Supplemental Information

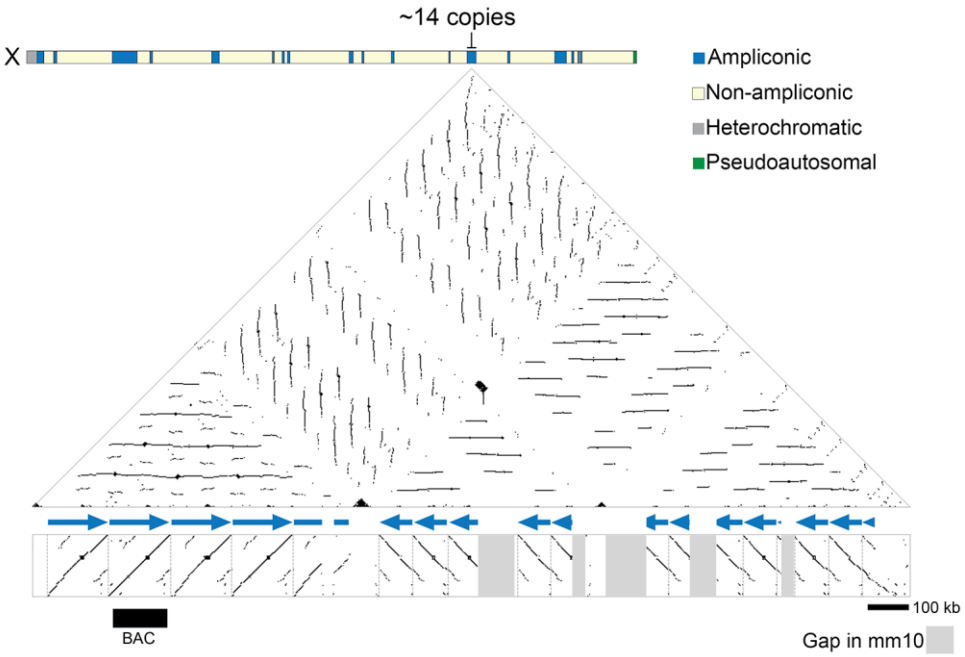
Genomic structure, evolutionary origins, and reproductive function of a large amplified intrinsically disordered protein-coding gene on the X chromosome (*Laidx*) in mice.

Martin F. Arlt, Michele A. Brogley, Evan R. Stark-Dykema, Yueh-Chiang Hu, Jacob L. Mueller

