

Contig Nx plot for 7 rat reference genomes

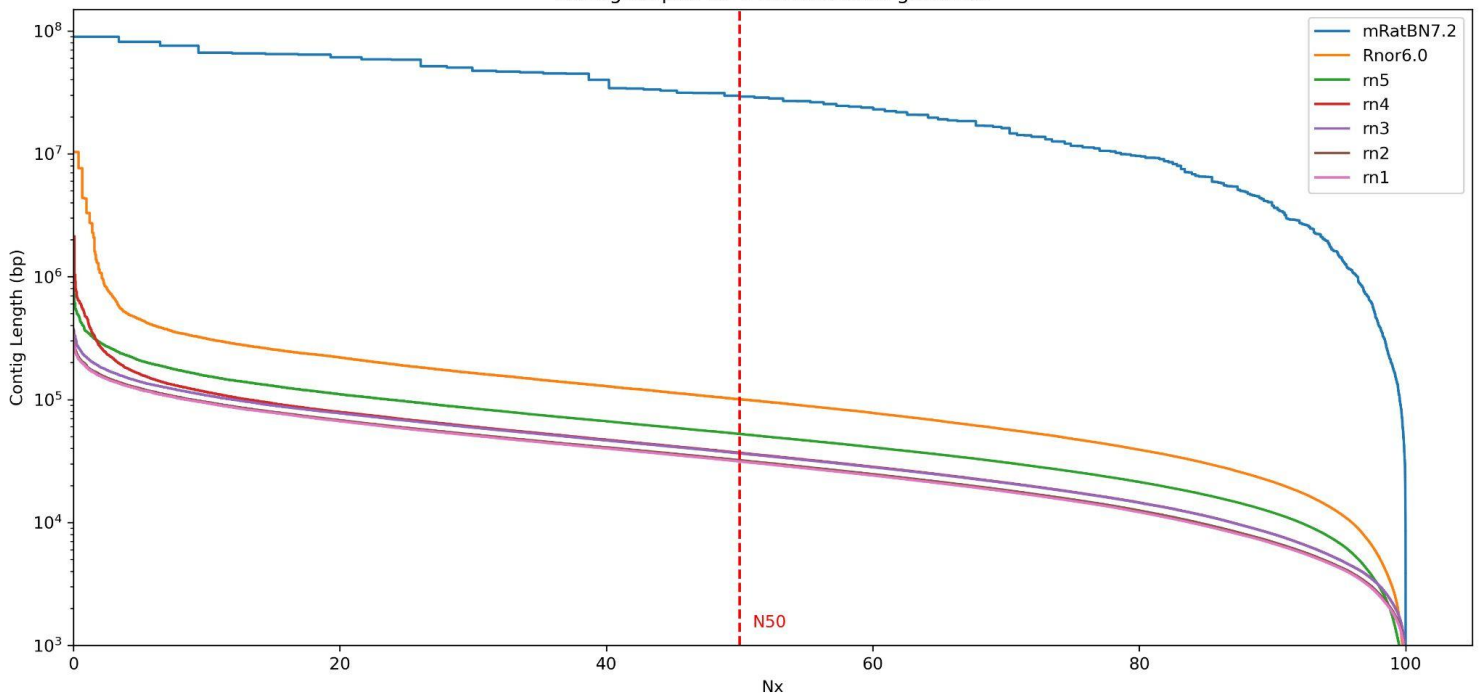


Figure S1. The Contig Nx plot for mRatBN7.2 and 6 prior rat reference genomes. The red dashed line indicates the contig N50 values. Small improvements in contig continuity are observed for the previous updates, with the most significant improvement (~290X) coming from mRatBN7.2

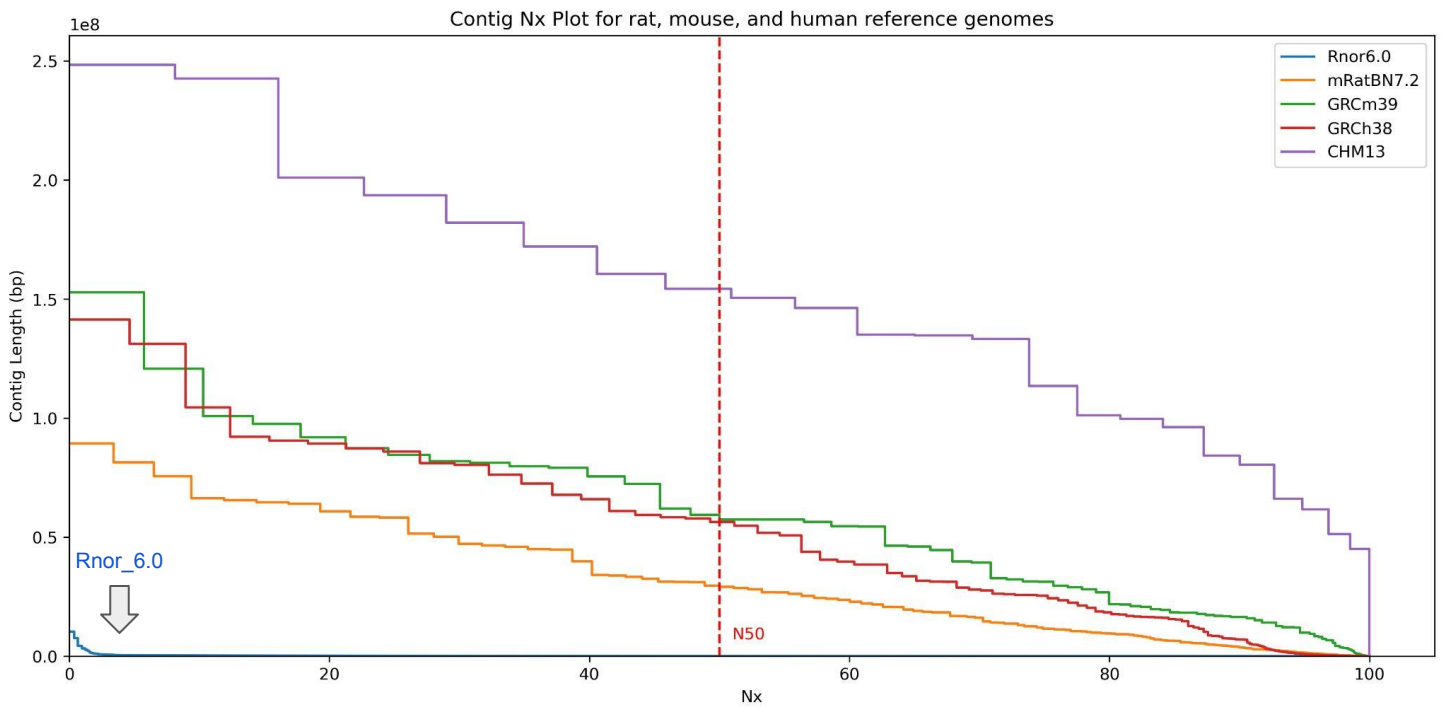


Figure S2. The Contig Nx plot for the Rat, Mouse, and Human reference genomes. The red dashed line indicates the contig N50 values. The Nx curve for Rnor6.0 is very low and partially overlapping with the X axis. The CHM13, being the first truly gapless human genome, is in its own tier in terms of assembly continuity. The GRCh38 released in 2014 and the GRCm39 released in 2020 have similar continuity. Although mRatBN7.2 is still lagging behind in continuity compared to the current human and mouse reference genomes, it represent a very significant improvement over the current rat reference genome Rnor_6.0 (Orange vs. Cyan line).

