Daxx maintains endogenous retroviral silencing and restricts cellular plasticity in vivo

Amanda R. Wasylshen1*, Chang Sun1,2†, Sydney M. Moyer1,2†, Yuan Qi3†, Gilda P. Chau††, Neeraj K. Aryal1,2, Florence McAllister4, Michael P. Kim5, Michelle C. Barton2,6, Jeannelyn S. Estrella7, Xiaoping Su3, Guillermina Lozano1,2*.

Tumor sequencing studies have emphasized the role of epigenetics and altered chromatin homeostasis in cancer. Mutations in DAXX, which encodes a chaperone for the histone 3.3 variant, occur in 25% of pancreatic neuroendocrine tumors (PanNETs). To advance our understanding of physiological functions of Daxx, we developed a conditional Daxx allele in mice. We demonstrate that Daxx loss is well tolerated in the pancreas but creates a permissive transcriptional state that cooperates with environmental stress (inflammation) and other genetic lesions (Men1 loss) to alter gene expression and cell state, impairing pancreas recovery from inflammatory stress in vivo. The transcriptional changes are associated with dysregulation of endogenous retroviral elements (ERVVs), and dysregulation of endogenous genes near ERVs is also observed in human PanNETs with DAXX mutations. Our results reveal a physiologic function of DAXX, provide a mechanism associated with impaired tissue regeneration and tumorigenesis, and expand our understanding of ERV regulation in somatic cells.

INTRODUCTION

One of the major findings from cancer genome sequencing projects was the large proportion of genetic mutations in epigenetic regulators found across broad cancer types. This includes enzymes involved in DNA methylation and histone modifications as well as chromatin remodelers (1, 2). One of the most unexpected discoveries was recurrent mutations in genes that encode histones. Missense mutations in the histone 3 variant H3.3 (encoded by the H3F3A and H3F3B genes) were first identified in the invariably fatal pediatric malignancy diffuse intrinsic pontine glioma and have subsequently been identified in several other cancers (3, 4). Moreover, the protein chaperones that are responsible for loading H3.3 into chromatin, DAXX and ATRX, are also mutated in cancers. Mutually exclusive loss-of-function mutations, predominantly indels, in DAXX or ATRX occur in 43% of pancreatic neuroendocrine tumors (PanNETs) (5). This high frequency of mutations suggests an important tumor suppressor role for this histone chaperone complex and, more generally, for its epigenetic regulatory axis. Biochemical and cell-based studies demonstrate that DAXX and ATRX mediate H3.3 deposition into heterochromatic regions of the DNA, including telomeres and pericentromeric DNA (6). Notably, all PanNETs with DAXX or ATRX mutations activate the alternative lengthening of telomeres pathway, implicating telomere dysfunction in tumorigenesis (7). Additional studies have demonstrated that DAXX, ATRX, and H3.3 cooperate with histone methylation to mediate the epigenetic silencing of transposable elements in the genome of mouse embryonic stem cells (8). Depression of endogenous retroviral elements (ERVVs) can lead to active transposition and affect the expression of endogenous genes through their strong long terminal repeat (LTR) promoters; however, our understanding of the mechanisms that maintain silencing in somatic tissues as well as how dysregulation affects normal tissue function and contributes to disease remains limited. Several other and often opposing functions have also been suggested for Daxx, including roles in promoting apoptosis, inhibiting apoptosis and affecting the DNA damage response (9), further emphasizing the need to evaluate Daxx function(s) in physiologically relevant models. Mouse models have been pivotal to understanding the physiologically relevant functions of genes and the molecular mechanisms downstream of their dysregulation that contribute to disease. Daxx deletion is lethal at an early embryonic stage in mice and embryos exhibit apoptosis, yet the reasons for this lethality remain poorly understood (10). We have therefore generated and characterized a conditional Daxx allele in mice, allowing for both spatial and temporal control of Daxx expression and the ability to investigate the essential biology downstream of Daxx loss. Using this model, we demonstrated that Daxx loss is well tolerated in the developing pancreas, with no robust phenotypic changes under homeostatic conditions and no impact on mouse survival. However, whole-transcriptome analysis indicated dysregulation of heterochromatin with Daxx loss. Because both genetics and environment interact to maintain tissue homeostasis, we combined Daxx deficiency with Men1 loss (a lesion that frequently co-occurs with DAXX mutation in human cancers), inflammatory stress, and the combination of the two. Comprehensive transcriptome and chromatin accessibility profiling revealed that Daxx loss leads to dysregulation of endogenous retrovirus silencing, with coordinate gene expression changes that are associated with altered cell state and impaired pancreas recovery and regeneration in vivo. An analysis of human PanNET RNA sequencing (RNA-seq) data provided supporting evidence of coordinate dysregulation of ERVs and protein-coding genes downstream of DAXX mutation.
RESULTS

\textit{Daxx} loss is well tolerated in the developing pancreas

The early lethality of \textit{Daxx}-null mice (10) requires a conditional \textit{Daxx} allele to study the effects of \textit{Daxx} loss in both a tissue- and temporal-specific manner. We previously identified a single-nucleotide deletion in the intron 2 loxP site in the available \textit{Daxx} \textit{tm2Led} allele, thereby rendering recombination inefficient in vivo (11). However, we were able to obtain a null allele (\textit{Daxx} \textit{Δ}3) from this mouse and confirmed the previously reported embryonic lethality, demonstrating that deletion of exon 3 does result in a null allele (11). To generate a functional conditional allele, we designed a CRISPR-Cas9 strategy to insert the missing nucleotide (fig. S1, A to C). We evaluated recombination of this new conditional allele (\textit{Daxx} \textit{fl}) compared with \textit{Daxx} \textit{tm2Led}, using the constitutive CMV-Cre \textit{Tg} driver. Robust induction of recombination was observed in vivo by polymerase chain reaction (PCR) genotyping of tail biopsies in \textit{Daxx} \textit{fl/+CMV-CreTg} mice compared with \textit{Daxx} \textit{tm2Led/+CMV-CreTg} mice (fig. S1D).

