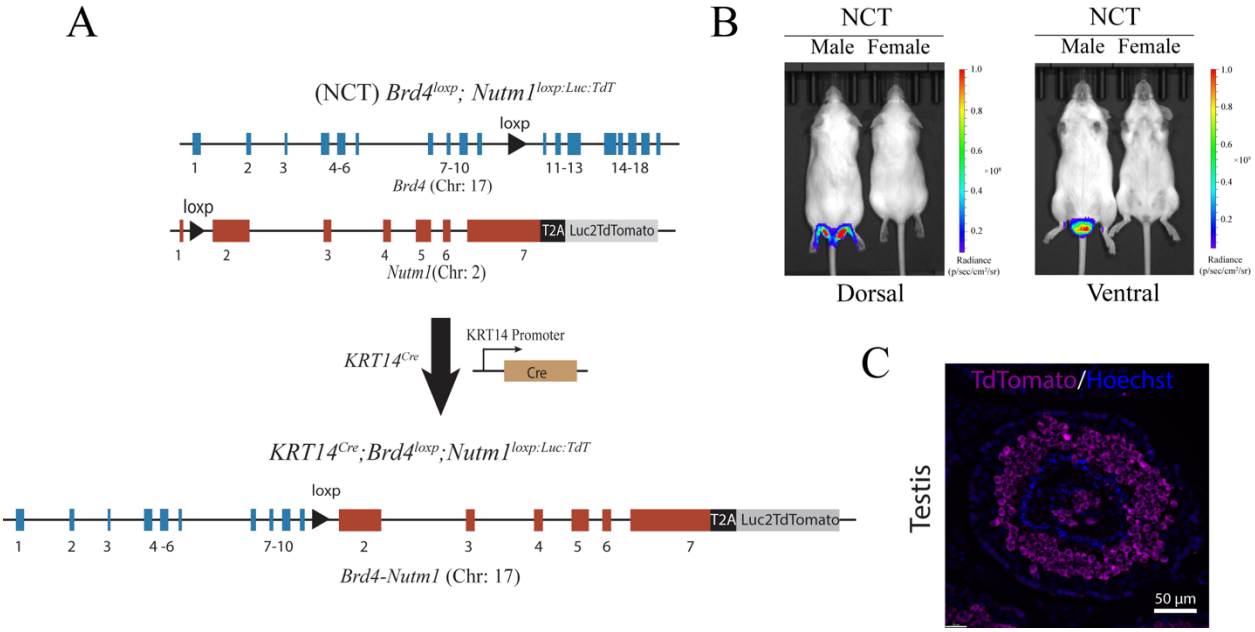


Figure 1. Design and validation of the NC translocator (NCT) mouse line. A. Schematic design for the NCT mouse line and the induction of NC through cre mediated recombination. B. Testis-specific bioluminescent (BLI) signals in male NCT mice. C. Immunofluorescent staining of the TdTomato reporter in NCT testis.