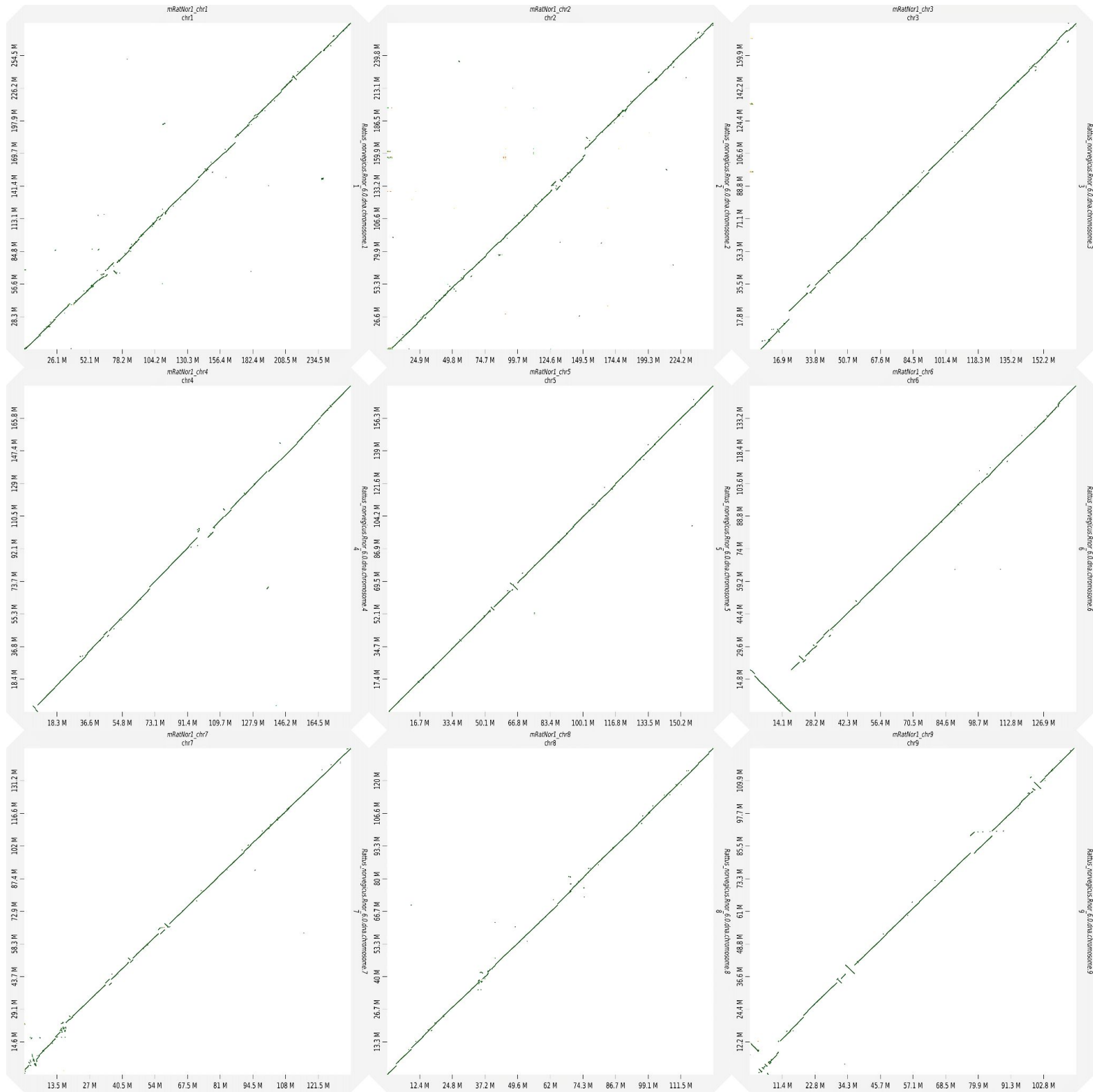


Figure S3. Chromosomal dot plots between Rnor_6.0 and mRatBn7.2. mRatBn7.2 is on the y-axis and Rnor_6.0 is on the x-axis.

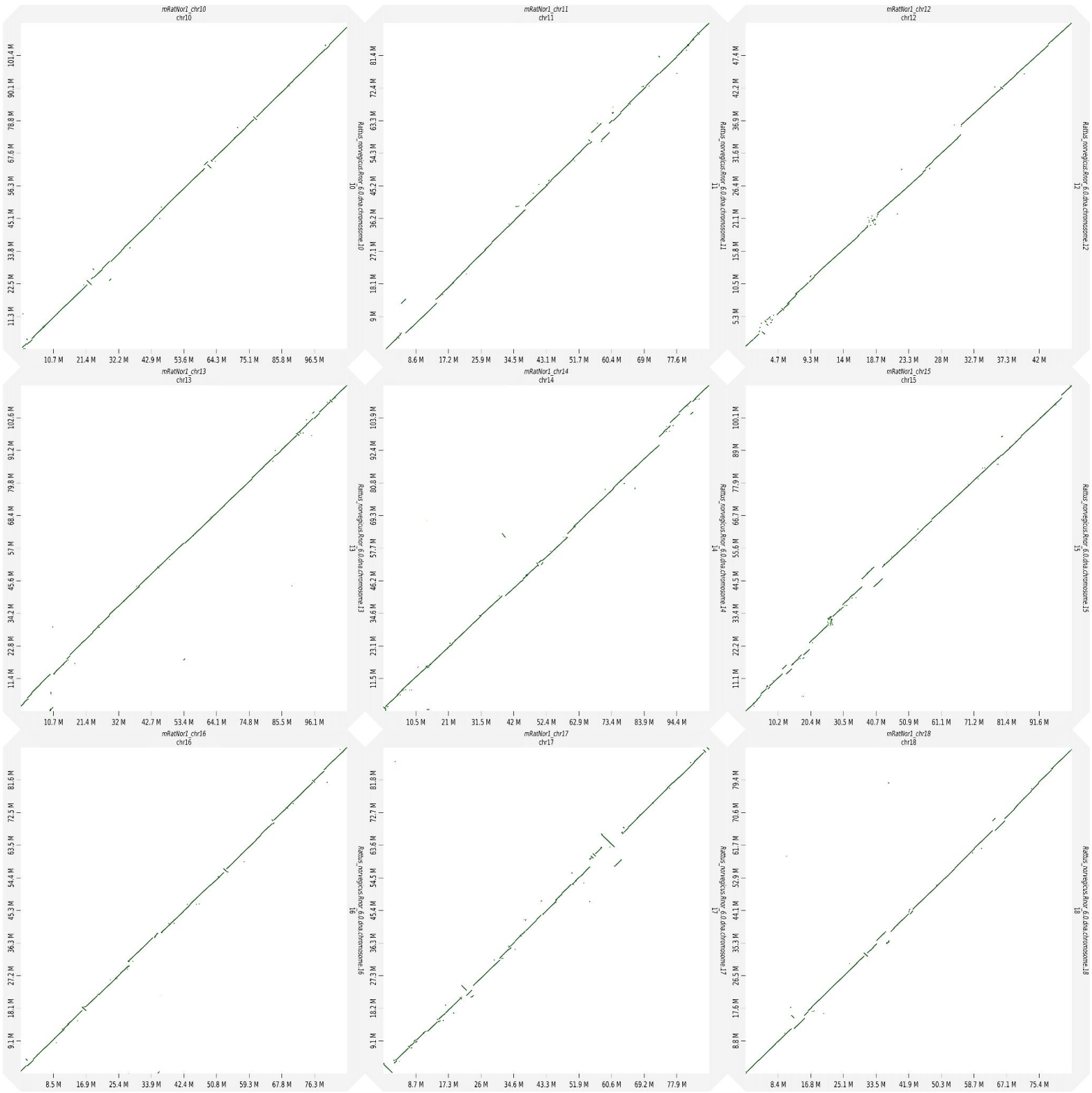


Figure S3 Continued.

