Truncating Prolactin Receptor Mutations Promote Tumor Growth in Murine Estrogen Receptor-Alpha Mammary Carcinomas

Short title: Prlr truncation promotes tumor growth in *Stat1-/-* mice

Obi L. Griffith[1,2,3,9], Szeman Ruby Chan[4,8,9], Malachi Griffith[1,3,5], Kilannin Krysiak[1], Zachary L. Skidmore[1], Jasreet Hundal[1], Julie A. Allen[4], Cora D. Arthur[4], Daniele Runci[4], Mattia Bugatti[7], Alexander P. Miceli[4], Heather Schmidt[1], Lee Trani[1], Krishna-Latha Kanchi[1], Christopher A. Miller[1,2], David E. Larson[1,5], Robert S. Fulton[1,5], William Vermi[4,7], Richard K. Wilson[1,2,3,5], Robert D. Schreiber[4,6,10,*], Elaine R. Mardis[1,2,3,5,*]

1. McDonnell Genome Institute, Washington University School of Medicine, 4444 Forest Park Ave, St. Louis, MO, 63108, USA.
2. Department of Medicine, Washington University School of Medicine, 660 S Euclid Ave, St. Louis, MO, 63110, USA.
3. Siteman Cancer Center, Washington University School of Medicine, 4921 Parkview Pl, St. Louis, MO, 63110, USA.
4. Department of Pathology and Immunology, Washington University School of Medicine, 660 S Euclid Ave, St. Louis, MO, 63110, USA.
5. Department of Genetics, Washington University School of Medicine, 660 S Euclid Ave, St. Louis, MO, 63110, USA.
6. Center for Human Immunology and Immunotherapy Programs, Washington University School of Medicine, 425 S Euclid Ave, St. Louis, MO, 63110, USA.
7. Department of Molecular and Translational Medicine, Section of Pathology, School of Medicine, University of Brescia, Piazza del Mercato, 15, 25121 Brescia BS, Italy.
8. Present address: Janssen Research & Development, Johnson and Johnson, 1400 McKean Rd, Spring House, PA 19477, USA.
9. Co-first author
10. Lead contact

* Correspondence: schreiber@immunology.wustl.edu (R.D.S), emardis@wustl.edu (E.R.M)

Supplemental Experimental Procedures

Whole Genome Sequencing
The yield and integrity of native genomic DNA was verified by a PicoGreen assay to determine mass (Invitrogen, Carlsbad, CA). Small insert dual indexed Illumina paired end libraries were constructed with the KAPA LTP sample prep kits according to the manufacturer's recommendations (KAPA Biosystems, Woburn, MA) with a few exceptions: (1) 100-500 ng of gDNA was fragmented using a Covaris E220 DNA Sonicator (Covaris, Woburn, MA) within a size range between 200-600 bp using the following settings: volume = 50 µL, temperature = 4 °C, duty cycle = 5, intensity = 4, cycle burst = 200, time = 90 seconds. (2) To prevent excessive over-amplification during PCR, cycle optimization was performed. 2 µL of the library was cycled as follows: 98 °C for 30 seconds, cycle – 98 °C for 10 seconds, 65 °C for 30 seconds, and 72°C for 30 seconds. After cycles 6, 8, 10, and 12 the program was halted and a 5 µL aliquot was collected. Each cycle amplification product was evaluated on a 2.2% agarose Flash Gel (Lonza Group, Basel, Switzerland) and the optimum cycle number determined. (3) Eight PCR reactions were amplified at the determined cycle number to enrich for proper adaptor ligated fragments. Libraries were fractionated on the LabChip XT using the DNA 750 chip (Perkin Elmer, Hopkinton, MA) collecting three unique fractions: 375 bp, 475 bp, and 675 bp with a +/- 5% covariance. Each fraction/library was assessed for concentration and size to determine molarity using the Qubit Fluorometer Quant-iT dsDNA HS assay (Life Technologies, Carlsbad, CA) and the Agilent BioAnalyzer High Sensitivity DNA Assay (Agilent Technologies, Santa Clara, CA), respectively. The concentration of each library fraction was verified through qPCR according to the manufacturer's protocol (Kapa Biosystems, Woburn, MA) to produce cluster counts appropriate for the Illumina HiSeq2000 platform. Each genome was loaded on HiSeq2000 version 3 flow cell according to the manufacturer's recommendations (Illumina, San Diego, CA). 2 X 101 bp read pairs were generated for each sample, yielding an average of 37.1x sequence coverage for the tumor genomes and 27.2x sequence coverage for the normal genomes (**Table S8**). The reference-aligned whole genome sequence data and sample details for 52 tumor and non-tumor mouse tissues have been submitted to NCBI SRA study SRP061941, BioProject PRJNA248457.

Reference alignment and somatic variant detection
Specifically, 'Reference Alignment' and 'Somatic Variation' analysis pipelines were used to identify somatic single nucleotide variants (SNVs), small insertions and deletions (indels), and copy number alterations (CNVs). Alignment was performed by BWA (v0.5.9) (Li and Durbin, 2009) against the mouse reference genome (mm9). Duplicates were marked by Picard Mark Duplicates (v1.46) [http://broadinstitute.github.io/picard/]. SNV calls were identified as the unique union of SamTools (version r963) (Li et al., 2009), Sniper (version 1.0.2) (Larson et al., 2012), VarScan (version 2.2.6) (Koboldt et al., 2012), and Strelka (version 0.4.6.2) (Saunders et al., 2012). Small insertion and deletion (indel) calls were identified as the unique union of GATK (1.0.5336) (McKenna et al., 2010), Pindel (version 0.5) (Ye et al., 2009), VarScan and Strelka. Copy number variants (CNVs) were called with CopyCat (https://github.com/chrisamiller/copycat). Annotation was performed by our custom annotator against Ensembl (version 67).

Validation and Extension Sequencing of Prlr by Sanger and MiSeq
Based on the region in which truncating *Prlr* mutations were observed in the discovery set (chr15:10258139-10258195; mm9) two sets of primers were designed to encompass this region with approximately 50 or 100 bp additional flanking sequence on each side, respectively. Primers

