

## Supplementary Table 1

### Mouse X-linked paralogous gene families with >85% protein identity

#### Ampliconic Genes

Gene Family Symbol	Gene Family Name	ENSMBL Gene ID	Comments
<i>Gmcl1</i>	<i>Germ cell-less homolog 1 (Drosophila)-like</i>	ENSMUSG00000073300	
<i>EG668965</i>		ENSMUSG00000073098	
<i>Ssxb</i>	<i>Synovial sarcoma, X member b</i>	ENSMUSG00000023165	
<i>Fthl17</i>	<i>Ferritin, heavy chain polypeptide-like 17</i>	ENSMUSG00000058160	
<i>Slx*</i>	<i>Sycp3 like X-linked</i>	ENSMUSG00000073253	
<i>Rhox</i>	<i>Reproductive homeobox gene</i>	ENSMUSG00000071772	
<i>4930567H17Rik</i>		ENSMUSG00000073141	
<i>Magea*</i>	<i>Melanoma antigen, family a</i>	ENSMUSG00000055746	
<i>Xlr</i>	<i>X-linked lymphocyte regulated</i>	ENSMUSG00000073125	
<i>EG434797</i>		ENSMUSG00000067771	
<i>Mageb</i>	<i>Melanoma antigen, family b</i>	ENSMUSG00000073069	
<i>Pabpc1s</i>	<i>Poly(A) binding protein, cytoplasmic 1-like 2A</i>	ENSMUSG00000046639	
<i>Tgif2lx</i>	<i>TGFB-induced factor homeobox 2-like, X-linked</i>	ENSMUSG00000061283	
<i>Pramel3</i>	<i>Preferentially expressed antigen in melanoma-like 3</i>	ENSMUSG00000067360	
<i>Ott*</i>	<i>Ovary-testis transcribed</i>	ENSMUSG00000072924	
<i>LOC207318</i>		ENSMUSG00000067218	
<i>4921511M17Rik*</i>		ENSMUSG00000067208	

#### Non-ampliconic Genes

<i>1700012L04Rik</i>		ENSMUSG00000073280	
<i>LOC278181</i>		ENSMUSG00000071798	
<i>LOC245376</i>		ENSMUSG00000068113	
<i>1700020N15Rik</i>		ENSMUSG00000059690	
<i>1700042B14Rik</i>		ENSMUSG00000057402	
<i>1700003E24Rik</i>		ENSMUSG00000067646	
<i>1700010D01Rik</i>		ENSMUSG00000071726	
<i>MGC58426</i>		ENSMUSG00000050435	
<i>4930524N10Rik</i>		ENSMUSG00000059663	

#### Excluded Gene Families

<i>60S Ribosomal-like</i>		ENSMUSG00000063836	Ribosomal repeat
<i>Activated spleen cDNA</i>		ENSMUSG00000073254	Retroviral repeat
<i>Ambiguous</i>		ENSMUSG00000070641	Not clustered
<i>Beta thymosin-like</i>		ENSMUSG00000072954	No methionine start
<i>Cxx1a/b/c</i>	<i>CAAX box 1 homolog a/b/c</i>	ENSMUSG00000067924	No detectable expres.
<i>HERV envelope-like</i>		ENSMUSG00000073275	Retroviral repeat
<i>3110007F17Rik</i>		ENSMUSG00000067402	No methionine start
<i>Phosphatase Regulatory Subunit11</i>		ENSMUSG00000073306	Not clustered
<i>Retrovirus peptidase</i>		ENSMUSG00000073149	Retroviral repeat
<i>Ribosomal protein L7a</i>		ENSMUSG00000046453	Ribosomal repeat
<i>Rps12*</i>	<i>Ribosomal protein S12</i>	ENSMUSG00000071696	Ribosomal repeat
<i>Smpsb</i>	<i>Similar to polypyrimidine tract-binding protein</i>	ENSMUSG00000068274	Inv. Rep. arms <8kb
<i>Tceal</i>	<i>Transcription elongation factor A (SII)-like</i>	ENSMUSG00000044550	Not clustered
<i>Ambiguous</i>		ENSMUSG00000068097	Autosomal match

(Supplementary Table 1 continued)

Clustered X-linked paralogs sharing >85% amino acid identity. Paralogs from the mouse X chromosome were retrieved using Ensembl's biomaRT tool. Paralogs were discarded if at least a single pair of gene family members did not cluster within one Mb of each other, contained >90% identity to autosomes, or have annotation concerns (e.g. lacks a methionine start). Five new ampliconic regions (indicated with an asterisk) were identified via this method. Fourteen of the 17 IRF-identified ampliconic regions were also detected; the three not present (Amp11, Amp13, and Amp14) were not detectable because only a single gene family member is annotated in Ensembl's set of protein-coding genes. The four novel/predicted genes (*Zfp161*, *ZxdB*, *LOC665542*, and *LOC207318*) identified in this study (Supplementary Table 4) are not shown here as they were also not present in the Ensembl annotation. Though *Rps12* is a ribosomal repeat, which is not expressed, it enabled the detection of a tandemly arrayed ampliconic region containing the *LOC665542* and *Srsx* genes.