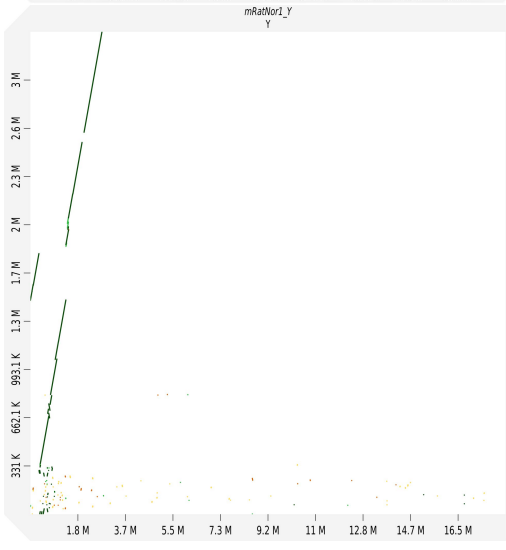
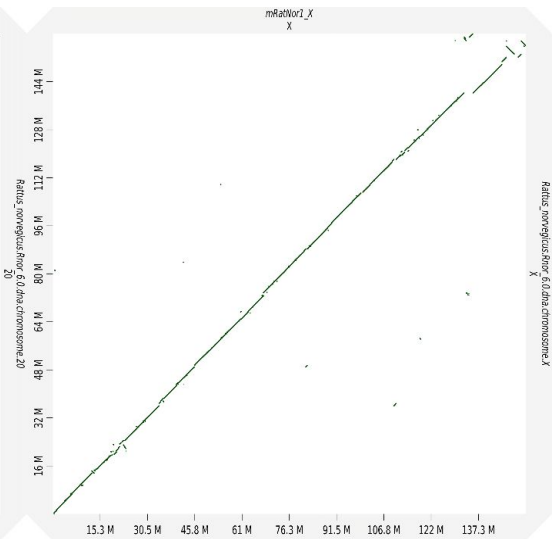
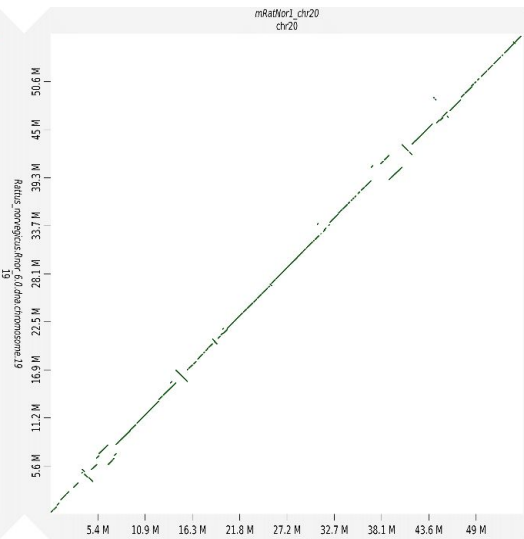
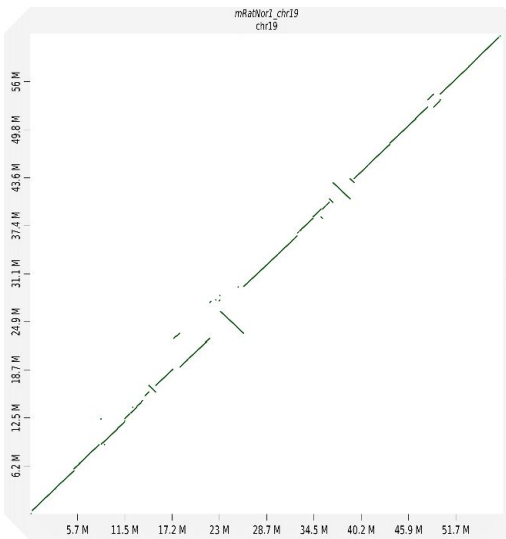
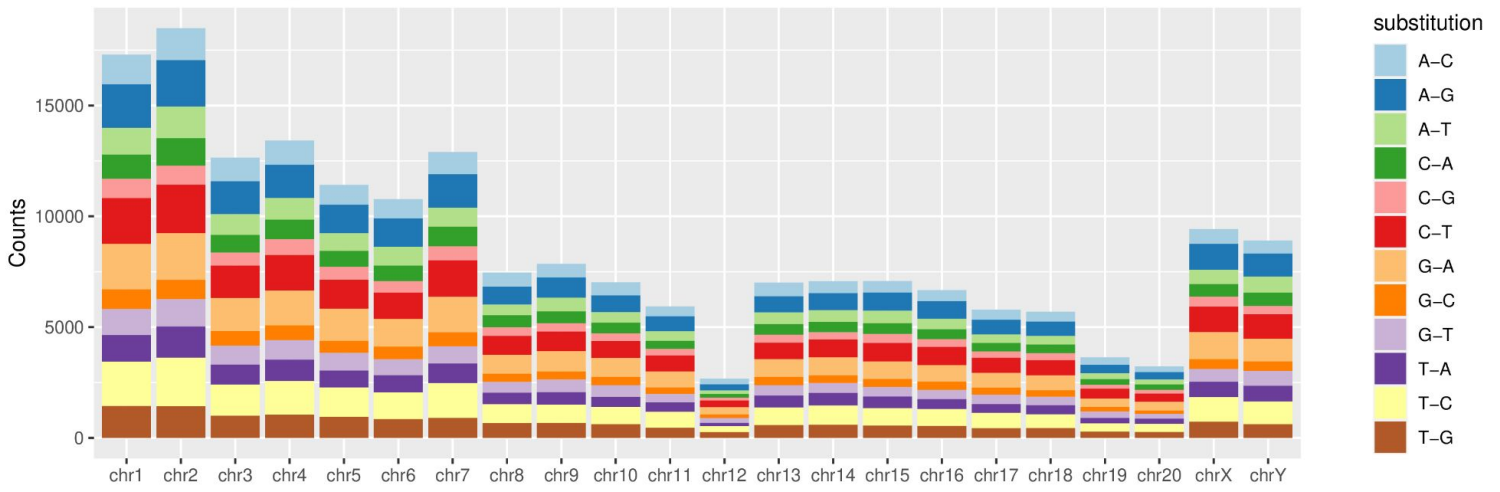
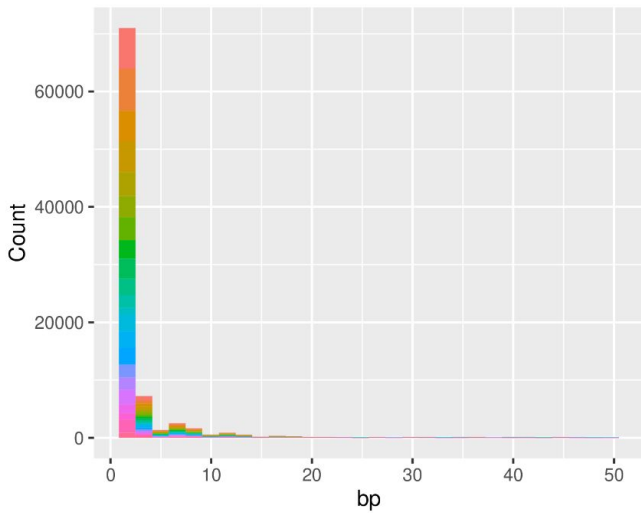


Figure S3 Continued

A. Number of single nucleotide variants between Rnor6.0 and mRatBN7.2



B. Size of indels between Rnor6.0 and mRatBN7.2.



C. Size of SVs between Rnor6.0 and mRatBN7.2.

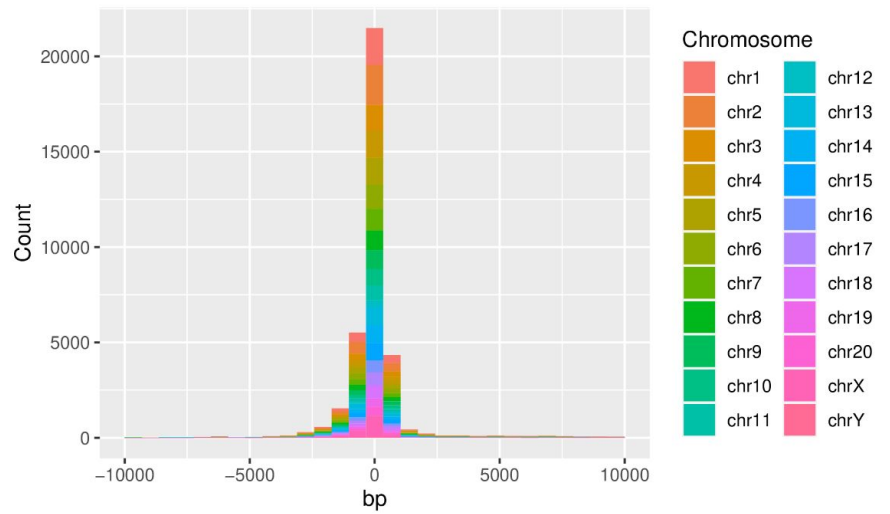


Figure S4: SNP and structural variants between Rnor_6.0 and mRatBN7.2. A) The substitution frequency per chromosome for Rnor6.0 and mRatBN7.2. **B)** The size of indels between Rnor 6.0 and mRatBN7.2. **C)** The size of SVs between Rnor6.0 and mRatBN7.2.