were designed using an in-house primer design graphical user interface (GUI) that does the following: (1) performs repeat and SNP masking; (2) reduces the amplicon size based on increasing GC-content; (3) blasts primers to verify a single unique genomic product (see **Table S10** for primer details). Primers were tailed (p1k / m13 reverse) for 3730 sequencing and ordered from Integrated DNA Technologies (IDT, Coralville, IA). Once received, primers were resuspended in 1X TE buffer. PCR reactions consisted of 5 ng DNA input, 6.25 µL of pooled and tailed primer pairs at 1.2 µM, and 6.25 µL of Kapa HiFi HotStart ReadyMix 2X (Kapa Biosystems, Wilmington, MA) at 2.5 mM. Amplification was performed in a Bio-Rad thermocycler (Bio-Rad Laboratories, Inc., Hercules, CA) with an initial denaturation (3 minutes at 95 °C) followed by 30 cycles of denaturation (20 seconds at 98 °C), primer annealing (15 seconds at 65 °C), extension (60 seconds at 72 °C) and then a final extension (60 seconds, 72 °C). After amplification, we removed reaction by-products using a 1.5:1.0 Ampure bead-to-sample ratio. A Lonza flash gel was run to confirm product. Two sequencing reactions were completed using a 2 µL DNA input, BigDye® (Life Technologies), and either the forward or reverse universal primer for each amplicon. Once complete, the sequencing reaction was precipitated using sodium acetate followed by a 70% ethanol wash (3 M NaOAc at 1/10th volume for a 300 mM final concentration and 2.5X volumes of EtOH). The DNA was dried down and then resuspended in EDTA and loaded on a 3730 DNA analyzer (Life Technologies). Bases were called from sequence trace files using phred and then assembled against a reference scaffold of the amplicon sequence using phrap (Ewing and Green, 1998; Ewing et al., 1998). Sequence variants were identified by manual review of assemblies and sequence traces in Consed (Gordon et al., 1998). We performed Sanger sequencing as described above on the original 22 tumor samples to validate *Prlr* variants that were called from WGS data and to extend the *Prlr* findings to 10 additional tumors and 35 non-tumor samples (**Figure 1 and Figure S2**). For two tumors in the original discovery set, additional FFPE samples were obtained, and sequenced on a MiSeq Illumina instrument using products of the same PCR protocol described above. Finally we sequenced an additional 9 FFPE samples by MiSeq, to evaluate the presence of Prlr mutations in DCIS tissues.

References

Boutin, J. M., Edery, M., Shirota, M., Jolicoeur, C., Lesueur, L., Ali, S., Gould, D., Djiane, J., and Kelly, P. A. (1989). Identification of a cDNA encoding a long form of prolactin receptor in human hepatoma and breast cancer cells. Molecular endocrinology *3*, 1455-1461.

Ewing, B., and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome research *8*, 186-194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome research *8*, 175-185.

Gordon, D., Abajian, C., and Green, P. (1998). Consed: a graphical tool for sequence finishing. Genome research *8*, 195-202.

Hu, Z. Z., Meng, J., and Dufau, M. L. (2001). Isolation and characterization of two novel forms of the human prolactin receptor generated by alternative splicing of a newly identified exon 11. The Journal of biological chemistry *276*, 41086-41094.

Kline, J. B., Roehrs, H., and Clevenger, C. V. (1999). Functional characterization of the intermediate isoform of the human prolactin receptor. The Journal of biological chemistry *274*, 35461-35468.

Kline, J. B., Rycyzyn, M. A., and Clevenger, C. V. (2002). Characterization of a novel and functional human prolactin receptor isoform (deltaS1PRLr) containing only one extracellular fibronectin-like domain. Molecular endocrinology *16*, 2310-2322.

Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L., and Wilson, R. K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome research *22*, 568-576.

Larson, D. E., Harris, C. C., Chen, K., Koboldt, D. C., Abbott, T. E., Dooling, D. J., Ley, T. J., Mardis, E. R., Wilson, R. K., and Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. Bioinformatics *28*, 311-317.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research *20*, 1297-1303.

Saunders, C. T., Wong, W. S., Swamy, S., Becq, J., Murray, L. J., and Cheetham, R. K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics *28*, 1811-1817.

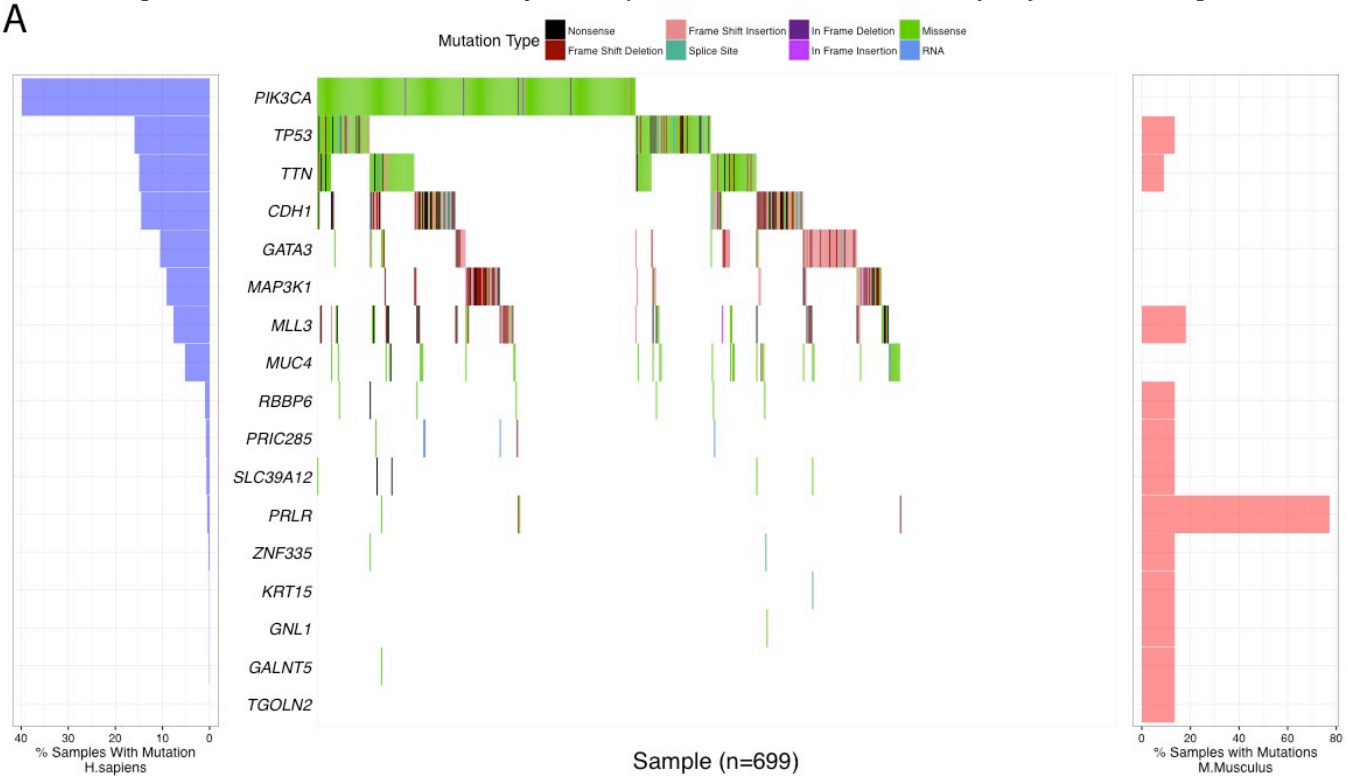
Trott, J. F., Hovey, R. C., Koduri, S., and Vonderhaar, B. K. (2003). Alternative splicing to exon 11 of human prolactin receptor gene results in multiple isoforms including a secreted prolactin-binding protein. Journal of molecular endocrinology *30*, 31-47.

Ye, K., Schulz, M. H., Long, Q., Apweiler, R., and Ning, Z. (2009). Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics *25*, 2865-2871.