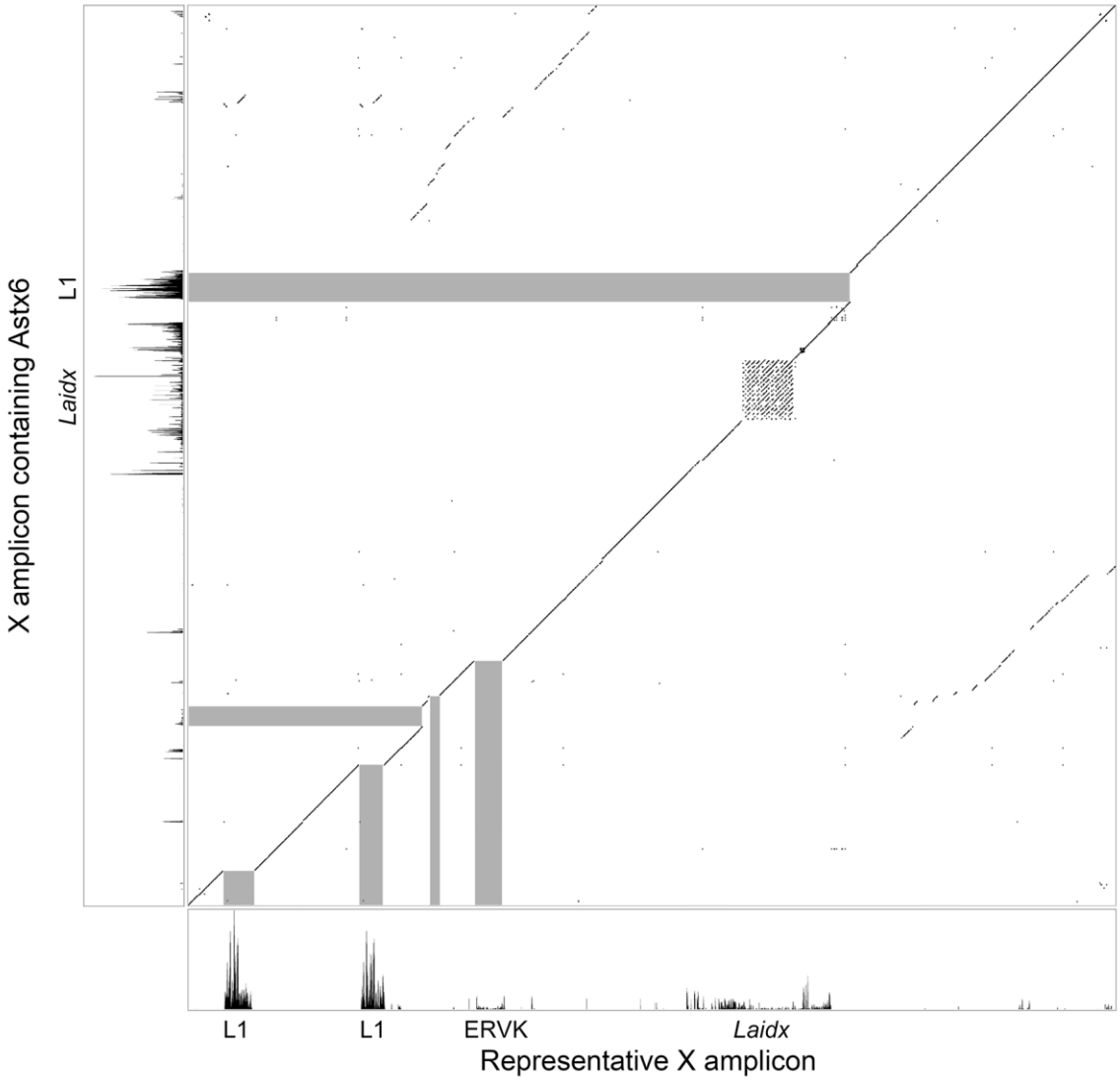
Supplemental Table 1: RT-PCR primers		
Primer Name	**Primer Sequence**	**Notes**
Gm17412 F	CGGACTTTACCCTCTGACCA	
Gm17412 R	CTTCCTTTCTGTTGGAGAGCA	
Srsx F	GATCCCCAGCTTTCTCAAGG	
Srsx R	TGCCAAAGCTTCTGGAGTCT	
SrsxA F	CCTGTCACCCAAGAGGTCAT	
SrsxB R	TATCCCTGTGGCATCTGTTG	
AstxE1-2 R	TCCAGGTCTTAGCTCATGAGG	Spans junction across *Laidx* exons 1 & 2
AstxE1-2 F	CCTCATGAGCTAAGACCTGGA	Spans junction across *Laidx* exons 1 & 2
AstxE9 R	AATCCATCGATGATGAGTGC	
Laidx1149 F	CAAAGATGATGGGGACACCT	Primer within 1149 bp tandem repeat unit
Laidx1149 R	CACCATCACCTCGAGAACCT	Primer within 1149 bp tandem repeat unit
Laidx1 R	GCAGGGTCAGGTCTCAAACT	
Laidx2 R	GAAGGCATGGAAAAGGAACA	
Laidx3 F	ATGGTGTGAAACCATCACTC	
Laidx4 R	CAGTCAGTGGTCACAACAAC	
Laidx5 F	ATGTTGGGAGGTTGTTGTGA	
Laidx6 R	CCACTTTTCTTGAATGGAGT	
Trim42 F	GAAGCATCGTCACCTCCTCT	
Trim42 R	CTTCTCGCATAGGCTGTGGT	

