Supplementary Table 2

X-ampliconic regions size estimates and breakpoint association

	Map position (Mb)	No. of RP-23 BES	No. of RP-24 BES	Predicted size from RP-23 BES	Predicted size from RP-24 BES	Average predicted region size (Mb) based upon RP-23 and RP-24 BES counts	# of between contig physical gaps	# of within contig physical gaps
Amp1	3-4.85	188	60	4.02	3.9	3.96	0	4
Amp4	23.1-29.8	192	82	4.11	5.32	4.72	9	0
Amp5	33.65-34.40	N/A	N/A	N/A	N/A	N/A	1	0
Amp6	49.77-49.89	N/A	N/A	N/A	N/A	N/A	1	0
Amp7	50.3-52.2	127	32	2.72	2.08	2.4	2	0
Amp17	119.5-121.6	101	29	2.16	1.88	2.02	2	1
Amp19	143-145.7	158	65	3.38	4.22	3.8	0	1

BAC-end sequence (BES) frequencies from a non-ampliconic region of the mouse X chromosome were used to estimate five massive X-ampliconic regions. The 10Mb (55Mb-65Mb) non-ampliconic control region contained 46.7 BES per Mb and 15.4 BES per Mb for the RP-23 and RP-24 libraries, respectively. Ampliconic regions Amp5 and Amp6 are contiguously assembled and therefore do not have BES counts or region size estimates. Twenty-four X-linked physical gaps (either within or between contigs) could be identified in the *Mus musculus* NCBI Build 37.1. Two pseudoautosomal region contigs were not included in this estimate. The majority (21) of the 24 gaps can be explained by their location within ampliconic regions. Nine of the 21 amplicon-associated physical gaps can be found within Amp4 (Figure 1a), which is likely to be due both to its large size (~4.7Mb) as well as to the difficulties in assembling regions with smaller repeat units (34.3kb and 25.7kb). Thus, even though the mouse X chromosome sequence was assembled using a BAC-based approach from an inbred C57BL/6J strain, the massive amplification of repeat units with high levels of identity within these ampliconic regions have led to their incomplete assembly.