Supplemental Figures

Figure S1. Comparison of recurrent mutations in *Stat1-/-* mouse tumors and human TCGA luminal breast cancers. Related to **Table 1, Results** and **Experimental Procedures.** Non-synonymous and RNA Mutations were compared between 699 TCGA luminal breast cancer samples and 22 whole genome sequenced *Stat1-/-* mouse tumor samples. Genes were mapped between species using biomaRt. Genes without a 1 to 1 mapping between *H.sapiens* and *M.musculus* were removed. (A) Genes displayed are recurrent in > 5% (35/699) of TCGA samples or > 13.6% (3/22) samples in the mouse cohort. Mutation types for the human dataset are indicated by color, displaying only the most deleterious mutation (following the order of the legend) per sample. The percent of samples with a mutation are displayed for the TCGA (blue bars) and mouse cohorts (red bars). (B) A Venn diagram displays the number of genes mapped between species with a mutation in any *Stat1-/-* mouse and ≥ 7 human (1%) tumor samples.

A



B

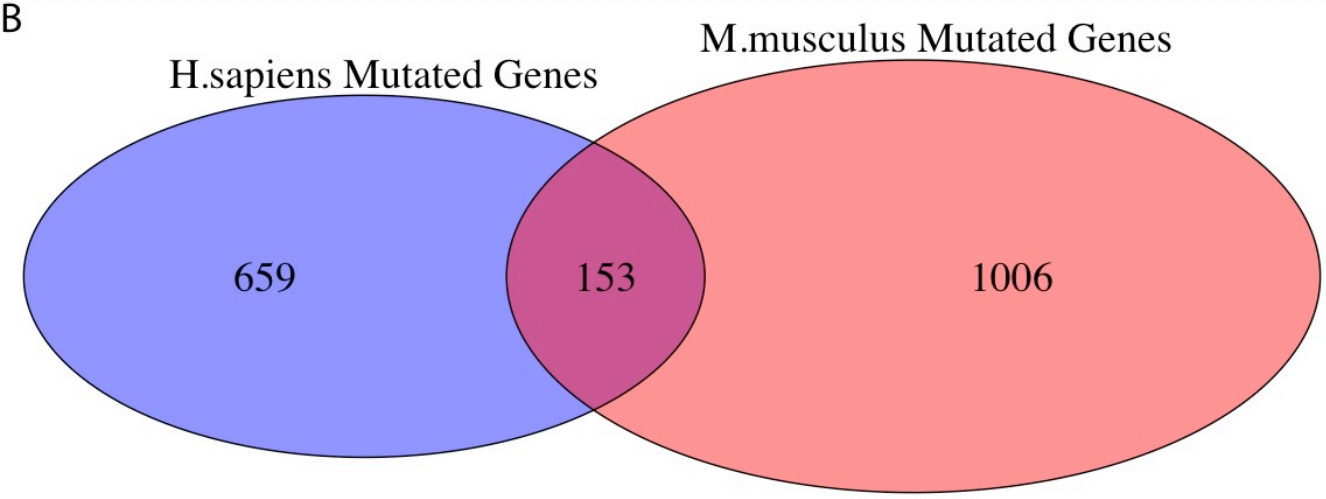


Figure S2. Summary of work flow in the discovery and validation of recurrent *Prlr* mutations. Related to **Figure 1, Results** and **Experimental Procedures.**
Analysis of whole genome sequencing data from discovery samples identified *Prlr* mutations in 21/22 primary *Stat1-/-* mammary tumors. Independent MiSeq or Sanger sequencing validated 22/23 *Prlr* mutations in these tumors. Overall, *Prlr* mutations were observed in 22/22 tumors in the discovery cohort with a single tumor harboring 2 *Prlr* mutations. *Prlr* mutation status was evaluated in the extension cohort using MiSeq or Sanger sequencing. Abbreviations: FFPE, formalin-fixed paraffin embedded; DCIS, ductal carcinoma in situ

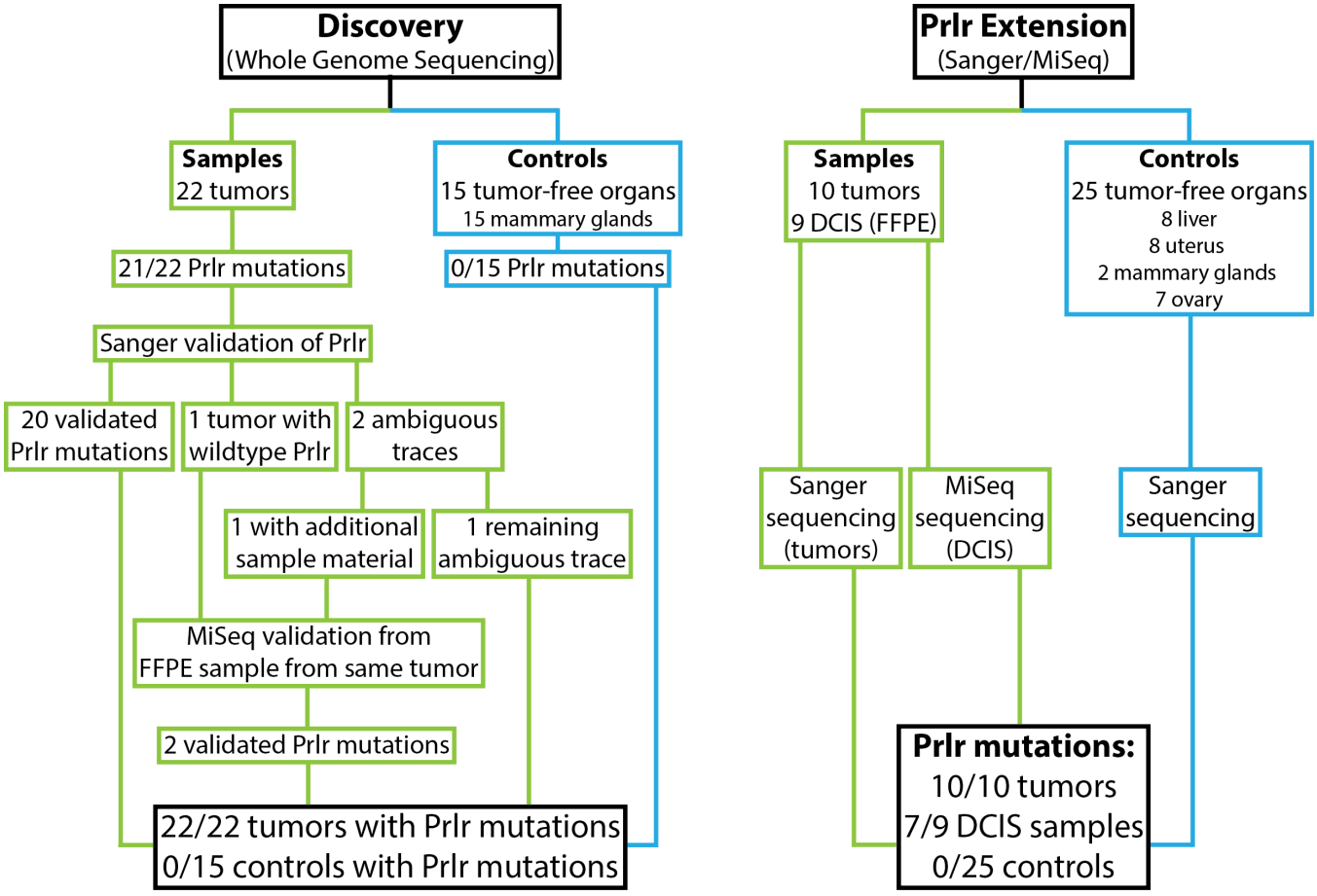


Figure S3. Activation of Stat3 and Stat5 in DCIS from *Stat1-/-* mammary glands. Related to **Figure 3** and **Results.**

Low (a-c) and high (d-f) power view obtained from sub-serial sections of mammary tissue from *Stat1-/-* mice and stained as indicated by labels. Most of the large atypical cells in DCIS displayed pStat3 (b, e) and pStat5 (c, f), as indicated by black arrow heads, whereas no reactivity was observed in normal epithelial cells (indicated by red arrow heads). Magnification 200x (a-c; scale bar 100 μm) and 600x (e-f; scale bar 33 μm).