Supplemental Table 2: *Laidx* RT-PCR assays		
Assay Name	**Primer 1**	**Primer 2**
Laidx 1	AstxE1-2 F	AstxE9 R
Laidx 2	SrsxA F	AstxE1-2 R
Laidx 3	Srsx F	Srsx R
Laidx 4	SrsxB R	Laidx1149 R
Laidx 5	Laidx1149 F	Laidx1149 R
Laidx 6	Gm17412 F	Laidx1 R
Laidx 7	Laidx3 F	Laidx4 R
Laidx 8	Laidx5 F	Laidx6 R
Laidx 9	AstxE1-2 F	Laidx6 R
Laidx 10	Laidx3 F	Laidx6 R
Laidx 11	Gm17412 F	Gm17412 R
Laidx 12	Gm17412 F	Laidx2 R

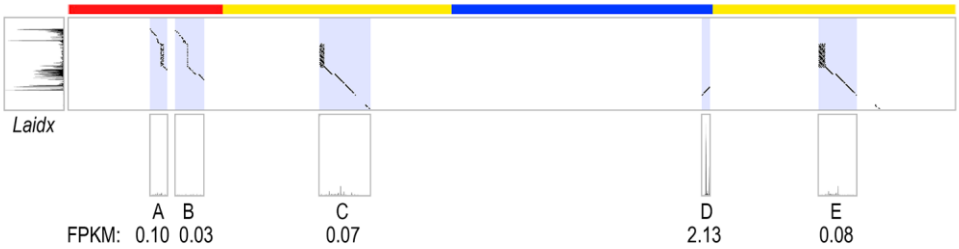
Supplemental Table 3: Oligos used to generate and genotype transgenic lines	
Oligo Name	**Oligo Sequence**
Laidx 5' gRNA	TGGCTGGGAGTGTGTCTTCA
Laidx 3' gRNA	CATGGGTTCTGGGTTAGTTA
Laidx 5' Donor	GTTACTATGTAGCTATTCGCTAGTTGTATTATTTAACACTAGTGAAGTTGTTTTAACTCTCAATCTCAGTTTCATGTTTATAATGTCCATGAATTCATAACTTCGTATAATGTATGCTATACGAAGTTATCGATAGACACACTCCCAGCCATAAATTTATAATCAGAAA
Laidx 3' Donor	ATATTAAATATTAATTAGGCACCTATGTGAGTACGGATTAGTGAAATGAACAAATATGTTTTACATGCTTATTGAGAAAAGCCATAAGCTTATAACTTCGTATAATGTATGCTATACGAAGTTATGAATTCTAACCCAGAACCCATGTCAAGGTTAACAGATTAA
5' *Laidx* F	TCCAATTTCAATTAACTACTCAACAAA
5' *Laidx* R	GTAGATTTAAAACTCTGTGCAAGATGA
3' *Laidx* F	AGGCACCTATGTGAGTACGGATTAGTG
3' *Laidx* R	GAGATCGTAGTTCTAGGCTTCATTACTCTG
Ube1xy F	TGGATGGTGTGGCCAATG
Ube1xy R	CACCTGCACGTTGCCCTT