Because \textit{DAXX} mutations are almost exclusively present in PanNETs, we chose to evaluate the physiologic consequences of \textit{Daxx} loss in the pancreas. Moreover, the pancreas is a robustly regenerative organ with inherent cell plasticity (12), making it an ideal system to study the effects of epigenetic dysregulation. The pancreas is specified between E8.5 and E9.5 in mouse development, largely driven by the expression of the homeobox transcription factor \textit{Pdx1} (13). We therefore used the \textit{Pdx1-Cre} \textit{Tg} \textit{(C)} line to drive deletion of \textit{Daxx} throughout the pancreas (14). \textit{Daxx} loss was well tolerated because \textit{Daxx} \textit{fl/Pdx1-CreTg} (DC) mice were born at the expected Mendelian ratio. We euthanized a small number of mice at 6 to 8 weeks of age to comprehensively evaluate pancreas development. We first confirmed successful recombination of the \textit{Daxx} \textit{fl} allele through quantitative PCR genotyping analysis (fig. S1E).

Western blot analysis of total pancreas lysates (Fig. 1A) demonstrated loss of protein expression in DC pancreases. Pancreas weight, as a percentage of total mouse weight, was the same in DC mice as in wild-type (WT) controls (Fig. 1B). In addition, the pancreas was histologically normal (Fig. 1C). To evaluate the long-term consequences of \textit{Daxx} loss in the pancreas, we aged a cohort of mice and found no reduction in survival (Fig. 1D) or significant alterations of \(\beta\) cell function as measured by nonfasting blood glucose in 12-month-old mice (Fig. 1E). We obtained similar results using a \(\beta\) cell–specific rat insulin II promoter–driven Cre recombinase (\textit{RIP-CreTg}) to induce \textit{Daxx} deletion. Combined, these data suggest that \textit{Daxx} is not required for pancreas specification or normal function in vivo.
To evaluate the pancreas at a molecular level, we next conducted basal transcriptome analysis through RNA-seq in adult (6- to 8-week-old) DC mice compared with WT controls. This analysis revealed significant differential expression of a single gene, Bglap3, a largely unstudied osteocalcin-related gene (15) in the Daxx-deficient pancreas compared with controls (log2FC = 2.54, $P_{adj} = 1.12 \times 10^{-5}$; Fig. 1F). We confirmed this gene expression difference in independent mice; Bglap3 expression was an average of 10-fold higher in the DC mouse pancreas compared with controls (Fig. 1G). Bglap3 also emerged as a transcriptional readout of endogenous retrovirus dysregulation in multiple cell types because it contained an intragenic ERV, IAP-d-int (Fig. 1H). Deletion of components of the Setdb1 histone methyltransferase complex in mouse embryonic stem cells and fibroblasts leads to loss of H3K9me3 and robust overexpression of Bglap3 (16, 17). We further interrogated our RNA-seq data for expression of the IAP-d-int (18) and found a significant increase in expression of the ERV as well ($P = 0.03$, Wilcoxon test; Fig. 1I). Combined, these data suggest that Daxx loss may alter heterochromatin, which we speculate may create a permissive environment and cooperate with additional environmental stresses or genetic lesions to affect biology in vivo.

Daxx and Atrx interact to form the H3.3 chaperone complex. Correspondingly, Atrx loss is also well tolerated in the developing pancreas (fig. S1F) and is associated with variable but increased expression of Bglap3 (fig. S1G), suggesting that deregulation of the Bglap3 locus is associated with the histone chaperone function of Daxx and Atrx.

The combined loss of Daxx and Men1 leads to cystic degeneration of the exocrine pancreas

Genetic lesions cooperate to affect pancreas diseases. For example, Kras mutation alone is insufficient to drive robust cancer phenotypes in mice, despite occurring in most pancreatic ductal adenocarcinomas (19). However, when combined with loss of the p53 tumor suppressor, Kras mutation promotes rapid and aggressive disease (20). We therefore evaluated the effects of Daxx loss in the context of Men1 deficiency. Our choice to use Men1 loss as a cooperating genetic lesion was based on the high frequency of concurrent mutations in DAXX and MEN1 in human PanNETs (5, 21). In the endocrine pancreas, Men1 functions as part of a multiprotein mixed lineage leukemia (MLL) histone methyltransferase complex, mediating H3K4me3 and promoting the expression of important cell cycle inhibitors, Cdkn1b and Cdkn2c (22, 23).

To evaluate the long-term consequences of combined Daxx and Men1 loss throughout the pancreas, we aged a cohort of mice with the following genotypes: Pdx1-CreTg5 (C), Daxxβ/βPdx1-CreTg5 (DC), Men1β/βPdx1-CreTg5 (MC), and Daxxβ/βMen1β/βPdx1-CreTg5 (DMC; fig. S2A). The C and DC mice from this cohort were the same mice presented in Fig. 1D and showed no differences in survival. Consistent with previous reports, homozygous inactivation of Men1 led to islet cell hyperplasia and PanNET development in the endocrine pancreas, but no major phenotypic changes to the exocrine pancreas in mice aged up to 2 years (Fig. 2A, a and b) (24). However, the combined loss of Daxx and Men1 caused many changes to the exocrine pancreas, detected by histological analysis. Representative sections are provided in Fig. 2A (c to f). Specifically, there was evidence of acinar atrophy, including fatty replacement of acinar cells (Fig. 2A, c) and acinar-to-ductal metaplasia (Fig. 2A, e), suggesting cellular plasticity and an altered stress response. The metaplastic areas were composed of ductal proliferation, often accompanied by cyst formation (Fig. 2A, c and d), chronic inflammation, and surrounding fibrosis, the latter characteristic of activated stellate cells. In some cases, epithelial changes were seen, including partial or complete lining of cells exhibiting enlarged, hyperchromatic, pencilate nuclei. In one case, the cells lining the cysts (partially in some cysts and completely in others) had morphology reminiscent of signet ring cells (i.e., globoid cells with abundant eosinophilic cytoplasm pushing the nucleus to the periphery; Fig. 2A, f). We quantified the frequency of different lesions in the pancreas. The cystically dilated ducts found in DMC pancreases were absent in MC pancreases ($P < 0.0001$, chi-square test compared with MC; Fig. 2B). Metaplasias (Fig. 2A, b and e) were identified in most mice from both genotypes (Fig. 2C); however, the pancreas area occupied by these lesions was increased in DMC mice (Fig. 2D). Moreover, the proportion of mice with a metaplastic area >1% of the total pancreas area was significantly higher in DMC mice than in MC mice ($P < 0.0001$, chi-square test compared with MC; Fig. 2D). These data demonstrate that Daxx loss, in combination with Men1 deficiency, results in an impaired homeostatic response in the pancreas.