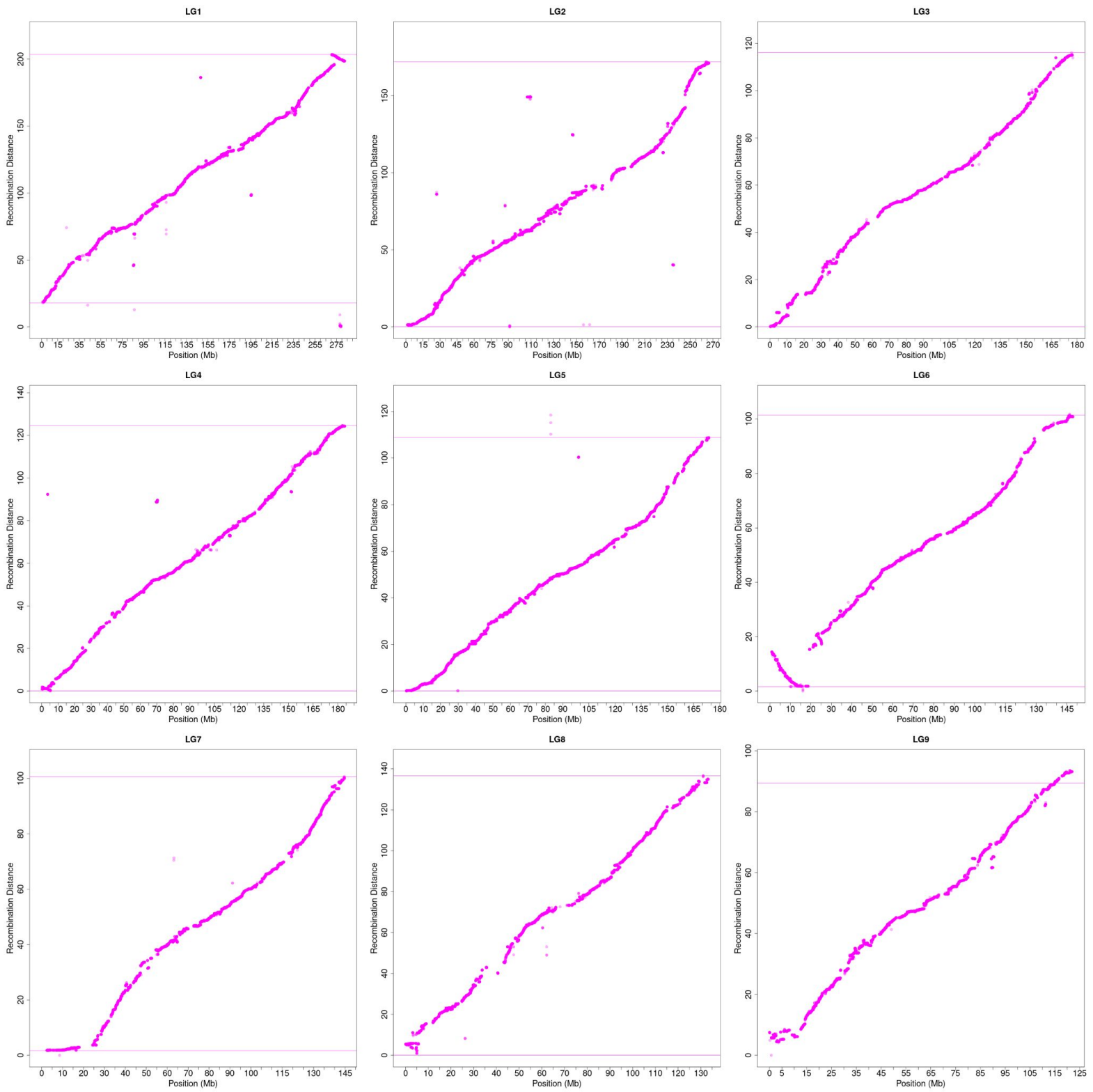


Figure S5. The order of genetic markers and their distances from a rat genetic map compared to their locations in Rnor_6.0.

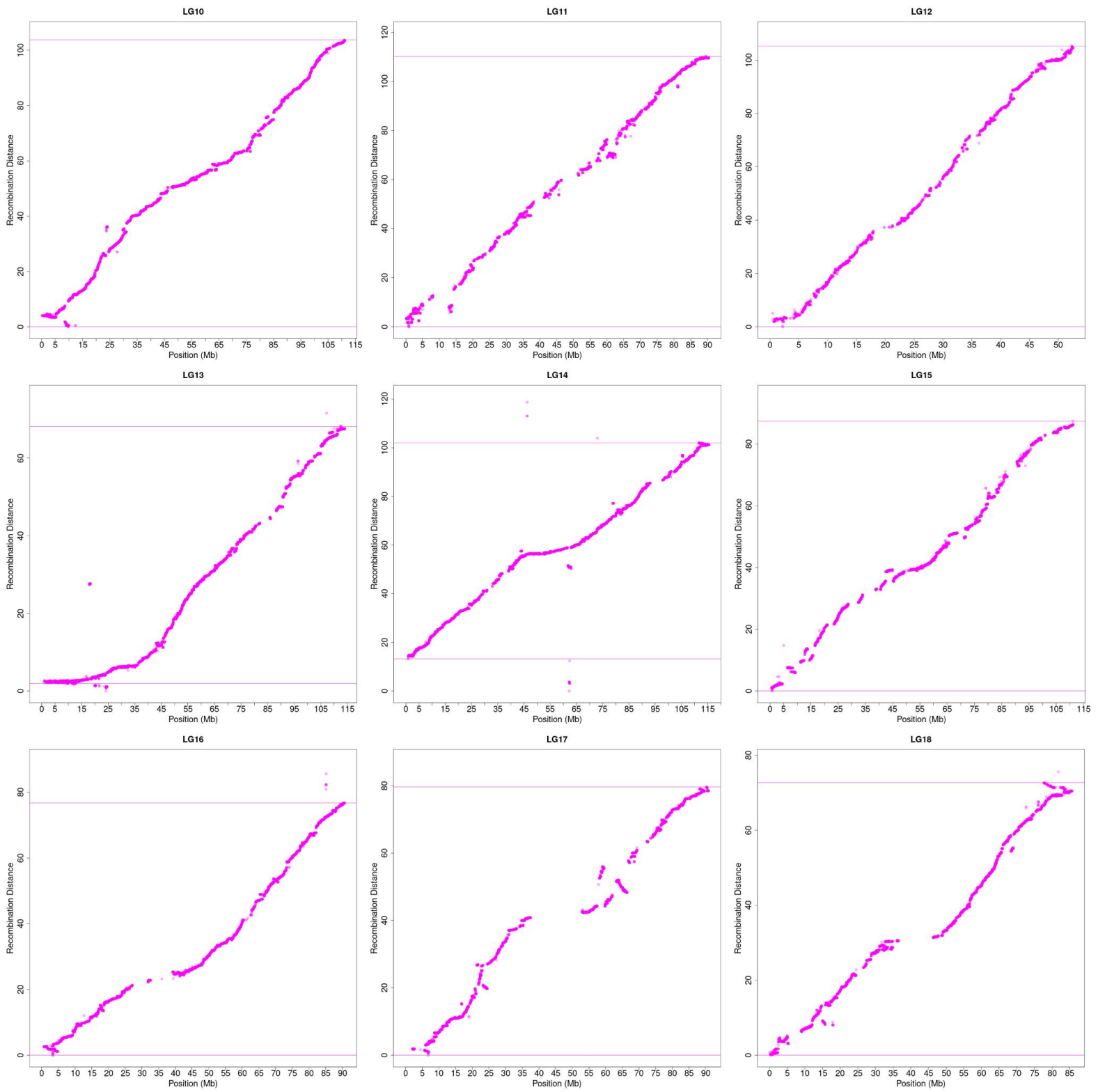
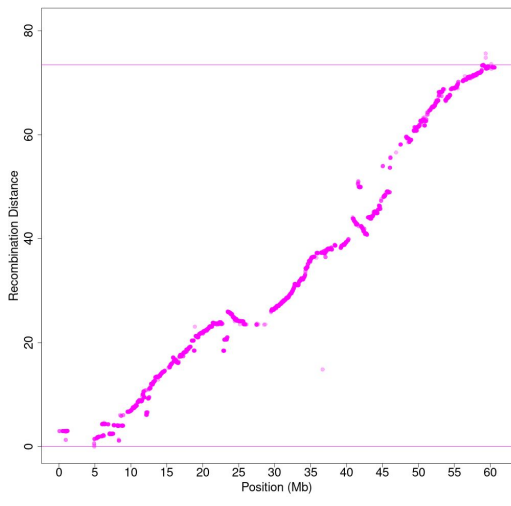


Figure S5 con.