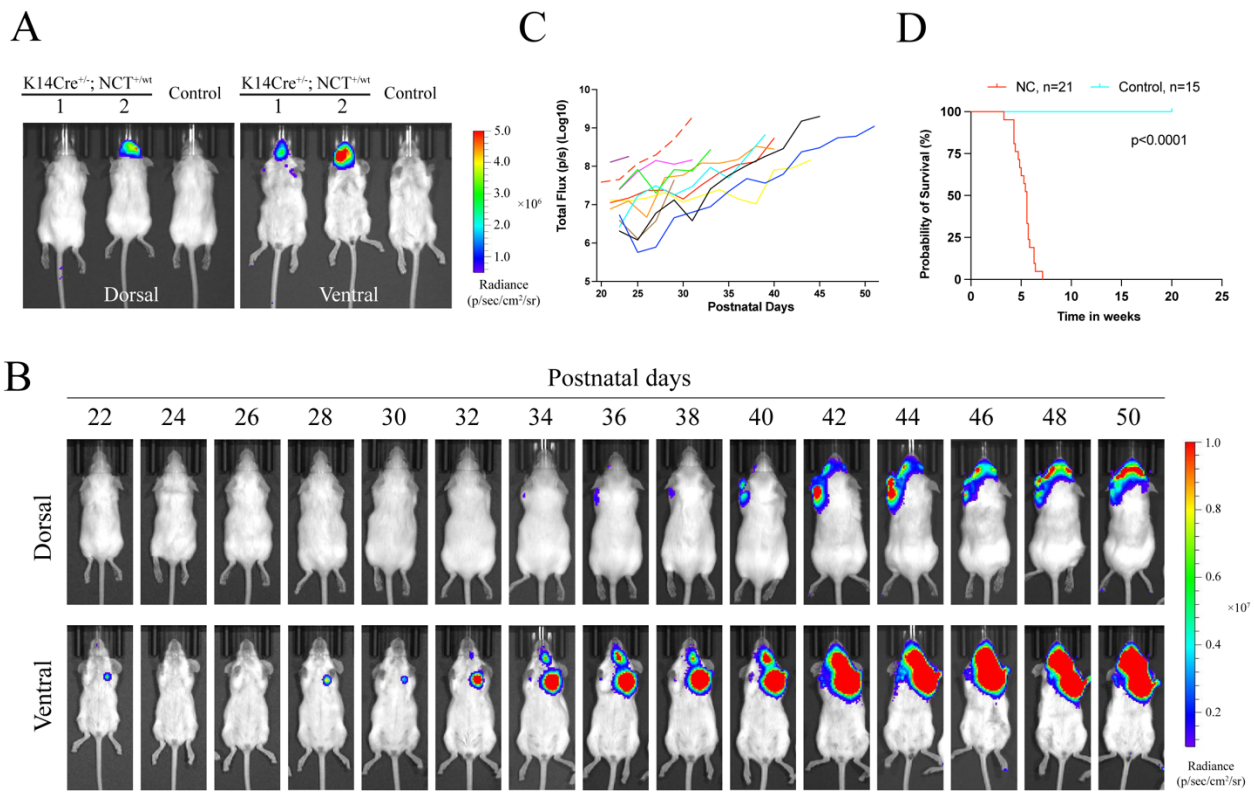


Figure 2: Tumor development in NC GEMM. A. BLI signal highlighting a head tumor in an NC mouse at weaning (postnatal day 21). B. A representative example of tumor development over time in an NC mouse tracked by BLI signals. C. Quantitative curves of BLI signals representing tumor growth in mouse head; each curve represents one tumor. D. Kaplan-Meier survival curve of control and tumor mice.