We speculated that the large endocrine tumor burden caused by Men1 deficiency (see below) induced tissue stress/injury, with the potential to obstruct vessels and ducts and induce damage and inflammation. Correspondingly, immune cell infiltration was significantly higher in DMC pancreases than in MC pancreases, as assayed by immunohistochemistry for the CD45 pan-immune marker (Fig. 2, E and F), indicative of an inflammatory response. These data uncovered an important role for Daxx in maintaining pancreas homeostasis in vivo.

This aged cohort also provided us with the opportunity to evaluate whether Daxx loss cooperates with Men1 deficiency to promote PanNETs in mice. Men1 loss, driven by several different pancreas Cre-recombinase lines, is sufficient on its own to promote multifocal hyperplasia and PanNET development, with a long disease latency (24–28). We observed no difference in overall survival between DMC and MC mice (fig. S2B), with mice from both genotypes developing well-differentiated PanNETs (fig. S2, C and D). Quantification of the endocrine compartment (islets) further revealed no significant changes in tumorigenesis (fig. S2E) and indicated that Daxx is not a strong endocrine tumor suppressor in mice.

Daxx loss cooperates with inflammation to affect the transcriptional state of the pancreas

Genetic alterations also cooperate with environmental stress to affect cellular and organisal responses. Experimental mouse models of inflammation and pancreatitis are well established (29, 30) and exhibit acinar-to-ductal metaplasia and immune cell infiltration, similar to the phenotypes that we observed in our aged cohort. Repeated intraperitoneal injection of the cholecytokinin analog caerulein is extensively used to study both acute and chronic pancreatitis. Caerulein induces acinar cells to locally secrete digestive enzymes, resulting in edema, infiltration of immune cells, and loss of identifiable acinar cells. The acinar cells undergo either cell death or a transient transdifferentiation to a metaplastic intermediate that expresses ductal markers. To induce chronic inflammation, we treated adult mice (6 to 10 weeks of age) with caerulein over 3 weeks, administered as four daily doses (hourly), 3 days per week (Monday, Wednesday, and Friday), as per established protocols (Fig. 3A) (31).
We included four different genotypes in these experiments: WT, DC, MC, and DMC mice. A small number of WT mice were euthanized 24 hours following the final caerulein dose to confirm pancreatitis (fig. S3, A and B). To evaluate the molecular response, we then euthanized mice 5 days following the final caerulein dose. In all genotypes, the pancreas weight (as a percentage of total mouse weight) was reduced in treated mice compared with untreated mice, but only the DMC pancreas was significantly reduced in size compared with WT-treated controls (fig. S3C), again suggesting cooperation between Daxx and Men1 loss. **P < 0.01, Student’s t test.

We next conducted global transcriptome analysis through RNA-seq of the total pancreas. We identified Daxx-dependent gene expression changes through comparisons of DC with WT and DMC with MC pancreases (Fig. 3B and table S1). What was immediately apparent was the much higher number of differentially expressed genes (DEGs) in caerulein-treated DC mice compared with WT mice, considering that the basal expression analysis identified only a single DEG. In addition, 80% of the DEGs were up-regulated in DC mice compared with WT mice, supporting the establishment of a more permissive chromatin state. We further focused on the comparison between DC and WT mice because it yielded a greater number of DEGs and is not complicated by Men1 loss. In addition, the DMC versus MC comparison identified too few genes for comprehensive pathway analysis. We conducted pathway analysis using the Molecular Signatures Database of Gene Set Enrichment Analysis and evaluated enrichments in the Gene Ontology biological processes. We identified pathways associated with tissue development, differentiation, and responses to stimuli (Fig. 3C). We then focused on the genes driving these pathways and found that several,
including transcription factors, are associated with specific cell populations in the pancreas according to single-cell transcriptome analysis (table S2) (33). Specific markers of pancreatic ductal cells, Mmp7 and Onecut2, were among the DEGs. In response to environmental or genetic stresses, acinar cells in the pancreas are known to undergo acinar-to-ductal metaplasia, and these data suggest that Daxx loss may contribute to establishing this transdifferentiated state. We also noted several genes associated with the mesenchymal cell types, which are not targeted by Pdx1-CreTg. The resident mesenchymal cell type in the pancreas is the stellate cell, which does not undergo recombination (34). Typically, these cells are quiescent, but they can be stimulated by cytokines to an activated state. Several established activators of stellate cells (Fgf1, Fgf2, and Tgfb3) are up-regulated in the DC pancreas, supporting paracrine activation of stellate cells. Once activated, stellate cells secrete factors that can promote growth and survival, as well as produce the fibrotic reaction that is characteristic of pancreatitis and pancreatic ductal adenocarcinoma. Single-cell sequencing defined a signature of activated stellate cells (35), and 6 of 11 of these genes were significantly up-regulated in our dataset (table S2). Ingenuity Pathway Analysis (36) of our DEGs identified hepatic fibrosis/hepatic stellate cell activation as the most significant pathway (Fig. 3D), and there are many similarities between pancreatic and hepatic stellate cells (34). Combined, these data demonstrated that, in the context of caerulein-induced pancreatitis, Daxx loss induced gene expression changes associated with pancreas cell state and suggested a role in restricting cellular plasticity.

Given the DEGs, and the potential paracrine activation of stellate cells in the pancreas, we hypothesized that the transcriptional changes and the potential cellular plasticity induced by Daxx loss could impair tissue repair and regeneration. We therefore repeated the chronic pancreatitis experiments, but this time, we allowed mice to recover for 4 weeks following the final dose. Pancreases from WT and DC mice completely recovered at the 4-week time point (Fig. 3, E and F). Pancreases from MC mice also recovered, despite the increased tissue damage in these mice. DMC mice, however, had a significant proportion of persistent metaplastic lesions in their pancreases. Metaplasia was confirmed by immunohistochemical staining for cytokeratin 19 (CK19) of pancreas sections 4 weeks following the final dose of caerulein. Image magnification is x20; scale bars, 100 μm. (F) Quantification of metaplastic areas as a percentage of total pancreas area (mean ± SD); ***P < 0.001, analysis of variance with Dunnett’s posttest comparing all groups with WT controls.