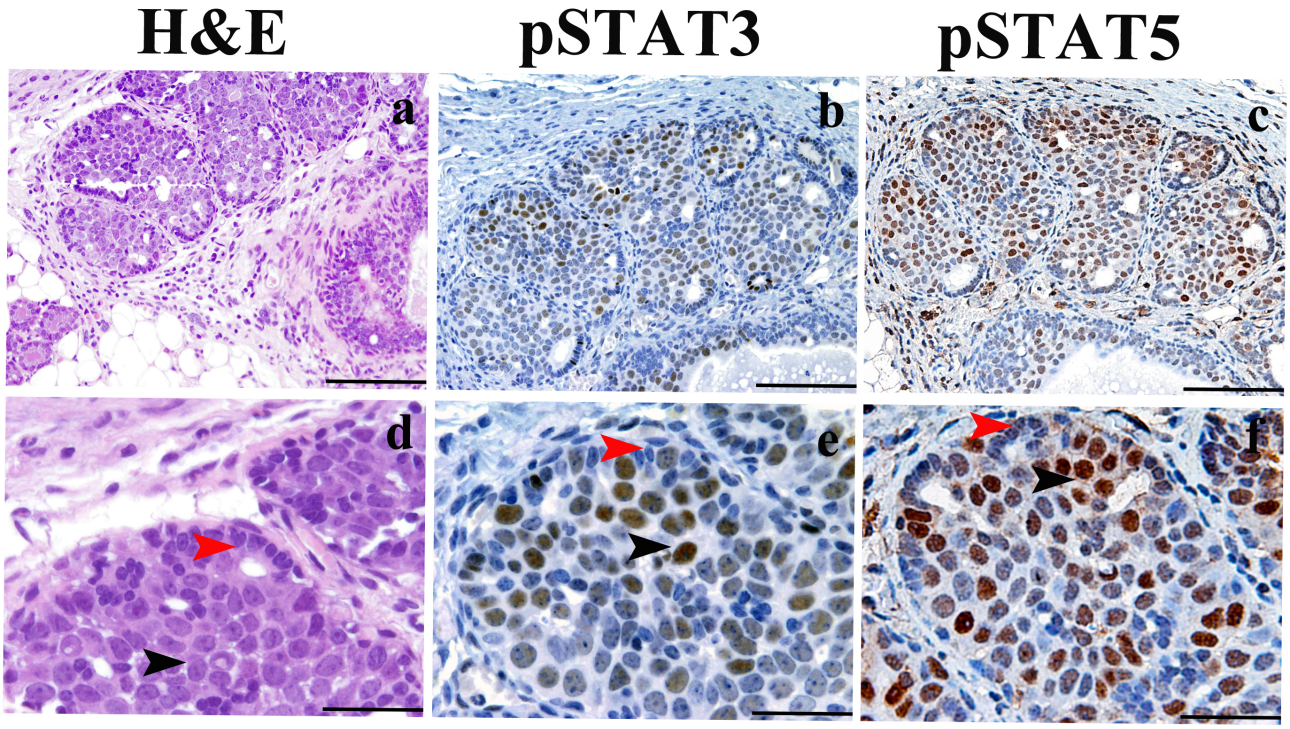


Figure S4. Predicted protein consequence of nonsense and frameshift mutations in *Prlr* for discovery and extension/validation data. Related to **Figure 2, Results, Experimental Procedures,** and **Discussion.**

A clustal omega multiple protein sequence alignment is shown to visualize the effect of *Prlr* truncating mutations. Only residues 301 to 350 of the wild type Prlr are shown based on the reference sequence ENSMUST00000124470 (top row). All residues up to amino acid (aa) position 315 are preserved relative to the reference protein sequence. However, starting at position 316, mutated Prlr proteins are predicted to have either altered amino acid sequences and premature truncation, resulting in predicted protein sizes of 317 to 349 aa compared to the reference full-length sequence of 608 aa.

```
ENSMUST00000124470      DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPGQGVKPTHLDPDSDSGHGSYD 350
TAC302_R_cerv_E316fs    DCEDLLVEFLEVDDN-----AANAIPFQRVSGSRC*------------------------ 330
TAC322_L_tho_L320fs     DCEDLLVEFLEVDDNEDE----------RR*----------------------------- 320
TAC247_L_cerv_L320fs    DCEDLLVEFLEVDDNEDE----------RPKSIRVKVL-----NPHT*------------ 332
OVX3L2_R_tho_L320fs     DCEDLLVEFLEVDDNEDE----------R*------------------------------ 319
OVX6R2_L_cerv_L320fs    DCEDLLVEFLEVDDNEDE----------R*------------------------------ 319
TAC314_L_tho_Y328fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEIRVKVLNPHT*------------ 337
B3R15_L_tho_G330fs      DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPVKVLNPHT*------------ 337
TAC266_L_tho_G330fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPVKVLNPHT*------------ 337
OVX13R1R2_L_ing_G330fs  DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPVKVLNPHT*------------ 337
TAC311_L_cerv_G330fs    DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPVKVLNPHT*------------ 337
B3R1R2L1_R_tho_P329fs   DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYRVKVLNPHT*------------ 337
TAC171_R_tho_P329fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYRVKVLNPHT*------------ 337
TAC270_L_cerv_P329fs    DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYRVKVLNPHT*------------ 337
TAC297_L_tho_P329fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYRVKVLNPHT*------------ 337
TAC300_L_tho_P329fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYRVKVLNPHT*------------ 337
TAC299_L_tho_D343fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPGQGVKPTHLDPDSELWSWKL* 349
SSM3_K334fs             DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPGQGVNPHT*------------ 337
TAC273_L_tho_K334fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPGQGVNPHT*------------ 337
TAC300_R_ing_K334fs     DCEDLLVEFLEVDDNEDE----------RLMPSHSKEYPGQGVNPHT*------------ 337
TAC246_R_ing_Y328*      DCEDLLVEFLEVDDNEDE----------RLMPSHSKE*---------------------- 327
TAC271_R_tho_Y328*      DCEDLLVEFLEVDDNEDE----------RLMPSHSKE*---------------------- 327
TAC301_L_tho_R319fs     DCEDLLVEFLEVDDNEDE----------PKSIRVKV------LNPHT*------------ 331
TAC272_L_tho_E318fs     DCEDLLVEFLEVDDNEDQ----------RVSGSRC*------------------------ 325
TAC312_L_tho_L320fs     DCEDLLVEFLEVDDNEDE----------RVSGSRC*------------------------ 325
TAC319_L_ing_L320fs     DCEDLLVEFLEVDDNEDE----------RVSGSRC*------------------------ 325
TAC270_L_ing_E318fs     DCEDLLVEFLEVDDNEDA---ANAIPFQRVSGSRC*------------------------ 332
TAC298_L_tho_E318fs     DCEDLLVEFLEVDDNEDA---ANAIPFQRVSGSRC*------------------------ 332
TAC268_L_tho_R319fs     DCEDLLVEFLEVDDNEDE---QCHPIPKSIR-------------VKVLNPHT*------- 336
TAC186_L_cerv_E318fs    DCEDLLVEFLEVDDNEDGWLAANAIPFQRVSGSRC*------------------------ 335
TAC273_R_ing_E318fs     DCEDLLVEFLEVDDNEDG*------------------------------------------ 318
SSM2_E319fs             DCEDLLVEFLEVDDNEDELRTSG*------------------------------------- 323
SSM1_E318fs             DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC183_R_tho_E318fs     DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC247_L_cerv_E318fs    DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC269_L_cerv_E318fs    DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC319_R_ing_E318fs     DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC323_L_tho_E318fs     DCEDLLVEFLEVDDNEDSG*----------------------------------------- 319
TAC274_R_ing_E318*      DCEDLLVEFLEVDDNED*------------------------------------------- 317
TAC299_R_tho_E318*      DCEDLLVEFLEVDDNED*------------------------------------------- 317
TAC311_L_ing_E318*      DCEDLLVEFLEVDDNED*------------------------------------------- 317
                        ***************
```