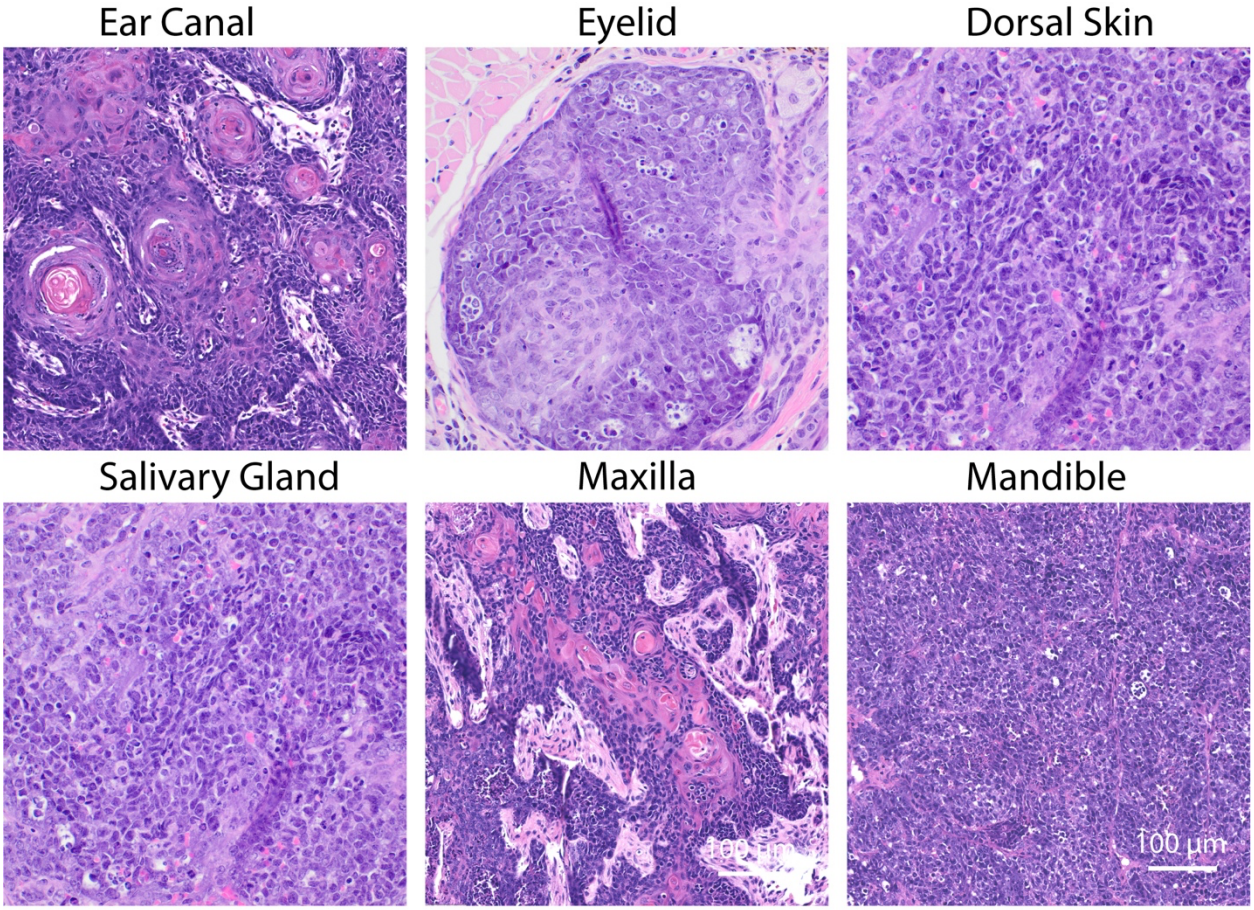


Figure 3. Representative histology characterization (H&E staining) of mouse NC developed at various anatomical sites.

