**SUPPLEMENTARY MATERIAL LEGENDS**

**Table S1.** Samples included in this study. The table provides the following information:

Serial ID.

Sample name: name provided by the investigator.

Type: inbred, F1, cell line, cross or unclassified.

Content Type: initial or Final.

Consensus strain: if a sample was used to build the consensus genotypes of one 241 inbred strain, that strain name is listed, if that sample was not used then zero.

Chromosomal sex marker selection: TRUE for samples used in selecting sex informative markers. FALSE for samples not used.

Chromosomal sex: XX, XY, X0, XXY or XX\*. The latter group are XX samples misclassified as XO.

Replicate: TRUE for technical replicates genotyped more than once. FALSE for samples genotyped only once.

Replicate name: An unambiguous name for that group of replicate samples.

Array\_calibration\_sample: Samples used for array calibration in the preliminary and production phase of the array

X chromosome intensity: median normalized intensity of chromosome X sex-informative markers.

Y chromosome intensity: median normalized intensity of chromosome Y sex-informative markers.

Median autosomal intensity: median intensity (r) of autosomal markers. Used as a normalization factor in chromosomal sex determination.

mean\_autosome\_xyraw: mean of the raw autosomal intensities (xraw + yraw) for the sample. Used as a normalization factor in construct detection.

H calls: Number of heterozygous calls for tier 1 and 2 markers (see below) in the autosomes and chromosome X.

H call on chromosome X: Number of heterozygous calls for tier 1 and 2 markers (see below) on chromosome X.

Autosomal N calls: Number of no calls for tier 1 and 2 markers (see below) in the autosomes.

N calls on chromosome X: Number of no calls for tier 1 and 2 markers (see below) in the X chromosome.

ks\_stat: Kolmogorov-Smirnov goodness of fit test statistic of the sample’s autosomal intensities against the autosomal intensity distribution of 200 random samples using markers from the preliminary array.

pd\_stat: Pearson's chi-squared test statistic of the sample’s autosomal intensities against the autosomal intensity distribution of 200 random samples using markers from the preliminary array.

Negative\_construct\_control: Indicates whether the sample was used as a negative control for construct analysis

BlastR: Sum of the autosome-normalized xraw intensity at 6 markers used to declare the presence or absence of the construct Blasticidin resistance.

Cas9: Sum of the autosome-normalized xraw intensity at 7 markers used to declare the presence or absence of the construct Cas9

Cre: Sum of the autosome-normalized xraw intensity at 15 markers used to declare the presence or absence of the construct Cre recombinase

DT: Sum of the autosome-normalized xraw intensity at 11 markers used to declare the presence or absence of the construct Diptheria toxin

IRES: Sum of the autosome-normalized xraw intensity at 6 markers used to declare the presence or absence of the construct Internal Ribosome Entry Site

Luc: Sum of the autosome-normalized xraw intensity at 10 markers used to declare the presence or absence of the construct Luciferase

SV40: Sum of the autosome-normalized xraw intensity at 18 markers used to declare the presence or absence of the construct SV40 large T antigen

bpA: Sum of the autosome-normalized xraw intensity at 8 markers used to declare the presence or absence of the construct Bovine growth hormone poly A signal sequence

chlor: Sum of the autosome-normalized xraw intensity at 9 markers used to declare the presence or absence of the construct Chloramphenicol acetyltransferase

g FP: Sum of autosome-normalized xraw intensity at 19 markers used to declare the presence or absence of the construct "Greenish" Fluorescent Protein (EGFP, EYFP, ECFP)

hCMV a: Sum of the autosome-normalized xraw intensity at 6 markers used to declare the presence or absence of the construct hCMV enhancer version a

hCMV b: Sum of the autosome-normalized xraw intensity at 11 markers used to declare the presence or absence of the construct hCMV enhancer version b

hTK pr: Sum of the autosome-normalized xraw intensity at 2 markers used to declare the presence or absence of the construct Herpesvirus TK promoter

iCre: Sum of the autosome-normalized xraw intensity at 8 markers used to declare the presence or absence of the construct iCre recombinase

r FP: Sum of the autosome-normalized xraw intensity at 5 markers used to declare the presence or absence of the construct "Reddish" fluorescent protein (tdTomato, mCherry)

rtTA: Sum of the autosome-normalized xraw intensity at 8 markers used to declare the presence or absence of the construct Reverse improved tetracycline-controlled transactivator

tTA: Sum of the autosome-normalized xraw intensity at 14 markers used to declare the presence or absence of the construct Tetracycline repressor protein

Refined background analysis: Samples used for the refined background analysis shown in Figure S9.

**Table S2.** Marker annotation. The table contains the following information:

1) Marker name.

2) Chromosome. The following types are allowed: 1-19, for the autosomes; X and Y for the sex chromosomes; PAR, for markers on the pseudoautosomal region; MT, for the mitochondria and 0, for genetic constructs.