Figure S5. Truncated Prlr is endogenously expressed and co-expression with full-length Prlr leads to colony formation and tumor growth. Related to **Figure 3** and **Results**.
A) Endogenous expression of full-length (FL) and truncated (T) Prlr in *Stat1-/-* mammary cell lines detected by immunoprecipitation and Western blot analysis. B) *Stat1-/-* MEFs were transduced with vector alone (Jak2) or vector expressing full-length (FL/FL-Jak2), truncated (T/T-Jak2) or both full-length and truncated Prlr (FL/T-Jak2). Cell surface expression of Prlr was demonstrated by flow cytometry (blue) with secondary streptavidin-PE alone used as control (gray). C) 50,000 MEFs transduced as indicated were plated in soft agar in triplicate. Colonies were enumerated after 3 weeks. Plates were photographed without magnification and at 50X magnification (inset). D) Nude mice implanted with indicated MEFs or Kras positive control were monitored for tumor growth. Animals were censored when progressively growing palpable tumors of at least 2 mm were detected. On Day 73, all remaining mice were sacrificed and evaluated for evidence of tumors prior to considering animals to be tumor-free. Co-expression of full-length and truncated Prlr in implanted MEFs resulted in significantly lower tumor-free survival than expression of full-length (log-rank test), truncated, vector alone or untransduced *Stat1-/-* MEFs. *p=0.03, **p=0.0005, ***p=0.0001, ****p<0.0001

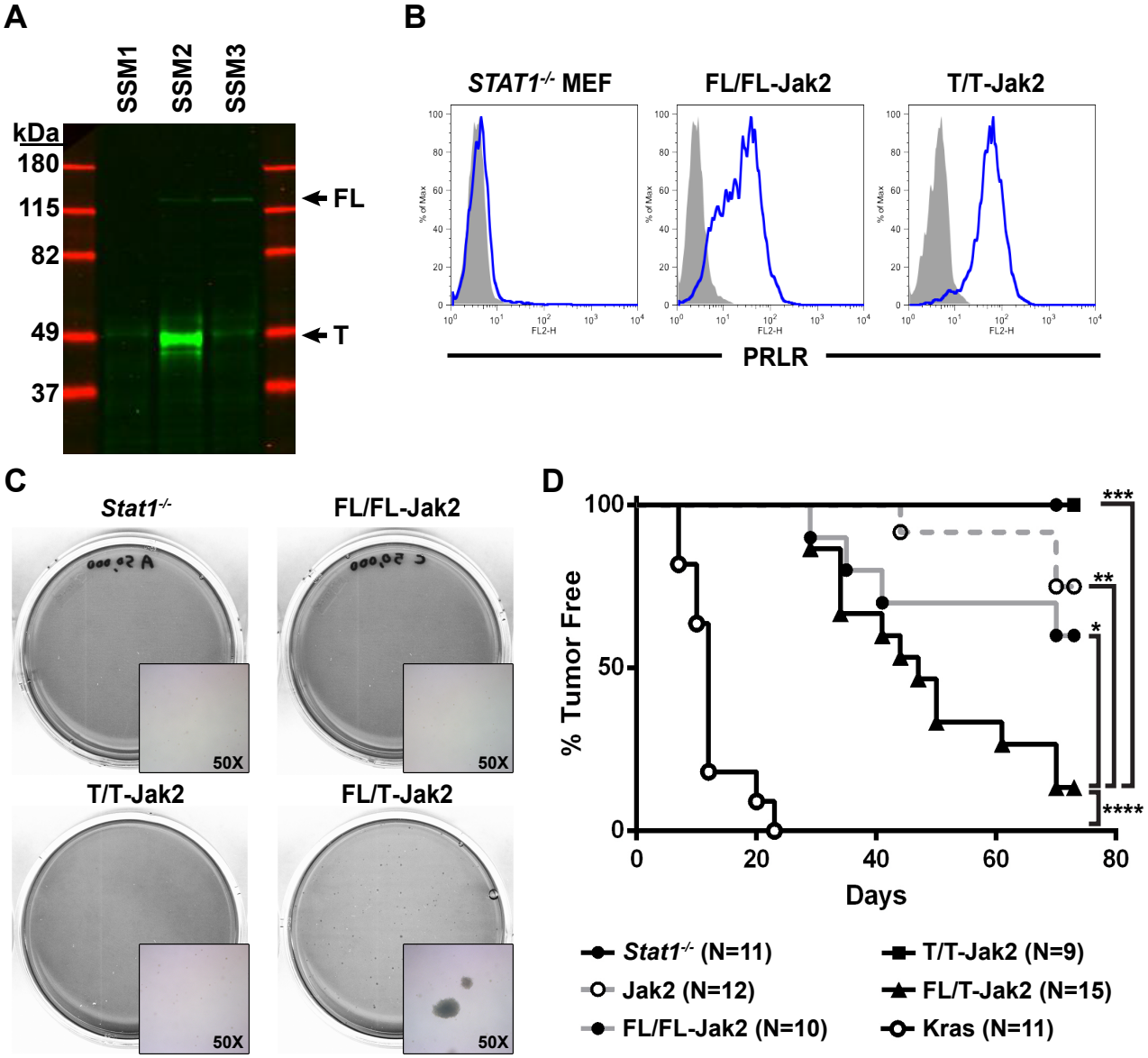
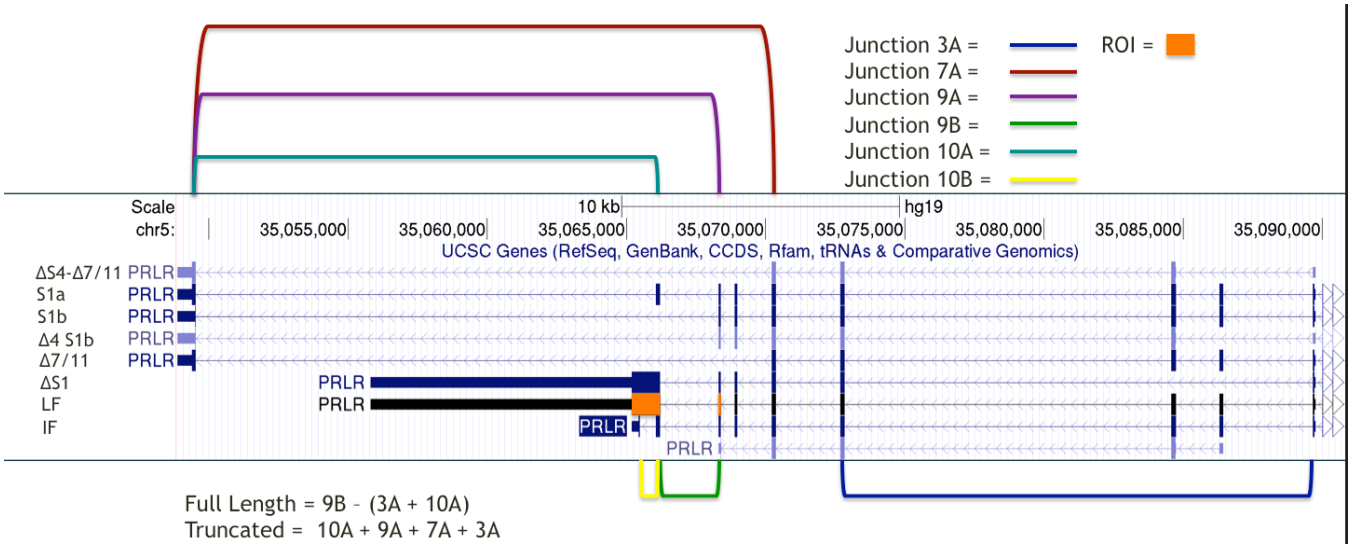