**Fig. 3. Daxx loss cooperates with inflammation and Men1 deficiency to affect the transcriptional state of the pancreas.** (A) Schematic representation of experimental outline. (B) Heatmaps representing Daxx-dependent gene expression changes in DC compared with WT mouse pancreases and in DMC compared with MC mouse pancreases following treatment with caerulein. (C) Gene Ontology (GO) biological processes pathway analysis of DEGs in DC mouse pancreases compared with WT controls. GSEA, Gene Set Enrichment Analysis. (D) Ingenuity Pathway Analysis of DEGs in DC mouse pancreases compared with WT controls. (E) Representative histologic images and immunohistochemistry for cytokeratin 19 (CK19) of pancreas sections 4 weeks following the final dose of caerulein. Image magnification is x20; scale bars, 100 μm. (F) Quantification of metaplastic areas as a percentage of total pancreas area (mean ± SD); ***P < 0.001, analysis of variance with Dunnett’s posttest comparing all groups with WT controls.
ATAC sequencing identifies chromatin accessibility changes associated with ERVs.

To date, the most important finding indicating the functional role for Daxx in vivo is the mutual exclusivity of DAXX and ATRX mutations in PanNETs, indicating that their shared function as a histone chaperone complex for H3.3 is important for tumor suppression. To understand the impact of Daxx loss on chromatin structure, we conducted ATAC sequencing (ATAC-seq) experiments because chromatin immunoprecipitation (ChIP) experiments for endogenous H3.3 have not provided sufficient signal for genome-wide ChIP sequencing analysis. A similar approach was recently taken to evaluate Atrx loss in glioma stem cells (37). Using established protocols for processing frozen tissues (38), we conducted ATAC-seq of pancreas samples from four MC mice and four DMC mice (two males and two females each) isolated 5 days following the final caerulein injection. We focused on these two genotypes because of the tissue repair and cell state defects (Fig. 3, E and F). The distribution of peaks across genomic regions was very similar between the two genotypes (Fig. 4A). We next identified 265 differentially accessible regions in the DMC pancreas compared with the MC pancreas, with 136 regions (51%) open and 129 (49%) closed (table S3). The relative proportions of open and closed regions were similar to what was reported for Atrx loss in glioma stem cells (37), although the number of regions that we identified was considerably smaller. As a positive control in our data, an 840–base pair (bp) region containing exon 3 of Daxx (region 140; table S3), which is deleted from the DMC pancreas and not from the MC pancreas, was identified as closed (false discovery rate = 0.01661). Overall, differentially accessible regions were enriched within 15 Mb of chromosome ends (P = 4.73 × 10⁻¹², Fisher’s exact test), consistent with the role of Daxx loading H3.3 into telomeres.

Because of the strong association between Daxx and H3.3 and heterochromatin, we next focused on differentially open regions in the DMC pancreas compared with the MC pancreas (table S3). The first and fourth most significant regions, 3.2 kb (false discovery rate = 1.08 × 10⁻⁸, Fig. 4B) and 2.0 kb (false discovery rate = 1.90 × 10⁻⁷), respectively, flank an approximately 5-kb region of chromosome 4qD2.2 containing the ERV RLTR4_MM-int with RLTR4_Mm LTRs (Fig. 4C). Daxx-mediated H3.3 deposition was previously linked to ERV silencing in mouse embryonic stem cells in vitro (8), and our data indicate that this function is conserved in somatic cells in vivo. In addition, the second most significant open region was just downstream...
of another RLTR4_Mm-int and an adjacent MuLV-int element, this
time contained in an intron of the Cep85 gene (Fig. 4D and table
S3). Five additional significantly open regions were also near these
ERV sequences (Fig. 4D). We next manually evaluated all open
regions for proximity to ERVs, and we found that 41 of 136 (30.1%) of
the open regions contained an ERV sequence, and an additional
23 of 136 (16.9%) were within 10 kb of ERVs (Fig. 4E and table S3).

Because ERV sequences occur in many, sometimes thousands,
copies in the genome, they present a computational challenge as
short sequencing reads of these repetitive regions cannot uniquely
map and are generally excluded from the analysis. This likely ex-
plains why our analysis captured the regions flanking the ERV, but
not the ERV itself. To address this limitation and directly evaluate
chromatin accessibility of transposable elements throughout the
genome, we assembled a mouse transposable element reference
genome with sequences representing all 219 elements annotated in
Dfam, an open collection of DNA transposable element sequence
alignments (39). These 219 elements collectively encompass ERV,
LINE (long interspersed nuclear element), SINE (short interspersed
nuclear element), and a small number of satellite repeats. We then
repeated the ATAC-seq analysis pipeline comparing caerulein-
treated DMC to MC pancreases, but this time, we mapped the reads
specifically to these transposable element sequences, allowing for a
genome-wide but not locus-specific analysis. In total, 213 peaks or
regions at least 200 bp in length were called in this analysis. Four
regions in ERVs were found to be differentially open in DMC pan-
creases (unadjusted $P < 0.05$; Fig. 4F), including one region in
MuLV-int that was highly significant (false discovery rate = 0.0060;
table S3 and Fig. 4F). Collectively, these ATAC-seq analyses dem-
strate that Daxx loss alters chromatin accessibility near ERVs in vivo.