LG19



LG20

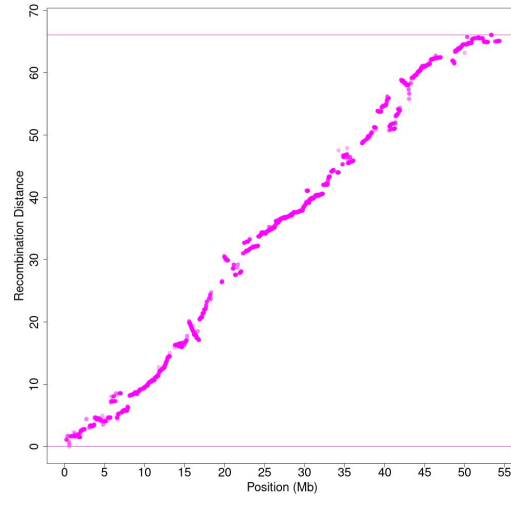


Figure S5 con.

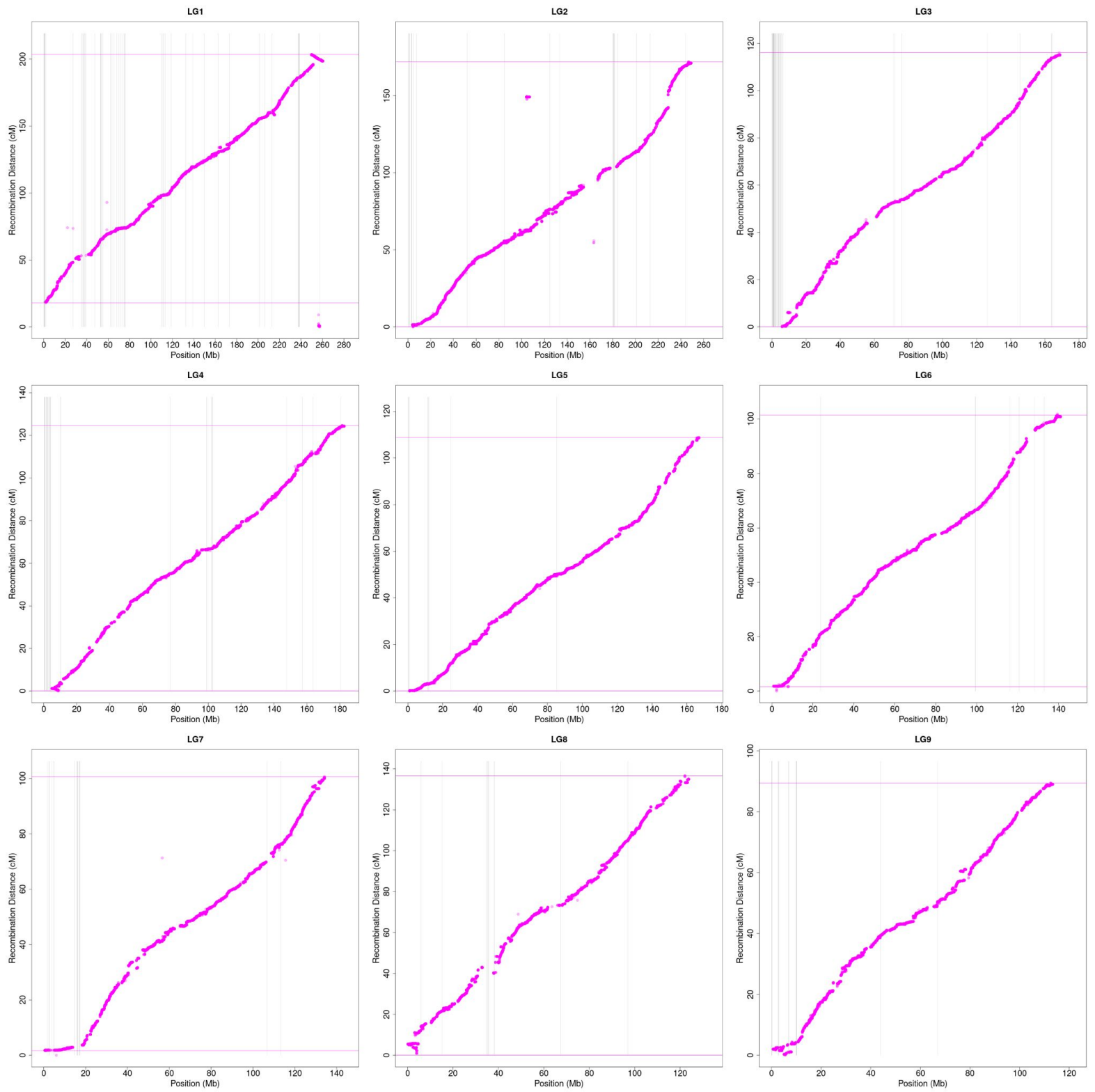


Figure S6. The order of genetic markers and their distances from a rat genetic map compared to their locations in mRatBN7.2

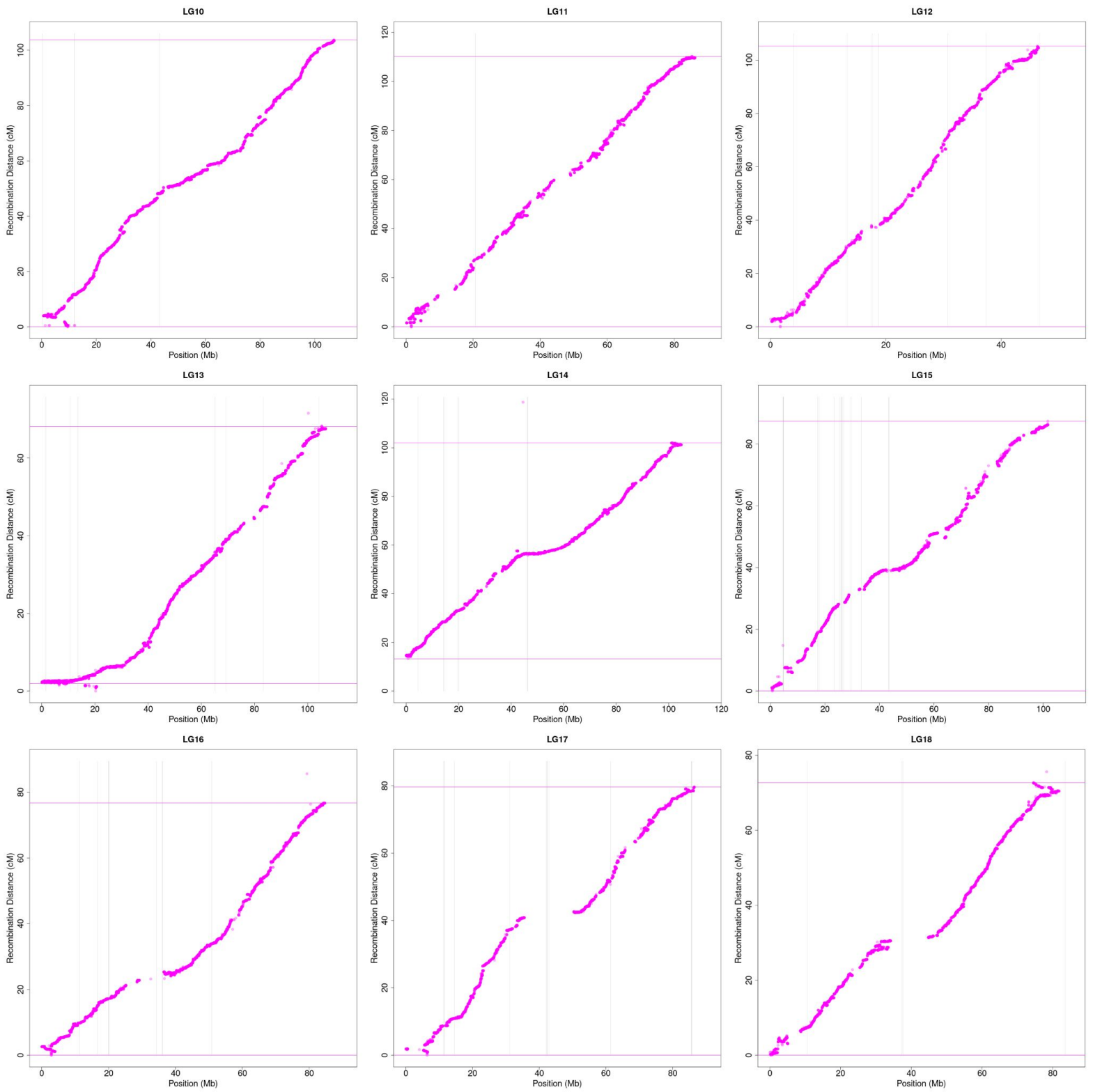
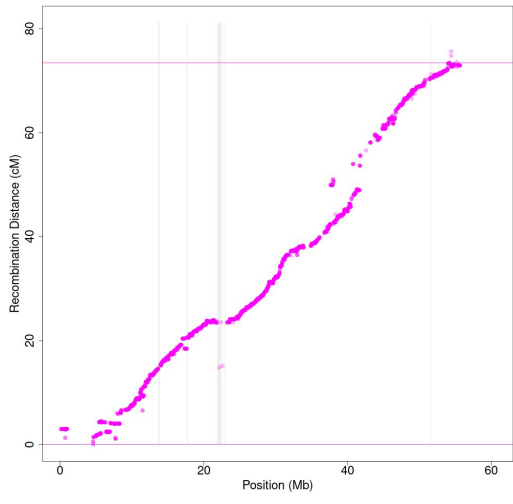