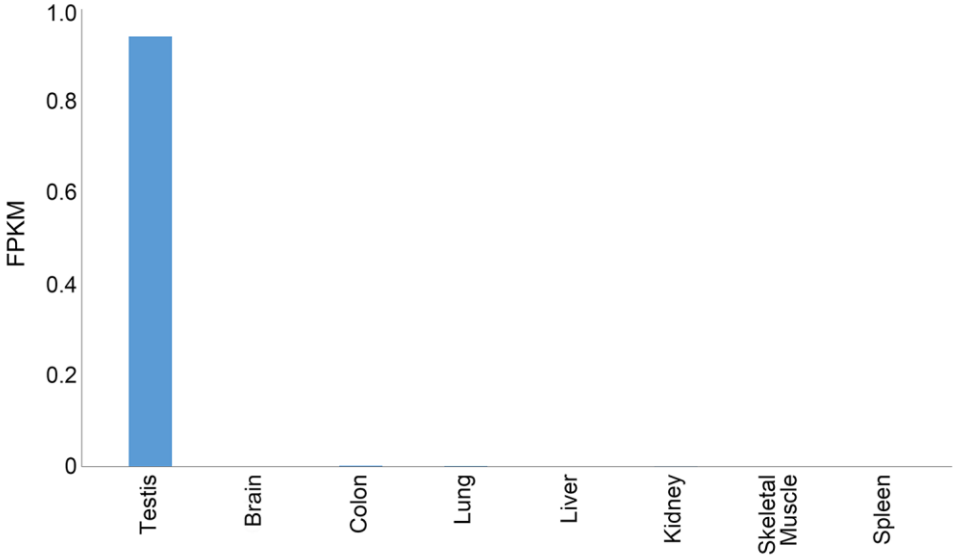
Supplemental Figure 1. Self-symmetry triangular dot plot of the *Srsx*-containing ampliconic region on the X (chrX:123,050,000-126,250,000; mm10). Each dot represents a perfect match of 100 nucleotides. Below the triangular dot plot is a dot plot comparing the consensus *Srsx*-containing X amplicon unit to the entire ampliconic region. Blue arrows indicate position and orientation of *Srsx*-containing X amplicon units. The position of PacBio-sequenced BAC RP23-106J7 is indicated. Vertical dotted gray lines mark the boundaries of each amplicon unit. Gray bars mark gaps in the mm10 reference genome sequence.



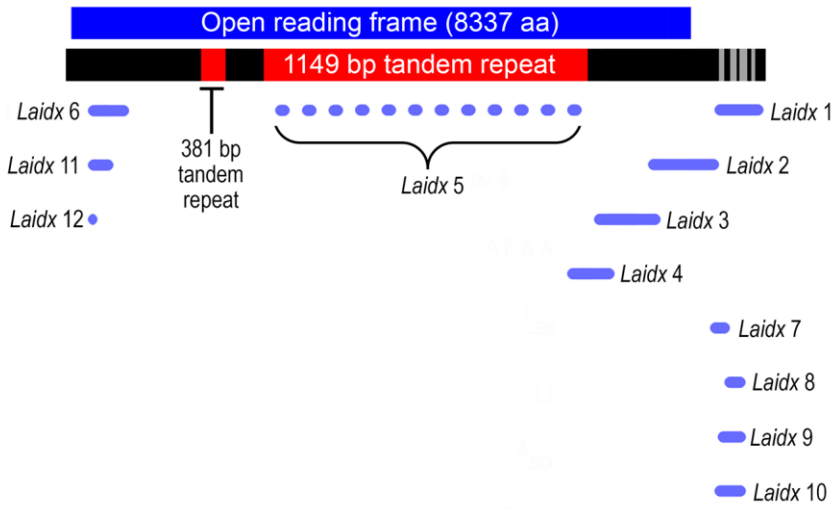
Supplemental Figure 2. Dot plot of DNA sequence identity between the consensus *Srsx*-containing X amplicon (chrX:123,326,277-123,555,768; mm10) and another copy of the amplicon (chrX:123,104,838-123,326,276; mm10), on the Y and X axes, respectively. Each dot represents 100% sequence identity in a 50 bp window. Gray highlights indicate sequence found in only one of the two amplicon units. Testis RNAseq reads for each region are illustrated along the respective axes.