**Daxx-dependent gene expression changes are associated with ERVs**

We next wanted to specifically investigate our transcriptome data
for associations with ERV dysregulation. To identify genes that
associate the strongest with Daxx loss in the caerulein-treated
pancreas, we compared the DEGs identified through the DC versus WT
and DMC versus MC analyses and identified nine common genes,significantly more than would be expected by chance (hypergeo-
metric $P = 9.88 \times 10^{-7}$; Fig. 5A and fig. S4). The differential expression
was significantly correlated for these nine genes (Pearson correla-
tion $R^2 = 0.9768$, $P < 0.0001$; Fig. 5B). The most significantly differen-
tially expressed of the overlapping genes was Bglap3, which showed
the most marked up-regulation with the combined loss of Daxx and
Men1 (Fig. 5C). IAP-d-int elements revealed an expression pattern
that paralleled Bglap3 (Fig. 5D), indicating coordinate regulation,
similar to what we observed in the basal context (Fig. 1, F to I) and
what has previously been reported following loss of H3K9me3 due to
either Zfp932 or Setdb1 depletion (17, 40). Of the other eight
genes, five showed a consistent pattern of up-regulation with loss of
Daxx: Ajuba, Neto1, Pcdh11x, Sult5a1, and Cd209a. All five of these
genes have an ERV, with retained internal proviral sequences, within
30 kb of the gene, and, in four of the five, the ERV is within only
10 kb (table S4).

To evaluate whether our DEG sets were globally enriched for
genes with nearby ERV elements, we compared our data with three
randomly generated gene sets (Random Gene Set Generator,
www.molbiotools.com) containing 100 genes. We manually anno-
tated ERVs with internal sequences up to 50 kb from the gene (table S4).
The proportion of genes with an annotated ERV was higher in our
DEGs compared to the randomly generated gene sets (Fig. 5E). This
was especially notable at very close distances to the gene (<5 kb).
Intragenic ERVs were present in 12.8% of the genes differentially
expressed when comparing DC mice with WT mice and 16.1% of
the genes differentially expressed between DMC and MC mice,
compared with only 3.3% of the genes in the randomly generated
genre sets. Our combined ATAC-seq and RNA-seq analyses strongly
indicate that dysregulation of ERVs occurs downstream of Daxx
loss, which subsequently results in robust gene expression changes
following inflammatory stress. Combined, these data demonstrate
that a major function of Daxx is to regulate the epigenome and that
dysregulation of ERVs leads to the coordinate dysregulation of
protein-coding genes in vivo.

Last, we globally interrogated our transcriptome data for differ-
ential expression of transposable elements. As in our analysis of
the ATAC-seq data, we used the assembled list of transposable elements
from Dfam to map RNA-seq reads, thereby giving a global (not
locus-specific) quantification of transposable element expression.
We evaluated transposable element expression in DC versus WT
and DMC versus MC mice, focusing on elements with a mean nor-
malized read count in WT mice greater than 10 (Fig. 5, F and G, and
table S5). The expression of a subset of ERVs was increased by Daxx
loss, whereas LINE and SINE expression remained unchanged,
indicating that this dysregulation is specific to ERVs. In the DC
pancreas compared with WT, there were three ERVs with increased
expression following Daxx loss (IAP-d, RLTR10D, and RLTR6-int,
unadjusted $P < 0.05$; Fig. 5, D, F, and H to J). Consistently, Obp2b
was the most significant DEG in DC mice treated with caerulein
(log2FC = 3.87, $P_{adj} = 2.07 \times 10^{-15}$), and it had an RLTR6-int element
less than 10 kb upstream of the gene. The number of ERVs up-
regulated with Daxx loss was even higher in the DMC versus MC
comparison. The same three ERVs from the DC versus WT com-
parison were up-regulated, as well as eight additional elements
(unadjusted $P < 0.05$; Fig. 5, G to J, and table S5). MuLV-int also
demonstrated significantly increased accessibility from our ATAC-seq
data, suggesting that MuLV-int elements are strongly dysregulated
in this context. Combined, these data further indicate that ERV dys-
regulation occurs downstream of Daxx loss in vivo.

**Transcriptome data from human PanNETs supports coordinate dysregulation of ERVs and protein-coding genes**

Transposable elements, including ERVs, evolved independently
and are not conserved between mouse and human germlines, with
different elements integrating in different regions of the genome.
We therefore wanted to determine whether the mechanisms down-
stream of Daxx loss and the coordinate dysregulation of ERVs and
coding genes might be conserved. We turned to a published study
of human PanNETs, which included RNA-seq data for 33 samples
(21). Because we cannot yet fully appreciate the contributions of
Atrx to our phenotypes and mechanisms in the mouse, we excluded
the three samples with ATRX mutations (SA514934, SA506710, and
SA506700) and one of two samples (SA528774) from a single
donor. We then conducted a differential gene expression analysis
between the five tumors with DAXX mutations and the 24 tumors
that were WT for DAXX (Fig. 6A). Despite the relatively low sample
size in this analysis, we identified four genes with significant gene
expression changes in DAXX mutant PanNETs compared with
controls (Fig. 6B and table S6). As a positive control, DAXX expression
was significantly reduced ($\log_{2} FC = -1.729$, $P_{adj} = 0.00114$) in
DAXX mutant tumors. This is consistent with published data showing
that DAXX mutant tumors are negative for DAXX protein
generating related ERV protein expression changes in (5). Two genes were found to
be significantly up-regulated in DAXXX mutant tumors: $FABP3$
($\log_{2} FC = 6.784$, $P_{adj} = 1.03 \times 10^{-7}$) and $SERINC2$ ($\log_{2} FC = 3.97$
$P_{adj} = 0.00131$). These genes are located adjacent to each other on
chromosome 1p35.2, and there is a large ERV locus just 18.8 kb