Figure S6. Definition of full-length vs. truncated-isoform ratio using RNA-seq junction data. Related to **Figure 4, Results** and **Experimental Procedures.**
Junction reads per million junction reads mapped (JPM) expression of the Full-Length (FL) and Truncated (T) isoforms of *PRLR* were calculated from raw counts and the ratio was defined as $\log_2((FL + .01)/(T + .01))$(Boutin et al., 1989; Hu et al., 2001; Kline et al., 1999; Kline et al., 2002; Trott et al., 2003). All isoforms could not be uniquely extracted and were simplified as defined below. Junction data was not available for junction 10B. JPM values for 3A and 10A were subtracted from the full-length JPM value to remove any contribution of the truncated isoforms to this calculation. The region of interest (ROI) corresponds to the "long" exon of PRLR and the upstream-most exon.

Junction per Million (JPM) = (Raw Count of Junction)/(Sum of all hg19 Junctions) * 1,000,000
Full-length (FL) = (LF, IF)
Truncated (T) = (S1a, S1b, Δ4 S1b, ΔS4-Δ7/11, Δ7/11, ΔS1)
Region of Interest (ROI) = (chr5:35065191-35066204;hg19 and chr5:35068318-35068387;hg19)



Supplemental Tables

Table S1. SNV/Indel MAF file. Related to **Table 1**, **Results,** and **Experimental Procedures**.

Table S2. Recurrently mutated genes. Related to **Table 1**, **Results,** and **Experimental Procedures**.

Table S3. Significantly mutated genes. Related to **Table 1**, **Results,** and **Experimental Procedures**.

Gene	Tot Muts	Covd Bps	Muts pMbp	P-value FCPT	P-value LRT	P-value CT	FDR FCPT	FDR LRT	FDR CT
Prlr	17	103884	163.64	0	0	0	0	0	0
Olfr1062	5	20944	238.73	5.52E-06	7.77E-12	2.21E-10	0.08566	1.21E-07	3.44E-06
Tgoln1	3	27253	110.08	0.00015	5.25E-10	3.48E-08	1	5.43E-06	0.000360
Taar7e	3	23768	126.22	0.00043	2.42E-09	8.76E-08	1	1.88E-05	0.000680
Esrrg	3	43384	69.15	0.00061	3.85E-09	1.93E-07	1	2.39E-05	0.001037
Gnl1	3	39899	75.19	0.00069	4.65E-09	2.00E-07	1	2.41E-05	0.001037
4930503E14Rik	3	12100	247.93	0.00186	2.17E-08	3.98E-07	1	8.41E-05	0.001767
Galnt5	3	62302	48.15	0.00185	1.97E-08	9.19E-07	1	8.41E-05	0.003569
Anp32b	2	26616	75.14	0.02078	8.60E-07	1.25E-05	1	0.002967	0.035214
4932431P20Rik	4	317999	12.58	0.00574	1.43E-06	1.58E-05	1	0.004442	0.040876
ENSMUSG00000089277	2	2331	858	0.00797	2.41E-06	2.10E-06	1	0.006755	0.007236
4930448N21Rik	2	9203	217.32	0.03802	2.61E-06	3.06E-05	1	0.006755	0.067775
Trp53	3	29673	101.1	0.03601	2.83E-06	2.52E-05	1	0.006759	0.060291
Slc39a12	3	73062	41.06	0.04052	3.28E-06	6.45E-05	1	0.007268	0.125116
Krt15	3	33307	90.07	0.01749	1.13E-05	1.08E-05	1	0.023384	0.033656
Brca1	2	121575	16.45	0.10509	1.57E-05	0.00038	1	0.030387	0.491181

Table S4. Sanger validation results for original WGS Prlr mutation calls. Related to **Table 2,** and **Results**.

Sample Name	Sanger Call	Start	Stop	Variant	Status	Supporting read count
B3R15 L. tho.	het del 1 G chr15:10258182	10258182	10258182	G/0	Validated	4
B3R1R2L1 R. tho.	het del 1 C chr15:10258180	10258180	10258180	C/0	Validated	4
OVX3L2 R. tho.	het del 1 G chr15:10258151	10258151	10258151	G/0	Validated	3
OVX6R2 L. cerv.	het del 1 G chr15:10258151	10258151	10258151	G/0	Validated	2
SSM1	hom del 1 G chr15:10258147	10258147	10258147	G/0	Validated	3
SSM2	het ins 11 TGAGGACGAGC chr15:10258139-10258140	10258139	10258140	0/TGAGGACGAGC	Validated	1
SSM3	het del 1 A chr15:10258195	10258195	10258195	A/0	Validated	3
TAC171 R. tho.	het del 1 C chr15:10258180	10258180	10258180	C/0	Validated	3
TAC183 R. tho.	hom del 1 G chr15:10258147	10258147	10258147	G/0	Validated	1
TAC186 L. cerv.	het ins 7 GATGGCT chr15: 10258147-10258148	10258147	10258148	0/GATGGCT	Validated	2
TAC246 R. ing.	wildtype	N/A	N/A	wildtype	Validated	4
TAC247 L. cerv.	ambiguous indel at	10258154	10258169	TAATGCCATCCCATTC/0	Ambiguous	1