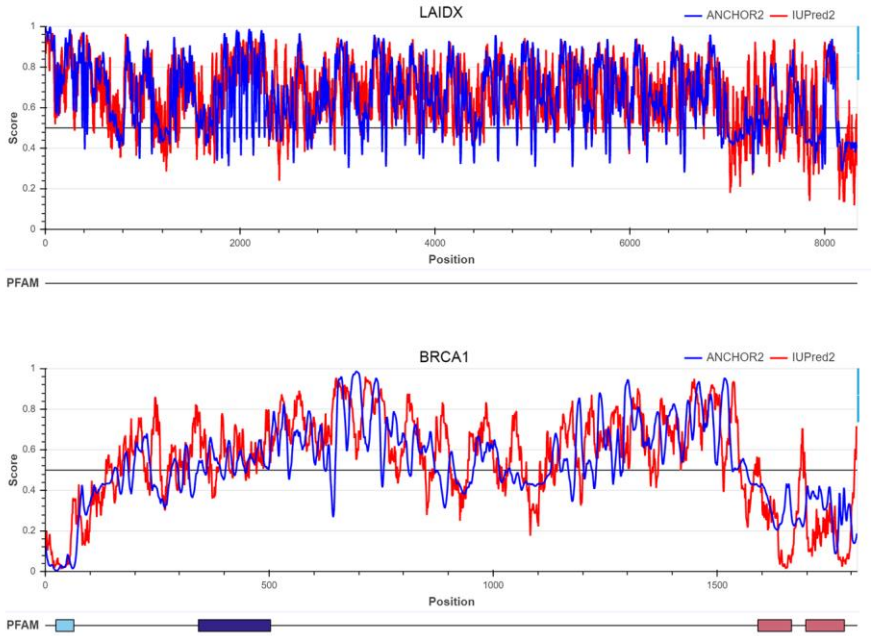
Supplemental Figure 3. Chromosome Y sequences homologous to *Laidx* (blue shading) are rearranged and contain five regions (A-E) with low levels of expression in round spermatids.



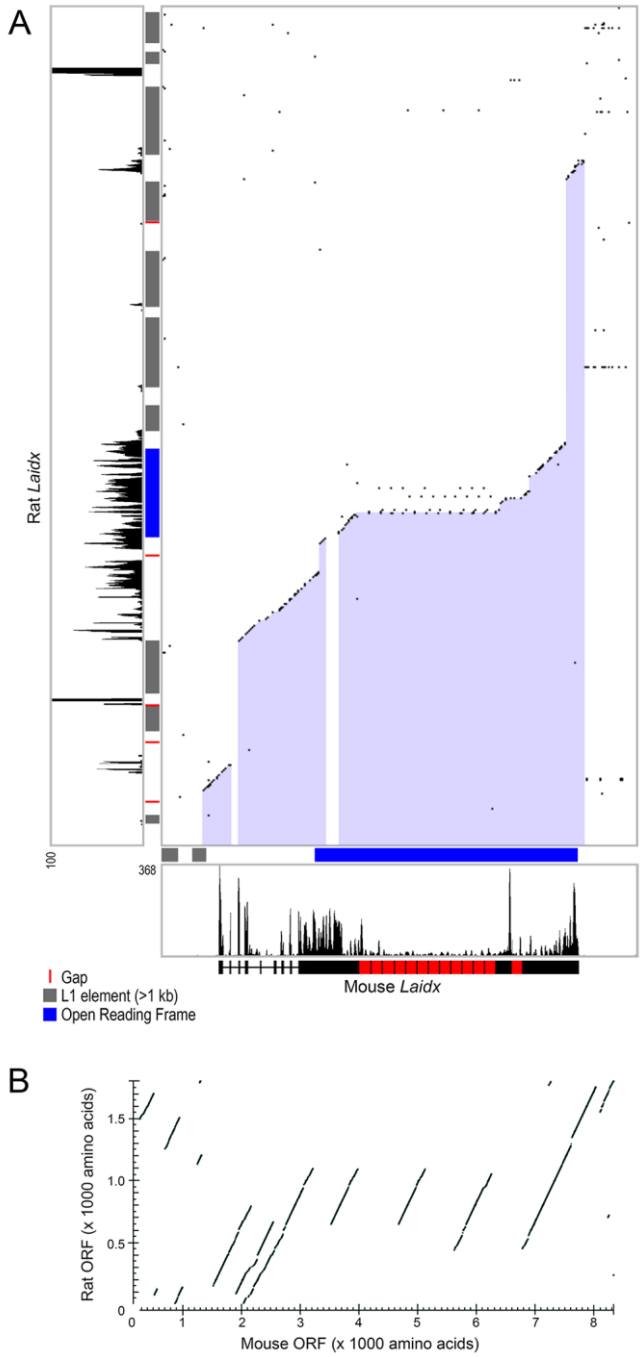
Supplemental Figure 4. *Laidx* is expressed exclusively in the adult mouse testis. mRNAseq analysis performed on publically available datasets (See Materials and Methods).