Fig. 5. Daxx-dependent gene expression changes are associated with ERV dysregulation. (A) Venn diagram depicting overlap of significant DEGs between mouse genotypes. Hypergeometric $P = 9.88 \times 10^{-7}$. (B) Correlated expression ($R^2 = 0.9768$) of the nine common Daxx-dependent gene expression changes in (A), with $Bglap3$ denoted in red. (C) $Bglap3$ normalized read counts from RNA-seq data (mean $\pm$ SD). Unadjusted $P$ values are indicated. (D) Normalized IAP-d counts from RNA-seq analysis (mean $\pm$ SD). Unadjusted $P$ values are indicated. (E) Percentage of DEGs with associated ERVs, compared with three randomly generated gene lists (mean $\pm$ SD). (F and G) Differential expression of transposable element families from RNA-seq data. Data points in red or purple reflect increased expression with an unadjusted $P < 0.05$. (H) Correlated ERV expression between DC versus WT and DMC versus MC pancreases. (I and J) Normalized RLTR10D and RLTR6-int counts from RNA-seq analysis. Unadjusted $P$ values are indicated.
variants, which cooperate to shape and restrict the transcriptional methylation, histone posttranslational modification, and histone
cocrine tumorigenesis.
in vivo, and implicating ERV dysregulation in pancreatic neuroen-
somatic regulation of ERVs beyond DNA and histone methylation
regulation of ERVs, thereby expanding our understanding of the ery
in vivo. The transcriptional changes are associated with dys-
profile and cell state, which ultimately impairs pancreas recov-
Men1 loss) to affect the transcriptional
and other genetic lesions (
state that cooperates with environmental stress (inflammation)
deletion creates a permissive chromatin
ings demonstrate that
Daxx, a chap-
erone for the H3.3 histone variant, we have generated a conditional
mouse model and used it to extensively interrogate the transcrip-
tome and biology downstream of Daxx loss. Combined, our find-
ings suggest that Daxx deletion creates a permissive chromatin
state that cooperates with environmental stress (inflammation) and other genetic lesions (Men1 loss) to affect the transcrip-
tional profile and cell state, which ultimately impairs pancreas recover-
y in vivo. The transcriptional changes are associated with dys-
regulation of ERVs, thereby expanding our understanding of the somatic regulation of ERVs beyond DNA and histone methylation
in vivo, and implicating ERV dysregulation in pancreatic neuroen-
docrine tumorigen.

Chromatin is maintained by a complex interplay between DNA methylation, histone posttranslational modification, and histone
variants, which cooperate to shape and restrict the transcriptional
landscape of specific cell types. Alterations in the epigenetic landscape can provide a more permissive (or less restricted) transcriptional
state. The concept of permissive chromatin was recently supported by a study evaluating the role of Setdb1 in silencing ERVs in mouse embryo fibroblasts (40). Specifically, the data sug-
gest that dysregulation of histone methylation alone is insufficient to drive robust transcriptional changes; however, the altered chromatin landscape can cooperate with context-dependent transcriptional factors to significantly affect the expression of ERVs. The basal
overexpression of Bglap3 in Daxx-depleted pancreases suggests epi-
genetic dysregulation in our model; however, we observed robust phenotypic and transcriptomic changes only in the context of tissue stress, suggesting a primed state that is waiting for a signal. These findings further suggest that proteins that are normally induced in response to injury may affect ERV expression or promoter/enhancer activity. Our RNA-seq data showed up-regulation of five pancreas cell type–specific transcription factors (table S2). This is an intrigu-
model to explain the specificity of DAXX mutations in human
PanNETs, implying that epigenetic dysregulation may cooperate with islet-specific transcriptional regulation to promote tumorigen.

In addition to creating a more permissive transcriptional land-
scape, mutations that alter chromatin maintenance mechanisms can have profound biological consequences, including giving cells the capacity to alter their state. A classic example of this is polycomb protein mutations in Drosophila, which cause transdifferentiation across cell lineages (43). Deposition of the histone variant H3.3 has previously been associated with cell fate transitions (44). Loss of H3.3 early during reprogramming of fibroblasts to induced pluripo-
tent stem cells enhanced the efficiency of the process, whereas H3.3 loss late in the process of differentiation results in an impaired ability to establish a new cell state. Our work suggests that Daxx restricts cellular plasticity in vivo. In the pancreas, cellular plasticity is exem-
plified by acinar-to-ductal metaplasia. Following inflammatory stress, loss of Daxx causes an increase in the expression of ductal genes, compensatory ductal proliferation, and a failure to restore the nor-
mal acinar compartment in the background of Men1 deficiency. These results parallel the H3.3 studies in induced pluripotent stem cells.
Additional work to understand the contributions of Men1 loss will be important because DAXX and MEN1 mutations frequently co-occur in PanNETs. Although Men1 loss may merely increase tissue damage in response to treatment with caerulein, as we have reported (32), there could also be a mechanistic link between the two mutations. Like Daxx, Men1 has several reported functions. Men1 regulates the epigenome as part of the MLL histone methyl transferase complex, which promotes the activating H3K4me3 mark (23). The loss of Men1 could thereby further perturb the chromatin landscape. In addition, Men1 can negatively regulate the activity of transcription factors, including Jund (45). This provides a simple model of cooperation in which an aberrantly active transcription factor could be coupled to an altered or permissive chromatin landscape.

In addition to understanding genetic cooperation between Daxx and Men1, it will also be of interest to explore how Daxx loss affects other mechanisms that maintain heterochromatin and the silencing of ERVs. Our results are similar to those studies evaluating loss of H3K9me3 (40, 46, 47). Depending on the cell type, Setdb1 depletion causes specific families of ERVs to be derepressed, as we also observed from our RNA-seq analysis of caerulein-treated pancreases. These data further suggest that additional cell-intrinsic factors contribute to the transcriptional outputs and downstream biology. In addition to histone methylation, DNA methylation also plays an important role in establishing and maintaining heterochromatin and ERV silencing. Comprehensive epigenomic profiling and integration will be essential to explore and understand the chromatin-based alterations downstream of Daxx loss. In addition, as longer read sequencing technologies advance, we will be in a stronger position to interrogate locus-specific transposable elements and their epigenetic regulation.

Last, this work has provided insights into the molecular mechanisms downstream of Daxx loss in vivo. Although our data strongly implicate the dysregulation of ERV silencing as an important biological consequence downstream of Daxx loss, transposable elements have evolved independently in the mouse and human genomes. Hence, the specific elements and their localization in the genome are not conserved between species. Therefore, the mouse may not be a faithful model for PanNETs with DAXX mutations. Moving forward, understanding how Daxx loss affects chromatin organization, transposable element silencing, and subsequent gene expression in humans will be essential to advance our understanding of relevant disease pathogenesis. We analyzed publicly available human PanNET RNA-seq data and identified significant up-regulation of two genes, SERINC2 and FABP3. These are adjacent genes nearby a large ERV locus. Moreover, the coordinate dysregulation of HERV10F-int and SERINC2 has been previously reported in lung cancer (42), and derepression of HERV10F-int was also found in HCC1944 breast cancer cells compared with human mammary epithelial cell controls (48). Even the small number of tumors analyzed in our study implicates transposable element derepression driving altered expression of endogenous genes in human PanNETs.