	chr15:10258154					
TAC266 L. tho.	het del 1 G chr15:10258182	10258182	10258182	G/0	Validated	1
TAC268 L. tho.	het del 4 GGCT chr15:10258151-10258154	10258151	10258154	GGCT/0	Validated	2
TAC269 L. cerv.	het del 1 G chr15:10258147	10258147	10258147	G/0	Validated	3
TAC270 L. cerv.	het del 1 C chr15:10258180	10258180	10258180	C/0	Validated	3
TAC270 L. ing.	het del 2 GA chr15:10258147-10258148	10258147	10258148	GA/0	Validated	3
TAC271 R. tho.	ambiguous snp T/A at 10258179	10258179	10258179	T/A	Ambiguous	2
TAC272 L. tho.	het del 23 CGAGCGGCTAATGCCATCCCATT chr15:10258146-10258168	10258146	10258168	CGAGCGGCTAATGCCATCCCATT/0	Validated	1
TAC273 L. tho.	het del 1 A chr15:10258195	10258195	10258195	A/0	Validated	3
TAC273 R. ing.	het del 4 CGAG chr15:10258146-10258149	10258146	10258149	CGAG/0	Validated	1
TAC274 R. ing.	het snp GT chr15:10258147	10258147	10258147	G/T	Validated	2

Table S5. Details of all Prlr mutations. Related to **Figure 2**, **Table 2** and **Results**.

Chr	Start	Stop	Reference	Variant	Type	Sample name	Strand	trv type	c position	amino acid change
15	10258182	10258182	G	0	DEL	B3R15 L. tho.	1	frame_shift_del	c.987	p.G330fs
15	10258180	10258180	C	0	DEL	B3R1R2L1 R. tho.	1	frame_shift_del	c.985	p.P329fs
15	10258151	10258151	G	0	DEL	OVX3L2 R. tho.	1	frame_shift_del	c.956	p.L320fs
15	10258151	10258151	G	0	DEL	OVX6R2 L. cerv.	1	frame_shift_del	c.956	p.L320fs
15	10258147	10258147	G	0	DEL	SSM1	1	frame_shift_del	c.952	p.E318fs
15	10258139	10258140	0	TGAGGACGAGC	INS	SSM2	1	frame_shift_ins	c.944_945	p.E319fs
15	10258195	10258195	A	0	DEL	SSM3	1	frame_shift_del	c.1000	p.K334fs
15	10258180	10258180	C	0	DEL	TAC171 R. tho.	1	frame_shift_del	c.985	p.P329fs
15	10258147	10258147	G	0	DEL	TAC183 R. tho.	1	frame_shift_del	c.952	p.E318fs
15	10258147	10258148	0	GATGGCT	INS	TAC186 L. cerv.	1	frame_shift_ins	c.952_953	p.E318fs
15	10258178	10258186	ATCCGGGTC	0	DEL	TAC246 R. ing.	1	in_frame_del	c.983_991	p.Y328*
15	10258147	10258147	G	0	DEL	TAC247 L. cerv.	1	frame_shift_del	c.952	p.E318fs
15	10258153	10258168	CTAATGCCATCCCATT	0	DEL	TAC247 L. cerv.	1	frame_shift_del	c.958_973	p.L320fs
15	10258182	10258182	G	0	DEL	TAC266 L. tho.	1	frame_shift_del	c.987	p.G330fs
15	10258151	10258154	GGCT	0	DEL	TAC268 L. tho.	1	frame_shift_del	c.956_959	p.R319fs
15	10258147	10258147	G	0	DEL	TAC269 L. cerv.	1	frame_shift_del	c.952	p.E318fs
15	10258180	10258180	C	0	DEL	TAC270 L. cerv.	1	frame_shift_del	c.985	p.P329fs
15	10258147	10258148	GA	0	DEL	TAC270 L. ing.	1	frame_shift_del	c.952_953	p.E318fs

15	10258179	10258179	T	A	SNP	TAC271 R. tho.	1	nonsense	c.984	p.Y328*
15	10258146	10258168	CGAGCGGCTAATGCCATCCCATT	0	DEL	TAC272 L. tho.	1	frame_shift_del	c.951_973	p.E318fs
15	10258195	10258195	A	0	DEL	TAC273 L. tho.	1	frame_shift_del	c.1000	p.K334fs
15	10258146	10258149	CGAG	0	DEL	TAC273 R. ing.	1	frame_shift_del	c.951_954	p.E318fs
15	10258147	10258147	G	T	SNP	TAC274 R. ing.	1	nonsense	c.952	p.E318*
15	10258184	10258184	G	0	DEL	OVX13R1R2 L. ing.	1	frame_shift_del	c.989	p.G330fs
15	10258181	10258181	C	0	DEL	TAC297 L. tho.	1	frame_shift_del	c.986	p.P329fs
15	10258148	10258149	AG	0	DEL	TAC298 L. tho.	1	frame_shift_del	c.953_954	p.E318fs
15	10258147	10258147	G	T	SNP	TAC299 R. tho.	1	nonsense	c.952	p.E318*
15	10258181	10258181	C	0	DEL	TAC300 L. tho.	1	frame_shift_del	c.986	p.P329fs
15	10258194	10258194	T	0	DEL	TAC300 R. ing.	1	frame_shift_del	c.999	p.K334fs
15	10258151	10258169	GGCTAATGCCATCCCATTC	0	DEL	TAC301 L. tho.	1	frame_shift_del	c.956_974	p.R319fs
15	10258142	10258149	AGGACGAG	0	DEL	TAC302 R. cerv.	1	frame_shift_del	c.947_954	p.E316fs
15	10258184	10258184	G	0	DEL	TAC311 L. cerv.	1	frame_shift_del	c.989	p.G330fs
15	10258147	10258147	G	T	SNP	TAC311 L. ing.	1	nonsense	c.952	p.E318*
15	10258222	10258223	0	A	INS	TAC299 L. tho.	1	frame_shift_ins	c.1027_1028	p.D343fs
15	10258151	10258173	GGCTAATGCCATCCCATTCCAAA	-	DEL	TAC312 L. tho.	1	frame_shift_del	c.956_978	p.L320fs
15	10258177	10258177	T	-	DEL	TAC314 L. tho.	1	frame_shift_del	c.982	p.Y328fs
15	10258151	10258173	GGCTAATGCCATCCCATTCCAAA	-	DEL	TAC319 L. ing.	1	frame_shift_del	c.956_978	p.L320fs
15	10258147	10258147	G	-	DEL	TAC319 R. ing.	1	frame_shift_del	c.952	p.E318fs
15	10258154	10258191	TAATGCCATCCCATTCCAAAGAGTATCCGGGTCAAGGT	-	DEL	TAC322 L. tho.	1	frame_shift_del	c.959_996	p.L320fs
15	10258147	10258147	G	-	DEL	TAC323 L. tho.	1	frame_shift_del	c.952	p.E318fs

Based on mouse Ensembl version 67, NCBI build 37.
Prlr gene (ENSMUSG00000005268) and transcript (ENSMUST00000124470).