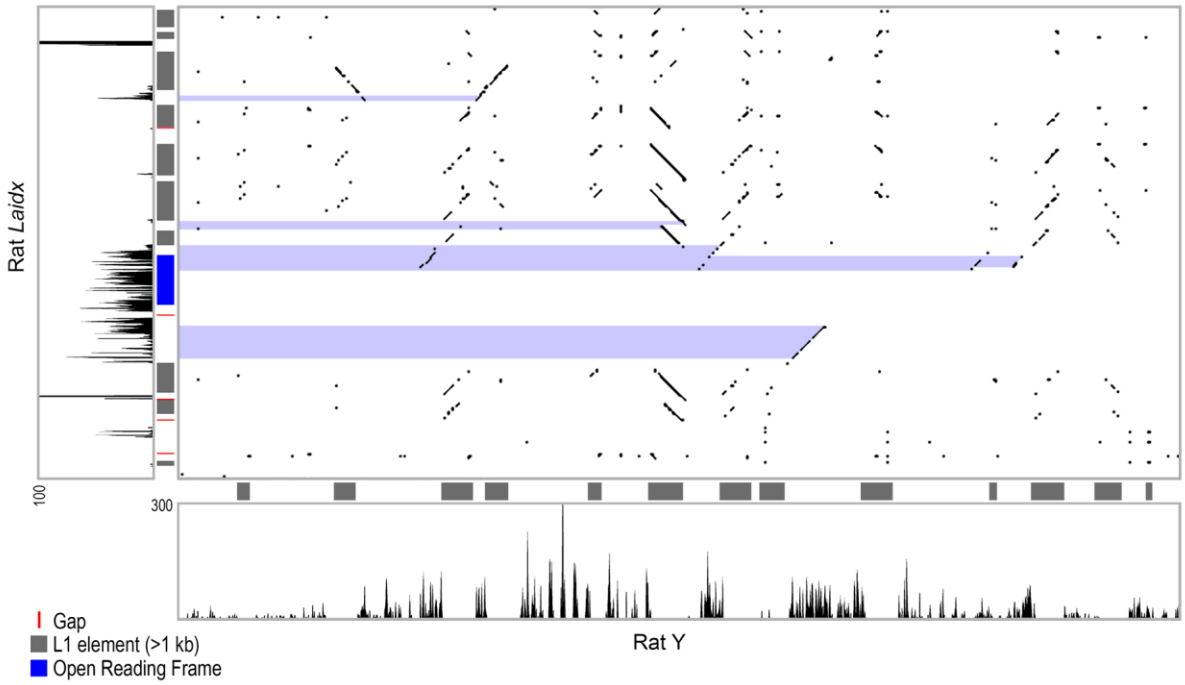
Supplemental Figure 5. RT-PCR characterization of *Laidx*. Schematic of predicted cDNA with tandemly repeated regions indicated in red. Alternating black and gray stripes represent exons. Purple bars indicate RT-PCR assays used to characterize *Laidx* (See Supplemental Table 2).