Figure S6 Continued

LG19



LG20

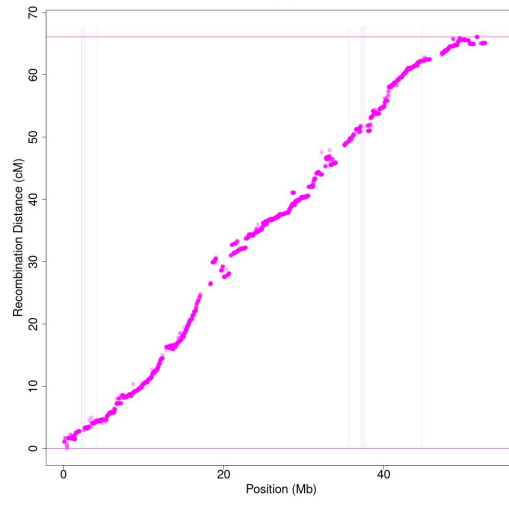


Figure S6 continued

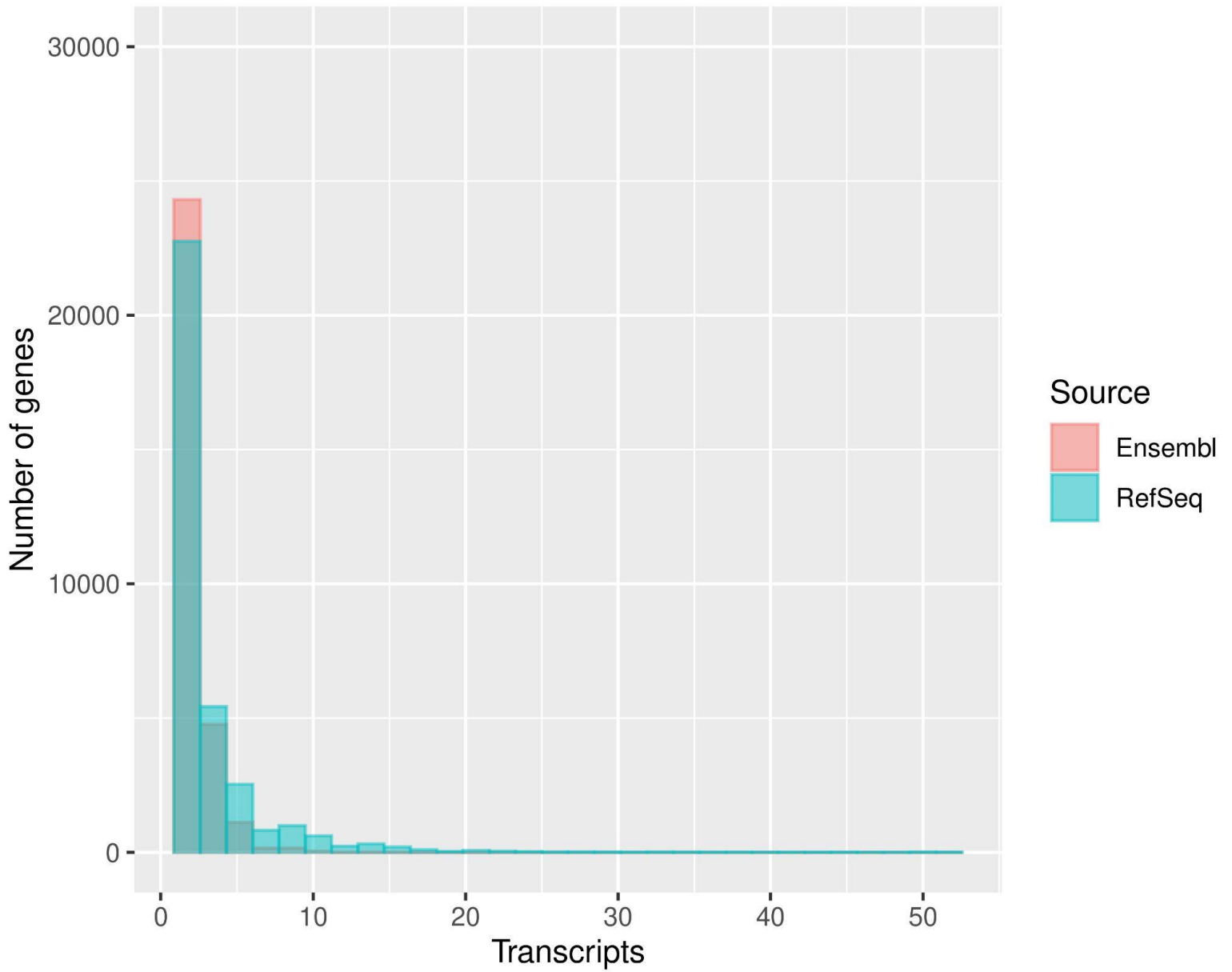


Figure S7. The number of transcripts annotated by Ensembl and RefSeq. Refseq annotated a greater number of genes with multiple transcripts than Ensembl. The average number of transcripts per gene was 2.9 for RefSeq and 1.8 for Ensembl.