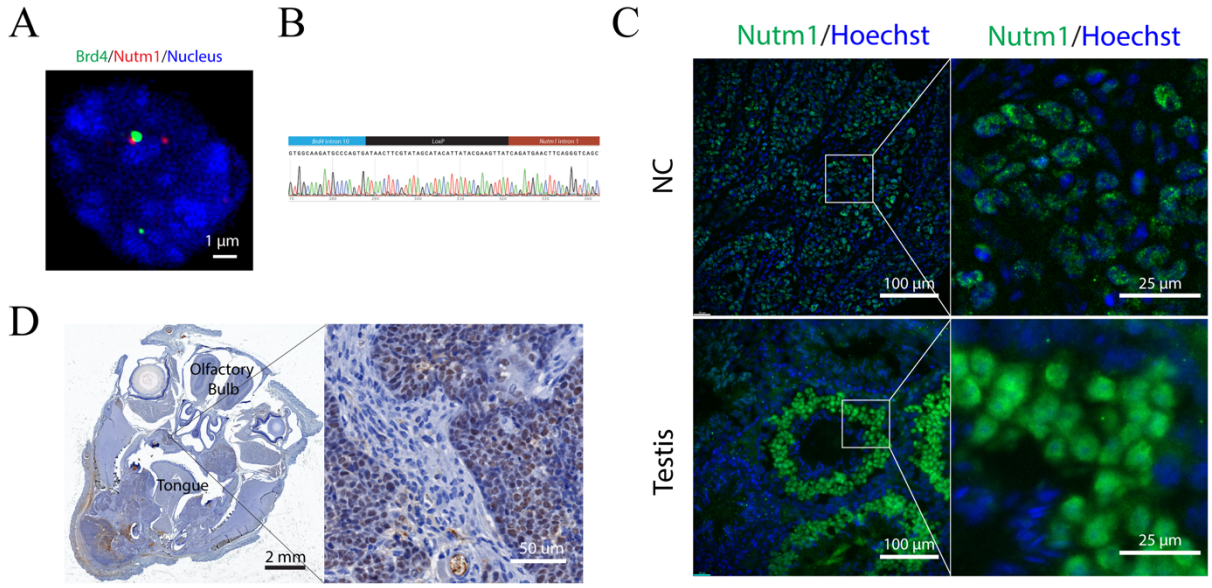


Figure 4. Cre-loxp-mediated chromosome translocation induced the formation and expression of the BRD4-NUTM1 fusion gene. A. Dual-color FISH analysis of the mouse *Brd4* locus (green) and *Nutm1* (red) in mouse NC cells. A representative picture of 20 cells. B. Sanger sequencing track showing the fusion junction within the Brd4 and Nutm1 introns. C. Immunofluorescence staining of the NUTM1 protein (representing the BRD4-NUTM1 fusion protein in NC) in mouse testis and NC tissues. D. Immunohistochemistry staining of the NUTM1 protein (representing the BRD4-NUTM1 fusion protein) in mouse NC. Representative staining of two NC bearing heads.