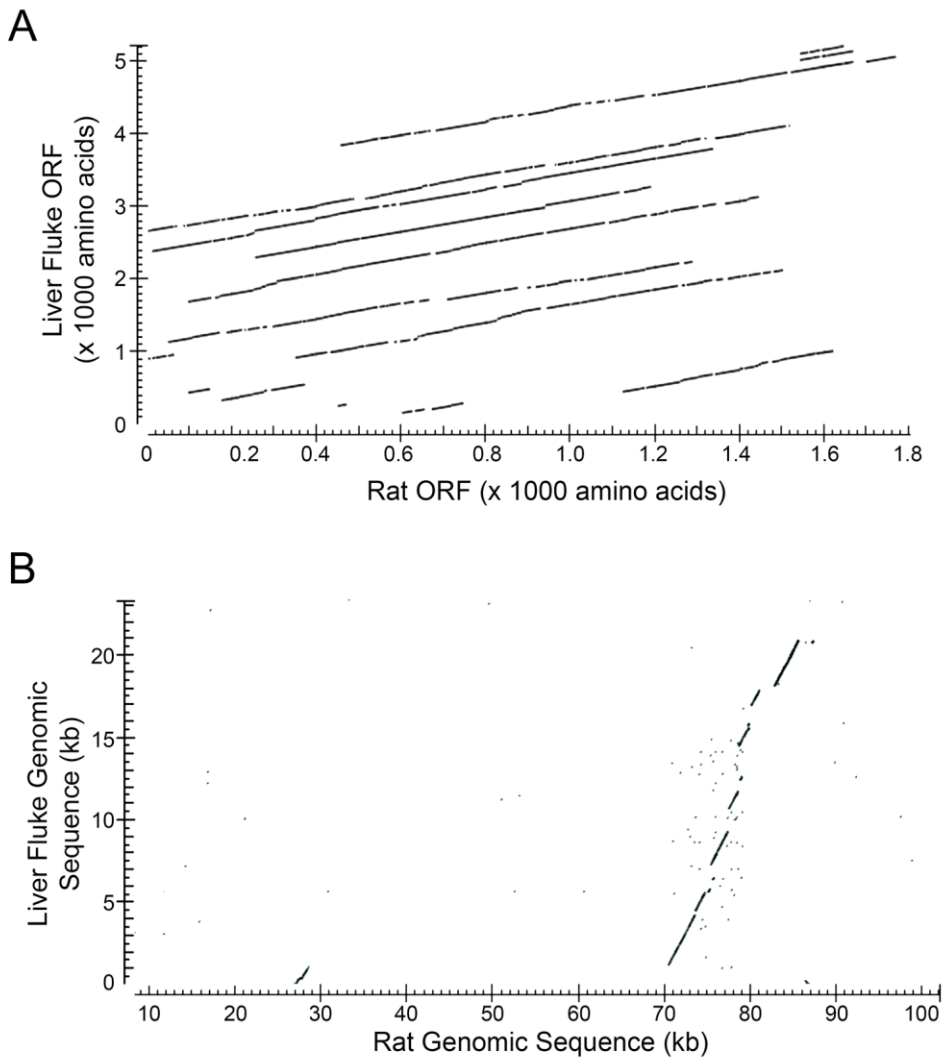
Supplemental Figure 6. LAIDX is an intrinsically disordered protein. IUPred2A (https://iupred2a.elte.hu) output shows high probability of disorder across the length of LAIDX (top). Mouse BRCA1 output is included for comparison (bottom).