Caeerulein-induced chronic pancreatitis
Chronic pancreatitis was induced using established protocols (31). Caeerulein (MedChemExpress) was reconstituted in sterile phosphate-buffered saline. Adult mice (6 to 10 weeks old) were injected with caerulein (individual dose, 50 µg/kg), given as four hourly intraperitoneal injections, three times per week for three consecutive weeks.

Histology and immunohistochemistry
Tissues harvested from mice were fixed in 10% neutral-buffered formalin and embedded in paraffin. Four-micrometer sections were stained with hematoxylin and eosin and examined using light microscopy. Tissue processing, paraffin embedding, sectioning, and hematoxylin and eosin staining were performed by the MD Anderson Department of Veterinary Medicine and Surgery Histology Laboratory. Immunohistochemistry was performed using standard methods, and tissue sections were stained with antibodies against CK19 (ab133496, Abcam) and CD45 (550539, BD Pharmingen). Visualization was performed using avidin-biotin complex and DAB (3,3′-diaminobenzidine) kits (Vector Laboratories) and slides were counterstained with hematoxylin.

MATERIALS AND METHODS

Mice
All mouse experiments were performed in compliance with the National Institute of Health guidelines for animal research and approved by The University of Texas MD Anderson Cancer Center Institutional Animal Care and Use Committee. Men1 conditional knockout mice (The Jackson Laboratory, stock no. 005109), Atrx conditional knockout mice, and Pdx1-CreERT2 (The Jackson Laboratory, stock no. 014647) mice were previously described (14, 25, 49). Mice were maintained on a mixed background (C57BL/6, FVB, 129). Unless otherwise stated, Cre-negative littersmates were used as functionally WT controls throughout these studies.

To generate the conditional DaxxΔ allele, we used sequence surrounding the mutant loxp site in the DaxxΔm2Led mice (Jackson Laboratory, stock no.008669) to score possible single-guide RNA (sgRNA) sequences using the CRISPR Design Platform (crispr.mit.edu). Our approach used a protospacer adjacent motif (PAM) sequence 15 bp upstream of the loxp mutation, within the retained flippase recognition target (FRT) site, and a single-stranded donor oligonucleotide that introduces a BamH1 sequence in place of the six nucleotides immediately upstream of the PAM sequence (fig. S1A). This allowed us to use restriction digest to screen pups for efficient homology-directed repair, and it destroyed the n20 sequence to prevent retargeting of the allele. The injection solution contained Cas9 mRNA (5 ng/µl) (PrecisionX hspCas9 SmartNuclease mRNA, Systems Biosciences), sgRNA (2.5 ng/µl), and single-stranded donor oligonucleotides (5 ng/µl), prepared in tris-EDTA buffer (5 mM tris, 0.1 mM EDTA). This was injected into the pronucleus of 200 to 250 DaxxΔm2Led/+ zygotes, which were subsequently implanted into pseudo-pregnant recipient WT females (20 to 25 per mouse). All injections and implantations were performed by the MD Anderson Genetically Engineered Mouse Facility. In all, 10 mice were obtained and screened. We observed nonhomologous end-joining in two mice (20%), and homology-directed repair in four mice (40%). Notably, three of four of these mice incorporated the BamH1 sequence but not the repaired loxP sequence, demonstrating that homology-direct repair can occur with as little as 25 bp of homology. We obtained one mouse with homology-directed repair that incorporated both the BamH1 sequence and the fixed loxP sequence (fig. S1, B and C). We confirmed germline transmission and subsequently screened the seven highest scoring potential off target sites as determined by the CRISPR Design Platform using PCR and Sanger sequencing and found no evidence of off-target cutting at these loci (data not shown). We then backcrossed for two generations to WT C57BL/6 mice.

Downloaded from http://advances.sciencemag.org/ on October 28, 2020
Image quantification
Acinar-to-ductal metaplasias and islets were quantified from whole pancreas sections scanned on an Aperio ScanScope XT (Leica Biosystems Inc.) and manually annotated using Aperio ImageScope software (Leica Biosystems Inc.). The percentage of total pancreas areas was calculated for each sample.

Western blot analysis
Tissue samples were flash-frozen, pulverized, and lysed in SDS lysis buffer [1% SDS, 6.5 mM tris-HCl (pH 6.8), 25% glycerol, 10% β-mercaptoethanol]. Protein extracts were separated by SDS–polyacrylamide gel electrophoresis, transferred to nitrocellulose membranes, and probed with antibodies against Daxx (M–112, Santa Cruz Biotechnology) and vinculin (V1931, Sigma-Aldrich). Proteins were visualized using Li-COR secondary antibodies and scanned on a Li-COR Odyssey Imager.

RNA isolation
Flash-frozen tissue was pulverized and homogenized in TRIzol RNA isolation reagent (Invitrogen). RNA was then purified as per manufacturer’s protocols or using the RNeasy Mini Kit (Qiagen).

RNA-seq and analysis
Barcoded, Illumina-compatible, strand-specific total RNA libraries were prepared using the TruSeq Stranded Total RNA Sample Preparation Kit (Illumina). Briefly 250 ng of deoxyribonuclease I–treated total RNA was depleted of cytoplasmic and mitochondrial ribosomal RNA using Ribo-Zero Gold (Illumina). After purification, the RNA was fragmented using divalent cations and double-stranded complementary DNA (cDNA) was synthesized using random primers. The ends of the resulting double-stranded cDNA fragments were repaired, 5′-phosphorylated, and 3′-A–tailed, and then Illumina-specific indexed adapters were ligated. The products were purified and then enriched with 12 cycles of PCR to create the final cDNA library.

The libraries were quantified using the Qubit double-stranded DNA HS Assay (Thermo Fisher Scientific), and then 16 libraries were multiplexed per pool. The library pool was quantified by quantitative PCR using the KAPA Library Quantification Kit (KAPABiosystems) and assessed for size distribution using the Fragment Analyzer (Advanced Analytical) and then sequenced on the Illumina HiSeq 4000 sequencer using the 76-nucleotide paired-end format. For the basal analysis, four mice of each genotype (eight in total) were sequenced in half of a sequencing lane. For the caerulein-induced gene expression analyses, 16 samples (four mice for each of four genotypes) per lane were sequenced on two lanes to increase sequence depth.