Table S6. Details of all known isoforms of PRLR according to Ensembl, UCSC and UniProt. Related to **Figure 4** and **Results**.

Table S7. Master sample sheet. Related to **Figure 1** and **Experimental Procedures**.

Table S8. Summary of WGS data quality. Related to **Figure 1** and **Experimental Procedures**.

Sample Name	Gbp Data Produced	Haploid Coverage	Total Reads	Duplicates	Mapped	Read Mapping rate
B3R15 tail	98.4637024	19.8	984637024	206399181	834461996	84.74818392

B3R15 L. tho.	129.468051	38.509	1294680510	117159572	1243606515	96.05508891
B3R1R2L1 tail	93.52087	17.278	935208700	217548506	779246918	83.32331789
B3R1R2L1 R. tho.	108.9314228	31.526	1089314228	80440642	1018346688	93.48511768
OVX3L2 tail	114.6703656	34.562	1146703656	100282677	1104080914	96.28302031
OVX3L2 R. tho.	118.6235328	34.335	1186235328	116885991	1110055169	93.57798936
OVX6R2 tail	116.7272886	29.229	1167272886	139010798	994458447	85.19502671
OVX6R2 L. cerv.	115.409363	32.942	1154093630	125139455	1077020815	93.32178837
SSM1	148.8810846	41.595	1488810846	205420313	1427896794	95.90854324
SSM2	134.3889544	38.475	1343889544	176826882	1293006018	96.21371219
SSM3	133.2712376	38.537	1332712376	159271169	1282370435	96.22259522
TAC171 R. tho.	124.5400856	37.301	1245400856	75625321	1173607476	94.23531952
TAC183 R. tho.	118.733256	34.326	1187332560	78295285	1106184358	93.16550352
TAC186 L. cerv.	122.5908556	36.411	1225908556	101466579	1156960781	94.37578157
TAC236 MG	67.177392	18.74	671773920	118418783	649717662	96.71671416
TAC237 MG	75.1392458	21.95	751392458	85646916	723535395	96.29260812
TAC238 MG	71.892757	20.2	718927570	123647626	697035285	96.95486918
TAC239 MG	55.7686952	16.385	557686952	79434218	541406240	97.08067188
TAC240 MG	60.7934208	17.749	607934208	89555920	590537505	97.13839051
TAC241 MG	85.466625	23.923	854666250	132107228	820859373	96.04443524
TAC242 MG	81.153794	21.825	811537940	147192188	772615038	95.20380994
TAC243 MG	70.8781072	19.51	708781072	129199592	687502699	96.99789204
TAC244 MG	67.776858	19.095	677768580	95026180	645439091	95.23001066
TAC245 MG	72.7946258	20.396	727946258	135899907	709284998	97.43645087
TAC246 R. ing.	125.3001016	35.808	1253001016	135434601	1191634543	95.10244028
TAC247 L. cerv.	127.9514356	37.301	1279514356	123751413	1216319815	95.0610526
TAC263 MG	138.047999	32.443	1380479990	239815859	1215651175	88.06003592
TAC263 tail	124.9547678	27.638	1249547678	217373427	1084998042	86.83126391
TAC264 MG	159.1427846	35.313	1591427846	299690249	1336302140	83.96875443
TAC264 tail	118.6124524	27.618	1186124524	251930780	1088694054	91.78581439
TAC265 MG	159.616026	37.612	1596160260	280650416	1404563681	87.99640714
TAC265 tail	133.2096788	24.884	1332096788	269070385	1035520810	77.73615396
TAC266 MG	224.6979652	48.6	2246979652	689278121	2109590342	93.88560062
TAC266 tail	187.7547706	42.292	1877547706	368155435	1653287121	88.05566515
TAC266 L. tho.	229.9043578	66.479	2299043578	224531078	2150781067	93.5511222
TAC268 MG	120.1834742	37.837	1201834742	80105347	1172453545	97.55530474
TAC268 tail	118.9858064	37.689	1189858064	65578870	1153901043	96.97804116
TAC268 L. tho.	112.4953512	33.684	1124953512	112005364	1086691684	96.59880808
TAC269 tail	104.9254376	29.111	1049254376	176108396	1011852849	96.43541854
TAC269 L. cerv.	101.7742068	30.929	1017742068	102614572	987654983	97.04374164
TAC270 L. cerv.	117.0297188	36.866	1170297188	68941240	1133862613	96.88672455
TAC270 L. ing.	120.442782	36.78	1204427820	93581306	1160065553	96.31673511
TAC270 tail	117.3161028	35.316	1173161028	107097272	1130809356	96.38995236
TAC271 tail	103.2880992	30.202	1032880992	94912847	973406841	94.2419164
TAC271 R. tho.	116.0329936	32.795	1160329936	122977932	1081360230	93.19420248

TAC272 tail	112.864483	33.928	1128644830	69371213	1056131141	93.57515428
TAC272 L. tho.	111.2546792	32.022	1112546792	115265413	1046458947	94.0597694
TAC273 L. tho.	122.280144	37.669	1222801440	73249513	1156665784	94.59146401
TAC273 R. ing.	123.6560642	37.616	1236560642	82326960	1161400870	93.9218693
TAC273 tail	89.7288922	22.954	897288922	150900438	819137924	91.29031953
TAC274 R. ing.	112.644696	33.934	1126446960	89913333	1071956470	95.16262266
TAC274 tail	92.3263574	26.762	923263574	106680274	878420409	95.14297257

Table S9. Summary of CNV calls. Related to **Experimental Procedures**.

Table S10. Primers for amplification and sequencing of Prlr truncation target region. Related to **Figure 2** and **Table 2**, and **Experimental Procedures**.

Primer details		Genomic location and sizes (mm9; C57BL/6J)				
Name	Sequence (5' to 3'; M13 sequencing tails in lower case)	Chr	Start	End	Product size	Amp size
M_15_02auf_1	tgtaaaacgacggccagtGAAGAACTGCTGAGTGCCTTGG	15	10,258,048	10,258,069	308	228
M_15_02auf_2	caggaaacagctatgaccGTGATCTCAGGGATGTGGAAGG	15	10,258,298	10,258,319		
M_15_02aug_1	tgtaaaacgacggccagtGGACTTGCTGGTGGAGTTCTT	15	10,258,104	10,258,124	155	76
M_15_02aug_2	caggaaacagctatgaccCACTGTCAGGATCTAGGTGTGT	15	10,258,201	10,258,222		

Note: product sizes without M13 tails are 272 and 119 respectively.

Supplemental Data Files

Appendix 1. IGV screenshots of Prlr mutations from the 22 discovery tumors. Related to **Figure 2**, **Table 2**, and **Results.**

Appendix 2. Sanger traces for Prlr validation of WGS findings. Related to **Figure 2**, **Table 2**, and **Results.**

Appendix 3. Sanger traces for Prlr extension to additional tumor and non-tumor samples. Related to **Figure 2**, **Table 2**, and **Results.**

Appendix 4. Screenshots from IGV for Prlr validation of two WGS samples from FFPE and extension to DCIS samples by MiSeq. Related to **Figure 2**, **Table 2**, and **Results.**