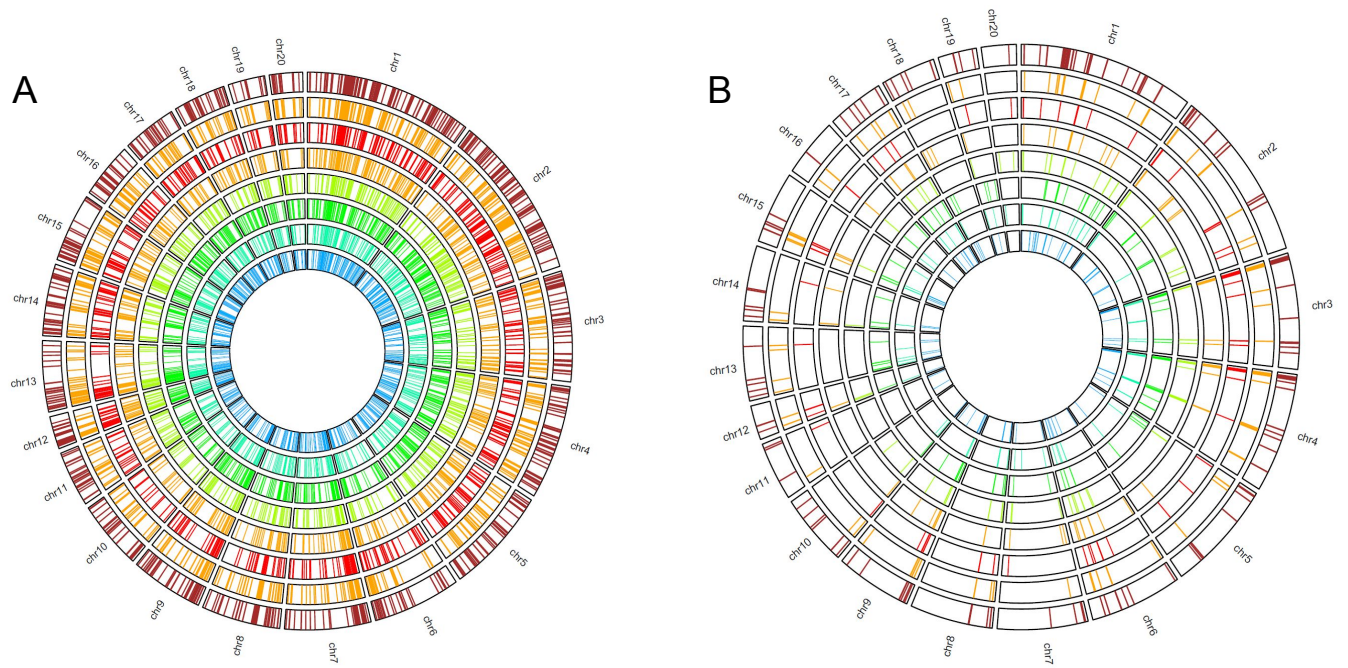


Figure S8. Locations of large deletions from multiple linked-read samples mapped the autosomes. From outer circle to inner circle are the following strains: BN, SHR/Olalpcv, BXH10, BXH8, BXH2, HXB17, HXB2, HXB21. **A)** Large deletions mapped to Rnor6.0 **B)** Large deletions mapped to mRatBN7.2.

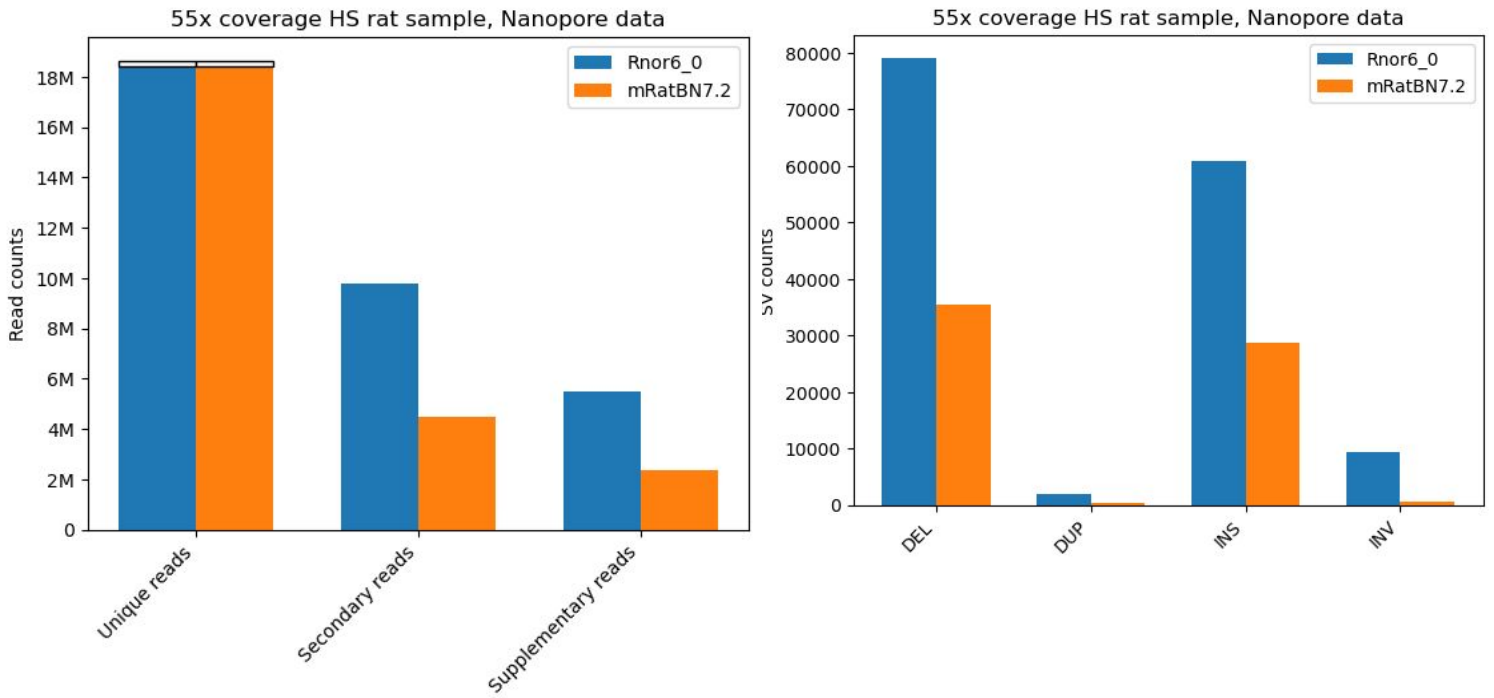


Figure S9. Mapping results of nanopore data of one HS rat with 55x coverage.

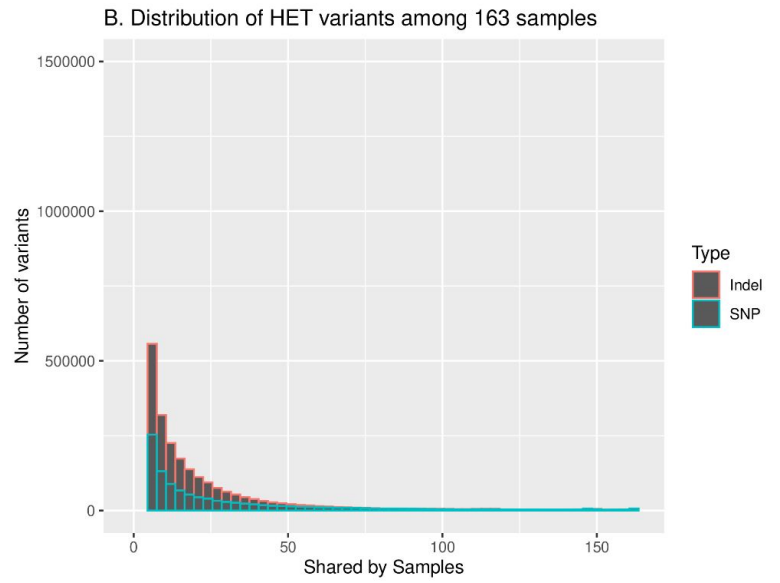
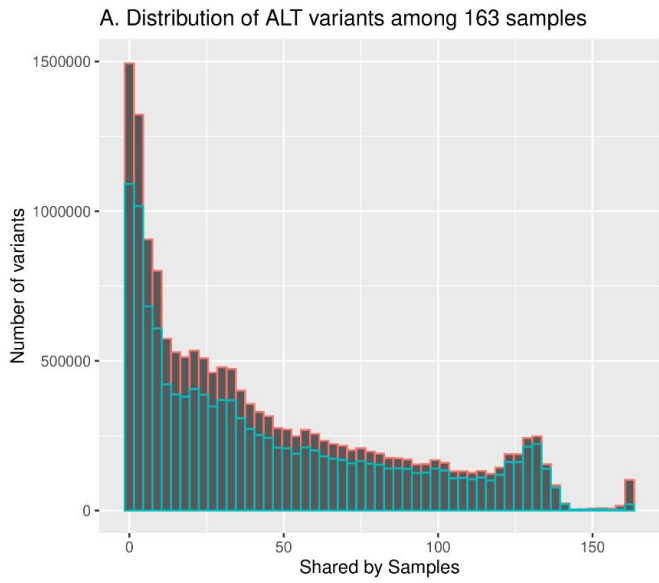


Figure S10. Distribution of the total number of variants shared by 1 or more samples across all 163 samples for A) homozygous SNPs/Indels and B) Heterozygous SNPs/Indels.