STAR (version 2.6.0b) was used to align reads to the GRCm38. p6 reference genome (50). Samtools (version 1.8) was used to sort, convert between formats, and calculate mapping statistics (51). FastQC (0.11.5) was used to check FASTQ read quality. Gene annotation was carried out using GENCODE M19 (Ensembl 94) annotation, which was downloaded from the GENCODE project (52). Aligned reads were summarized at the gene level using high-throughput sequencing count (53). Differential gene expression was determined using R (3.5.1) and Bioconductor package DESeq2 (54). Read count was first prefiltered to remove extremely low expressed genes. DESeq2 was then used to carry out read count filtering, normalization, dispersion estimation, and identification of differential expression. DESeq2 modeled the counts using a negative binomial distribution, followed by the Wald test. The final adjusted P value was obtained using the Benjamini and Hochberg method.

ATAC-seq and analysis
Flash-frozen pancreatic tissue was pulverized and used for ATAC-seq following the OMNI-ATAC protocol (38). Briefly, tissue was added to a prechilled Dounce Homogenizer with cold homogenization buffer. Nuclei were pelleted and purified using an OptiPrep gradient and then collected. A total of 50,000 nuclei were used in the transposition reaction as per standard protocol using TDE1 (tagmentation enzyme, Illumina). DNA was purified using a Zymo DNA Clean and Concentrator kit. Barcodes were added through PCR amplification. Following library quantification, samples were pooled with all eight samples sequenced on a single lane of an Illumina HiSeq 4000, with 76-nucleotide paired-end sequencing.

Quality assessment of sequencing data was performed using FastQC (0.11.5) to check FASTQ read quality. Adapter sequences at the 3′ end of reads were removed using trim galore (0.4.5) and cutadapt (1.18). Reads were aligned to the GRCm38.99 reference genome using Bowtie2 (2.3.4.2) (55). Samtools (1.8) was used to sort, convert between formats, and filter alignment reads (51). Chromosome M (mitochondria) was removed from the alignments, and duplicated reads were also removed using PICARD (2.9.0). Reads that were mapped to multiple locations were removed, and only reads that mapped to unique chromosomal locations and paired properly were retained. Peaks were called using MACS2 (2.1.2) (56), and annotation was carried out using HOMER (4.10) (57). To count reads, we used paired-end mode with a maximum fragment length of 800 bp, a window size of 20 nucleotides, and a spacing of 20. When merging adjacent windows into the same genomic region, a cutoff of 10 bp was used. Windows that were less than 10 bp apart were considered adjacent and were grouped into the same cluster. To identify differentially accessible chromatin regions, we used the R package csaw (1.16.1), which carried out a window-based approach to compare between groups of samples (58).

Transposable element analysis of ATAC-seq and RNA-seq
The sequences for each of the transposable element families were extracted from Dfam, using the chromosome location of the nonredundant element with the smallest e-value as the reference sequence. A reference genome for transposable elements was then assembled specifically for this project, and sequencing reads were aligned to this genome using Bowtie (2.3.4.2). Analysis followed the same RNA-seq and ATAC-seq pipelines as described above.

Statistical analysis
Data are presented as means ± SD. All statistical analyses were performed using GraphPad Prism 6 software, and P < 0.05 was considered statistically significant. Comparisons between two groups were made using the Student’s t test, and comparisons among multiple groups were made using analysis of variance with the Dunnett’s multiple comparisons test to compare experimental samples to WT controls.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/32/eaba8415/DC1

View/request a protocol for this paper from Bio-protocol.

Acknowledgments: The Atrx conditional knockout mouse was provided by R. Gibbons. We thank S. Arur, N. Copeland, N. Jenkins, and A. Maitra, as well as members of the Lozano laboratory for helpful discussions. We also acknowledge Scientific Publications, Research Medical Library at MD Anderson for editing this manuscript. Funding: This research was supported by grants from the Neuroendocrine Tumor Research Foundation and NIH (NIH/NCI R21 CA208463) to G.L. and a fellowship from the Canadian Institutes of Health Research to A.R.W. Sequencing experiments were performed with The MD Anderson Advanced Technology Genomics Core and gene targeting to generate the Daxx allele was conducted by The MD Anderson Genetically Engineered Mouse Facility. Both core facilities were supported, in part, by a Cancer Center Support Grant from the National Cancer Institute (CA016672).

Author contributions: A.R.W. and G.L. conceived the initial concept, developed the project, interpreted data, and wrote the manuscript. Experiments were conducted by A.R.W., C.S., S.M.M., G.P.C., and N.K.A. All computational analyses were done by Y.Q., with supervision from X.S.F.M., M.P.K., and M.C.B. contributed to experimental design and interpretation. J.S.E. performed pathological evaluation of mouse tissue samples. All authors reviewed and provided suggestions on the manuscript. Competing Interests: G.L. is a current member of the Board of Scientific Advisors for the Neuroendocrine Tumor Research Foundation. N.K.A.’s contributions were made while affiliated with The University of Texas MD Anderson Cancer Center. As per the requirement of the journal, he is currently an employee of AstraZeneca and may or may not own stock options. The other authors declare no competing interests. Data and materials availability: RNA-seq and ATAC-seq data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database with accession nos. GSE145175, GSE150017, and GSE150011.

Submitted 10 January 2020
Accepted 18 June 2020
Published 5 August 2020
10.1126/sciadv.aba8415

Daxx maintains endogenous retroviral silencing and restricts cellular plasticity in vivo
Amanda R. Wasylishen, Chang Sun, Sydney M. Moyer, Yuan Qi, Gilda P. Chau, Neeraj K. Aryal, Florencia McAllister, Michael P. Kim, Michelle C. Barton, Jeannelyn S. Estrella, Xiaoping Su and Guillermina Lozano

Sci Adv 6 (32), eaba8415.
DOI: 10.1126/sciadv.aba8415