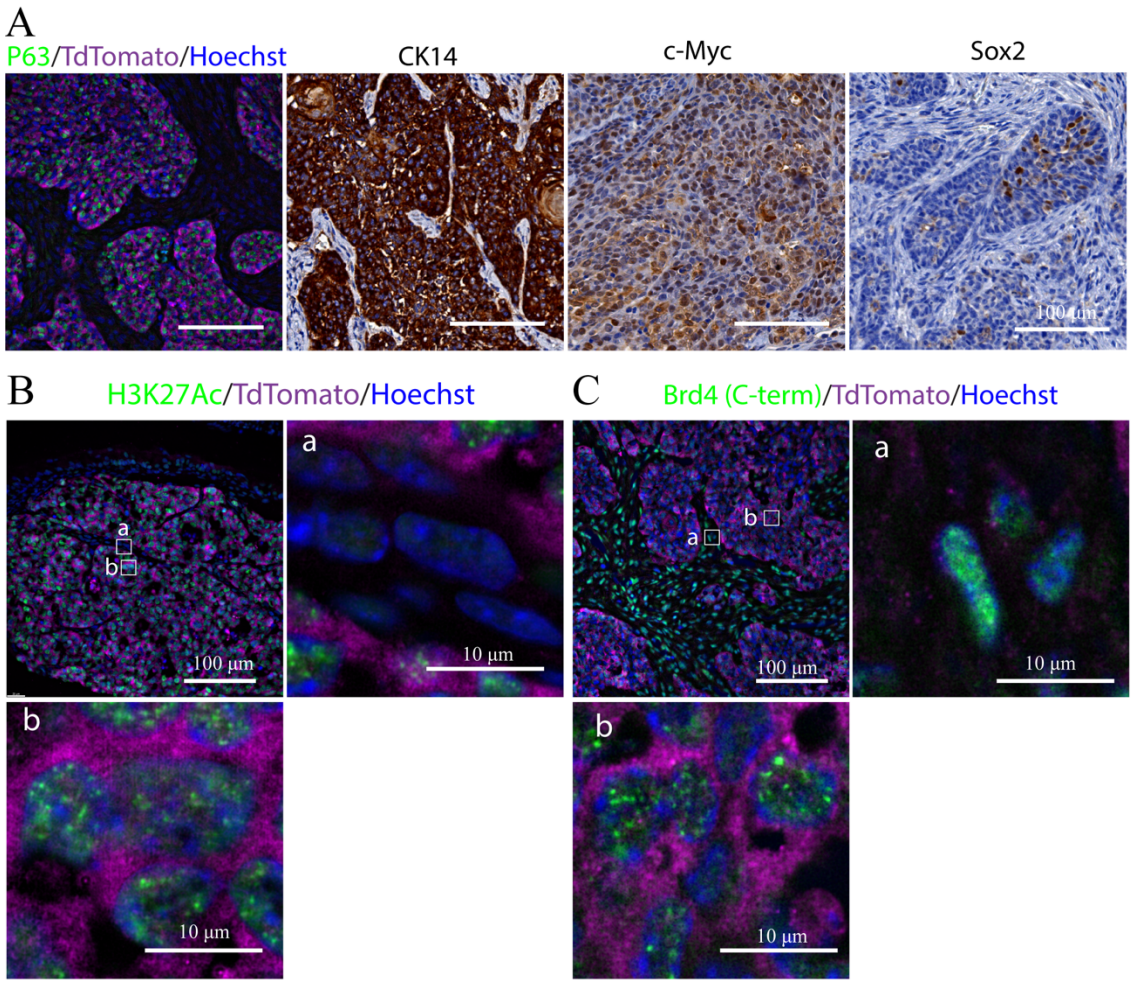


Figure 5: Molecular features of mouse NC: A. IF and IHC analysis of basal progenitor markers P63, K14, Sox2 and cMYC proteins in mouse NC. The TdTomato stains in IF demarcate NC cells from stromal cells. Pictures are representative images of at least three independent stains on different tumors. B. IF staining of H3K27Ac in mouse NC. The TdTomato stains in IF demarcate NC cells from stromal cells. a. higher magnification of stromal cells; b. high magnification of NC cells. C. IF staining of Brd4 in mouse NC. The TdTomato stains in IF demarcate NC cells from stromal cells. a. higher magnification of stromal cells; b. high magnification of NC cells.