3) Position in bases in build 38.

4) Strand. +, indicating the probe sequence is found in the 5' to 3' order (on the forward strand) in the reference genome immediately preceding the variant. -, indicating that the reverse complement of the probe sequence is found in the 5' to 3' order (on the forward strand) in the reference genome, immediately following the variant and NA, when not available.

5-6) Sequences A and B. Sequence A for one bead probes is the sequence of the marker probe without the SNP and for two bead probes, the sequence of the marker probe including the SNP. Sequence B: for one bead probes, not applicable; for two bead probes, the alternative sequence of the marker probe including the SNP.

7-8) Reference Allele and Alternate allele. Columns denoting the genotype call for the reference and alternative alleles

9) Tier. For biallelic SNP markers, tier was assigned based on observed genotype call types (homozygous reference, homozygous alternate, or heterozygous) at each marker across a set of 3,878 samples used for array QC and validation. Tier 1 markers were those for which we observe all three call types. Tier 2 markers were those for which we observe two of the three call types. Tier 3 markers were those for which we observe only one call type. Tier 4 markers were those markers for which we observe no calls (N) in every sample. For construct markers, tier is assigned based on the capability of a marker to detect a given construct. Informative tier makers are those for which the marker has been validated to test for the presence or absence of a given construct based on intensity. Partially informative tier makers were those for that could potentially be used to test for the presence or absence of a given construct based on intensity. Those markers which have not been tested were assigned the tier “Not tested”.

10) rsID.

11) Diagnostic. Name of the construct, substrain or strain group that the maker is diagnostic for. In all other cases is empty.

12) Diagnostic type. Substrain, strain group or construct.

13) Diagnostic information: Abbreviated name of the construct, name of substrain or list of substrains in which we observed the diagnostic allele. In all other cases is empty.

14) Partial diagnostic: 1, for diagnostic alleles that are not fixed. 0, in all other cases.

15) Diagnostic allele. Whether the reference or alternative allele is the diagnostic

16) Positive threshold. Threshold value to declare the presence of a given construct

17) Negative threshold. Threshold value to declare the absence of a given construct.

18) Uniqueness measured using Bowtie.

19) X chromosome markers used to determine the presence and number of X chromosomes. 1, chromosome X markers used in sex chromosome determination. 0, in all other cases.

20) Y chromosomes markers used to determine the presence of a Y chromosome. 1, chromosome Y markers used in sex chromosome determination. 0, in all other cases.

21) Flags. SPIKE, markers added in the final iteration of the array. Empty in all other cases.

22) Diagnostic Birth. The population where a diagnostic allele was first seen (E2, E3, E4 in the BxD, CC for the Collaborative Cross, or Consensus for the MiniMUGA inbred consensus)

23) Diagnostic Fixation. The population where a diagnostic allele is inferred to be fixed (E3, CC, Consensus, or Not Fixed (segregating))

**Table S3.** List of inbred strains with consensus genotypes grouped into four classes: classical, wild-derived, CC and BXD. Number of samples genotyped to build the consensus.

**Table S4.** Examples of the rules for consensus genotypes calls. \*, denotes the diagnostic allele.

**Table S5.** Consensus genotypes for the strains listed in Table S3. The table provides the marker name, chromosome, position, and consensus genotypes for 10,819 biallelic SNPs. Construct markers are not included in this table.

**Table S6.** Construct probe design annotation. The table includes the following information: marker names, number of probes assigned to a given construct, initial target construct (Target.Construct); probes passing our pipeline (Good.Probes), unique probe sequences (Ind.Probes), Construct ID in the report (Report.Code), and initial sequence accessions used to design probes (Seed Sequences).

**Table S7.** Samples with sex chromosome aneuploidy, mosaicism, and misclassified chromosomal sex. The table provides sample ID, sample name, chromosomal sex, mosaic status, error in chromosomal sex determination, parental origin of the sex chromosome nondisjunction, rationale for assignment, dam, sire, type, sample\_x\_intensity, sample\_y\_intensity, median\_autosomal\_intensity, sample\_h\_calls, sample\_n\_calls, H\_calls\_on\_X, ks\_stat, and pd\_stat.

**Table S8.** Number of SNPs with N and not N genotype calls in the autosomes and X chromosome in the F1 male with sex chromosome aneuploidy and mosaicism (TL9348). See Figure 3.

**Table S9.** The table provides the two parental strains of a RCC, the strain group to which they belong, and the fraction of the genome covered at different linkage disequilibrium metrics.

**Figure S1.** Sex effect on normalized intensity for markers on chromosome X. The left panel represents the range of intensities (mean +/- standard deviation) at 269 markers considered informative based on the lack of overlap between the distribution of intensities in males (blue) and females (red). The right panel represents the range of intensities (mean +/- standard deviation) at 426 makers that are not considered sex informative.