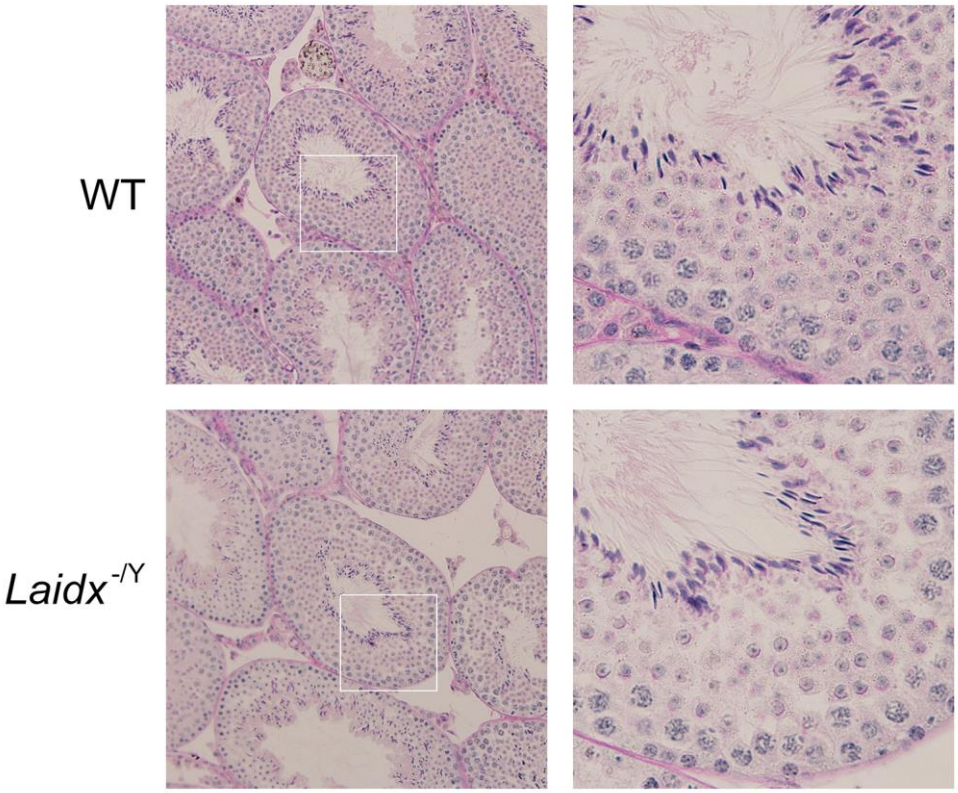
Supplemental Figure 7. (A) Dot plots of DNA sequence identity at the *Laidx* locus in mouse (X-axis) and rat from BAC CH230-1D6 (Y-axis). Each dot represents 100% sequence identity in a 15 bp window. Blue highlights indicate regions of sequence identity between the two sequences. Rat and mouse testis RNA-seq reads for each region are illustrated along the respective axes. Red lines represent sequence gaps, condensed to 50 bps. Blue boxes represent *Laidx* open reading frames. Gray boxes represent rodent LINE1 (L1) elements. (B) Blastp alignment of the 8337 amino acid mouse LAIDX protein and the 1806 amino acid rat LAIDX protein.



Supplemental Figure 8. Dot plots of DNA sequence identity at the *Laidx* locus in rat from BAC CH230-1D6 (Y-axis) and the rat Y chromosome BAC RNAEX-9O8. Each dot represents 100% sequence identity in a 25 bp window. Blue highlights indicate regions of sequence identity, excluding repetitive elements. Rat testis RNA-seq reads for each region are illustrated along the respective axes. Red lines represent sequence gaps, condensed to 50 bps. Blue boxes represent *Laidx* open reading frames. Gray boxes represent rodent LINE1 (L1) elements.



Supplemental Figure 9. *Laidx* sequence comparisons between rat and *Clonorchis.* (A) Blastp alignment of the 5280 amino acid *Clonorchis* protein CSKR_14446s and the predicted 1806 amino acid rat LAIDX protein. (B) TBLASTN alignment of rat X chromosome BAC CH230-1D6 with *Clonorchis* genomic sequence encoding CSKR_14446s.



Supplemental Figure 10. Periodic-acid Schiff stained testis histology from WT and *Laidx*$^{-/Y}$ males demonstrating normal development and morphology.