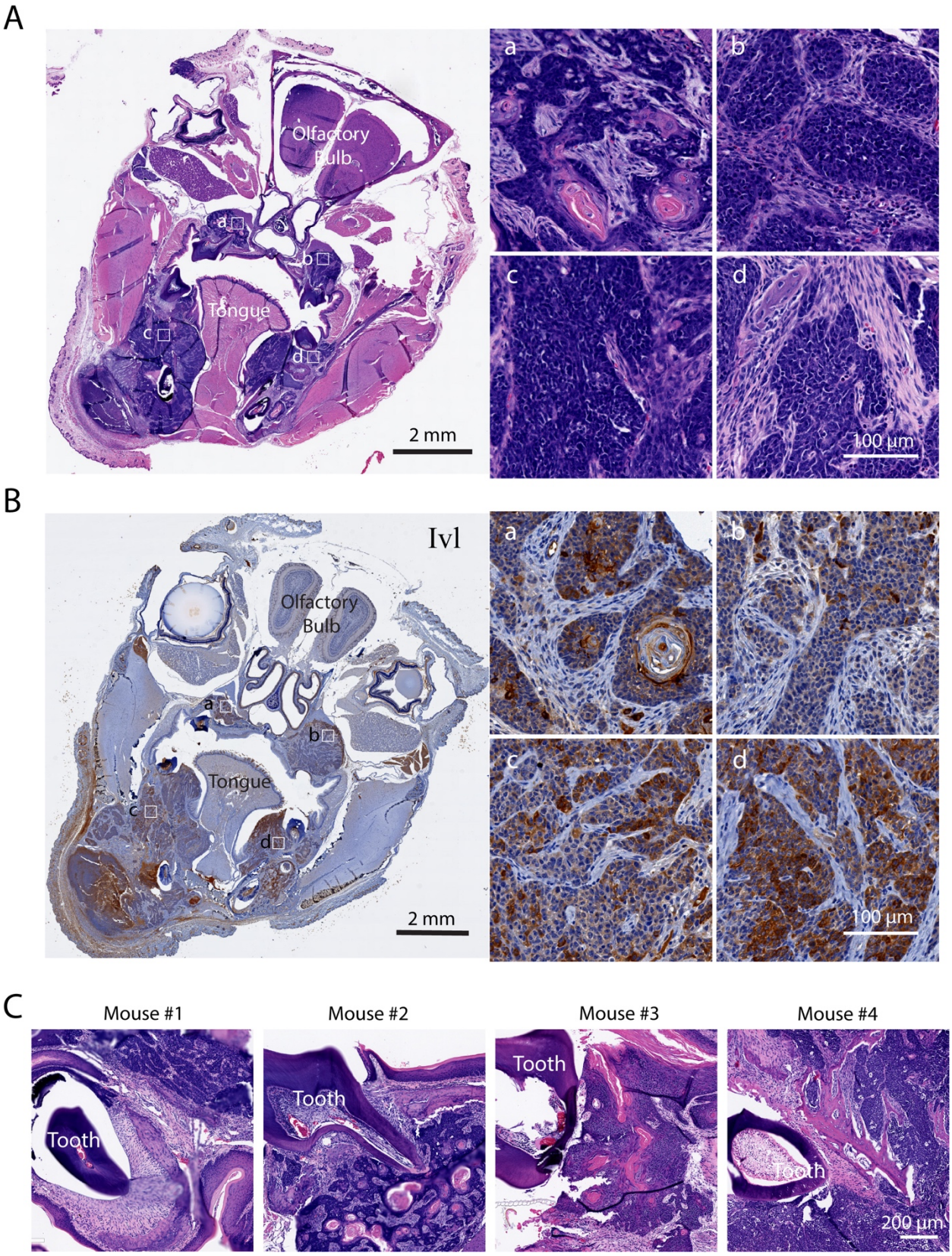


Figure 6. Histopathological heterogeneity of mouse NC. A. Histopathological characterization (H&E) of multiple tumors in the same NC bearing head. B. IHC of Ivl

proteins on the same NC bearing head. C. Histopathologic features (H&E) of tumors arising within the periodontal ligament in 4 different mice.