**Figure S2.** Intensities at construct markers present in MiniMUGA. a) For each marker negative controls are shown in black and experimental samples including potential positive controls are shown in blue. Markers are grouped according to construct. Marker names are colored based on their performance: black, validated markers; blue, markers excluded due to lack of intensity range and discrimination; purple, markers excluded because of lack of discrimination between negative controls and experimental samples; and red, markers excluded due to lack of high correlation with any other markers from the same construct. b) Heatmap of the correlation values between all black and red construct markers.

**Figure S3.** Alignments of validated construct markers. For each construct the file provides a short summary, the alignment of the working probes, the target DNA and protein sequences. The alignment of forward (black) and reverse (blue) probes is shown with the nucleotide used for “genotyping” (A) shown in red background for forward probes and in blue (T) for reverse probes. Mismatches are shown in purple background.

**Figure S4.** Examples of normal and abnormal intensity distributions**.**  Intensity distributions for six samples with low pd\_stat and six samples with high pd\_stat on the autosomes and chromosome X. Colored histogram bars are the intensity values distribution on the corresponding chromosome. Colored lines are the kernel density estimates for these data. Black lines are an attempt to fit the actual data to a normal curve.

**Figure S5.** Threshold determination for chromosomal sex using pd\_stat. The distribution of pd\_stat values in the 6,899 samples is shown on the y axis. The pd\_stat is Pearson's chi-squared test statistic of the sample’s autosomal intensities against the autosomal intensity distribution of 200 random samples. The x axis shows the ks\_stat for better contrast. The ks\_stat is Kolmogorov-Smirnov goodness of fit test statistic of the sample’s autosomal intensities against the autosomal intensity distribution of 200 random samples. Samples in yellow were incorrectly identified as XO but are in fact XX (aka XX\*, Tables S1 and S6). Samples in green are from mouse species other than *Mus musculus*. Samples in blue are labeled Aneuploid by our algorithm. We manually established a threshold to capture all the misclassified samples and samples from other species.

**Figure S6.** Chromosome Y duplications. Spatial distribution of normalized intensity at SNPs in the proximal end of the Y chromosome in C3H/He, DBA/1 and C57BL/6 samples. The range of intensities are shown in orange in cases where we had multiple samples with the duplication while samples with normal Y chromosome are shown in blue. Duplicated regions are shown in red and transition regions with uncertain copy number are shown in pink. The bottom of the figure shows the location of the MiniMUGA markers and genes.

**Figure S7.** Inbreeding thresholds. The figure shows in red the distribution of observed H calls in 385 samples representing 85 classical inbred strains. It also shows in blue the distribution of predicted number of H calls in 3,655 F1 hybrids using the consensus genotypes from 86 classical inbred strains. Tier 1 and 2 markers on the autosomes, X chromosomes and PAR were used. Thresholds for inbred, close to inbred and outbred are shown as vertical bars.

**Figure S8**. MiniMUGA Background Analysis Report for the following four female cell lines: C2Cl2, GPG C3-Tag-T1-Luc, MLE12, and C57BL/6J.

**Figure S9.** Relationship between reported and refined background analyses.Each dot represents one of 107 samples with primary and secondary backgrounds identified in the initial report (>99.8% genome explained). The x axis is the percentage of the genome assigned to the secondary background in the reported initial analysis. The y axis is the percentage of the genome assigned to the secondary background in a refined analysis which includes only markers that are informative between the primary and secondary backgrounds and any markers where the sample in question has a different genotype call. The analysis is based on the number of markers. The plot shows a tight linear regression line fit with a slope of ~2.6. This multiplier can be used to generate a more realistic estimate of the true contributions to the genetic background of a sample.

**Figure S10.** Age and breeding history of four mouse samples from the B6.129-Nox4tm1kkrJ congenic line maintained through breeding at UNC. Green triangles note the position of the generate allele. Red bars denote the ancestral allele for diagnostic SNPs fixed by E3 in the BXD panel. Pink bars denote ancestral alleles for diagnostic SNPs fixed by the start of the CC. Light blue bars denote diagnostic alleles at diagnostic SNPs fixed by E3. Lighter blue bars denote diagnostic alleles at diagnostic SNPs fixed by the start of CC. Grey bars denote ancestral alleles at post-CC diagnostic SNPs. Dark blue bars denote diagnostic alleles at post-CC diagnostic SNPs. Split bars denote heterozygosity.

**Figure S11**. *De novo* X chromosome duplication. The range of intensities for females and males are shown in pink and blue, respectively. The sample with the duplication is shown as black line. Genotypes for the parental CC strains and the test sample are shown at the bottom as well as the first marker included in the duplication (asterisk) and the extent.