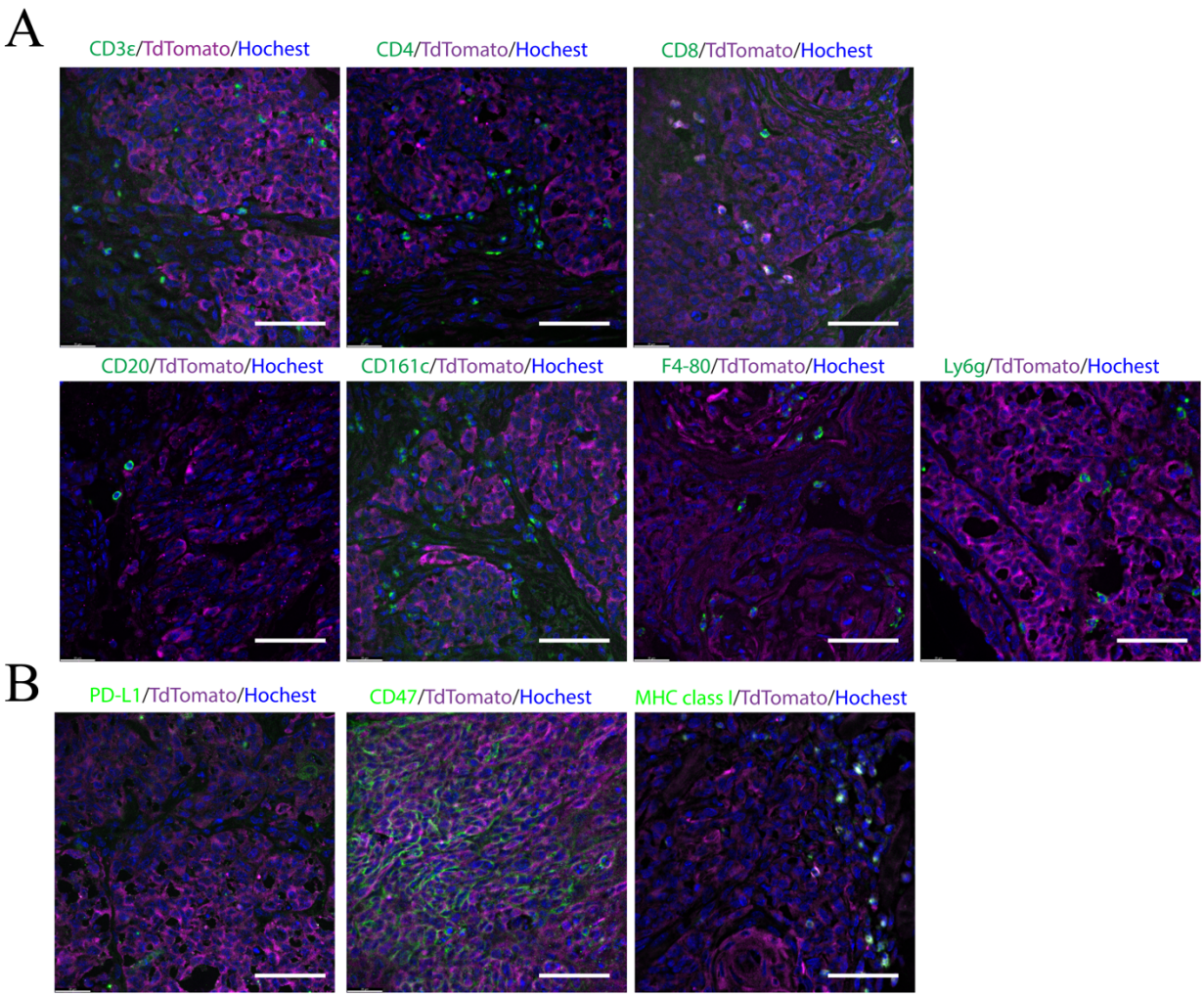
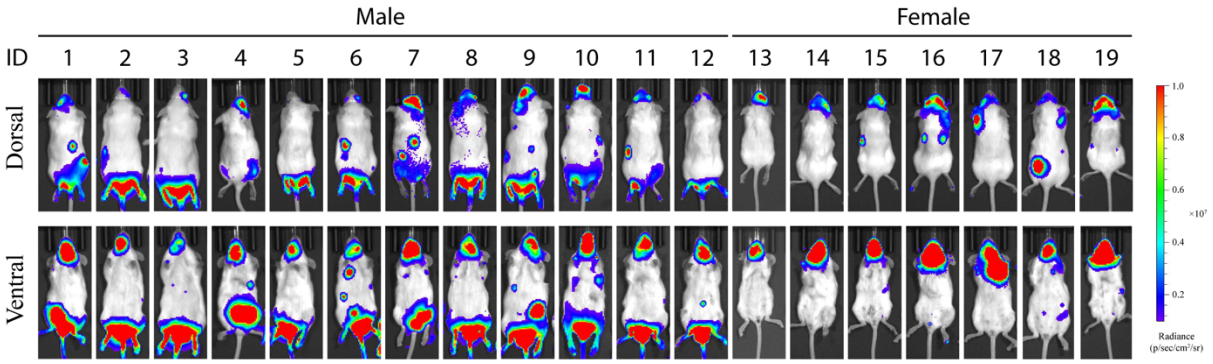


Figure 7. Limited immune infiltration in mouse NC. A. IF staining of markers of immune cells on NC tissues. The TdTomato stain in IF demarcate NC cells from stromal cells. Pictures are representative images of at least three independent stains on different tumors. B. IF staining of immune checkpoint proteins and MHC-Class I on NC tissues. The TdTomato stain in IF demarcate NC cells from stromal cells. Pictures are representative images of at least three independent stains on different tumors.



Supplementary Figure 1: BLI image of 19 NC mice at humane endpoint.

Tumor Distribution (n=27)

Maxilla	Mandible	Skin	Salivary Gland	Eyelids	Tongue	Ear Canal
13	20	16	2	2	2	1

Supplementary Table 1: Anatomical locations of NC tumor in mice.