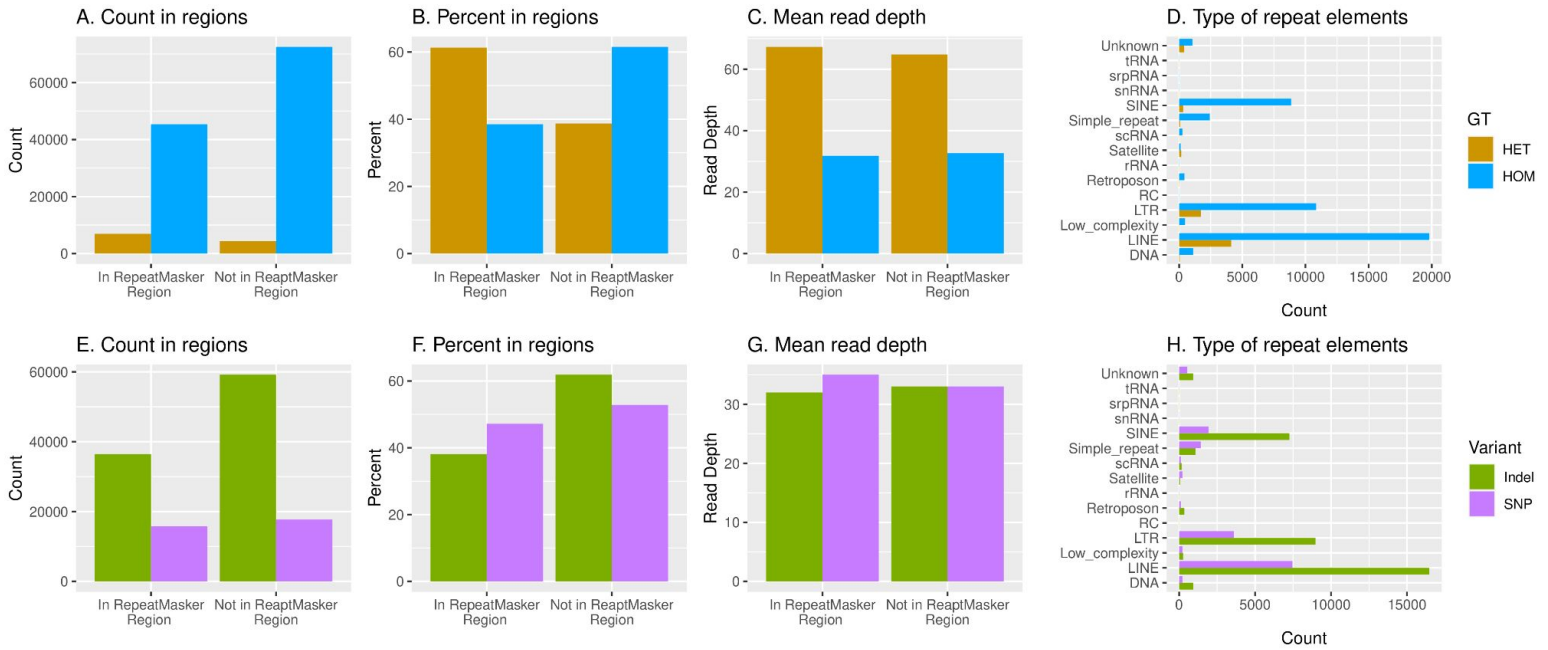


Figure S11. Potential errors on mRatBN7.2 are concentrated in repeat regions. We identified 129,186 variants that are shared by more than 156 out of 163 samples with the same genotype and zygosity. They are indicative of errors in mRatBN7.2. A majority of these variants are located in repeat regions. **A)** Count of these variants by zygosity and regions identified by Repeatmasker. **B)** Percent of these variants by zygosity and by regions identified by Repeatmasker. **C)** Read depth of variants by zygosity and by regions identified by Repeatmasker. **D)** Types of repetitive elements overlapping with these variants. **E-H).** Similar to A-D but classified by variant type (i.e., SNP vs Indel)

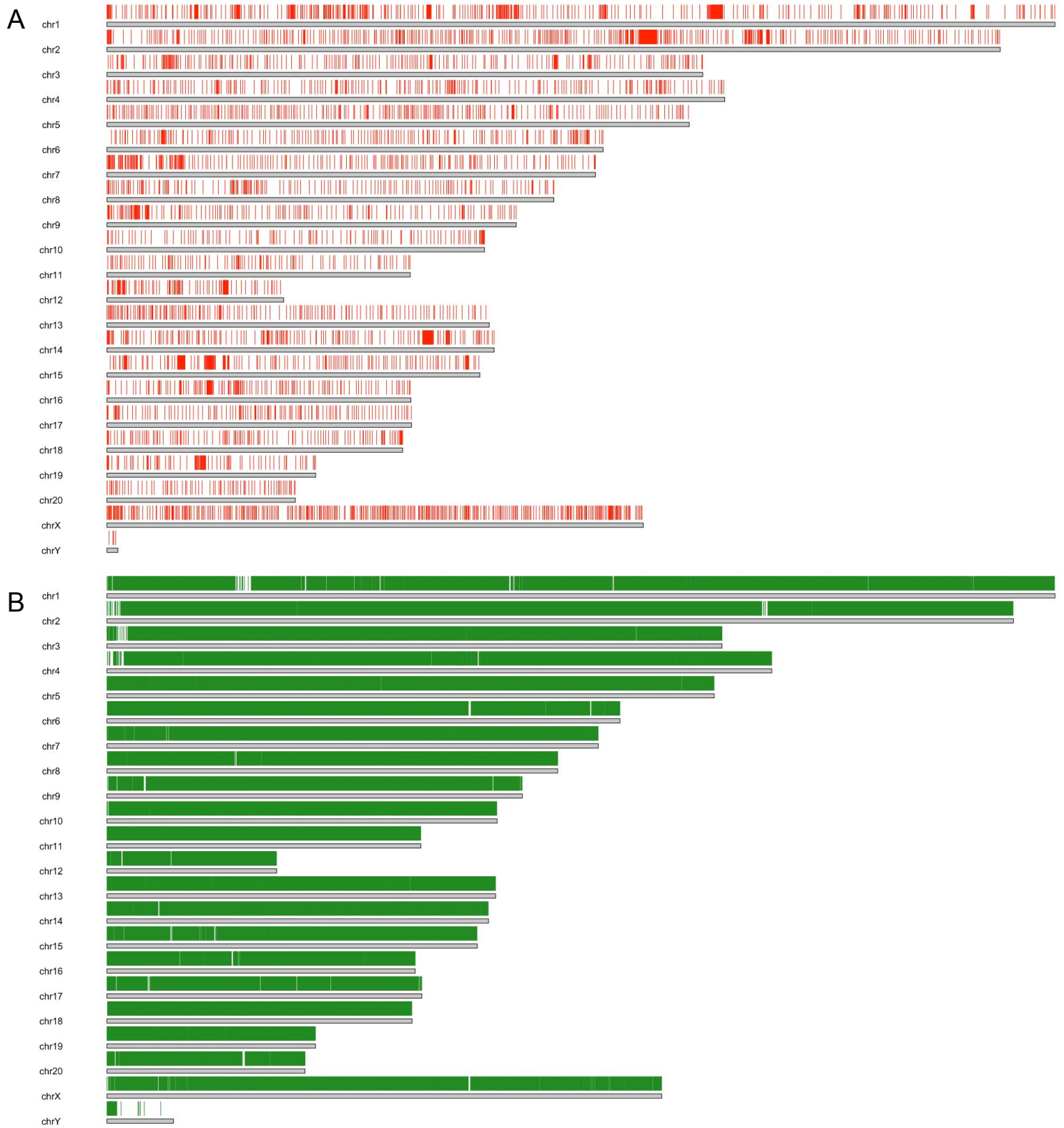


Figure S12. Simulated liftover analysis: unliftable and lifted site distribution. An evenly 1000 bp spaced bed file covering Rnor_6.0 is generated and then lifted to mRatBN7.2. Out of the 2,782,023 sites, 92.04% are liftable, 7.96% are not liftable. **A)** The distribution of the unliftable sites are plotted on Rnor_6.0. **B)** The distribution of the lifted sites are plotted on mRatBN7.2. The data are subsampled by 50 for better visualization.

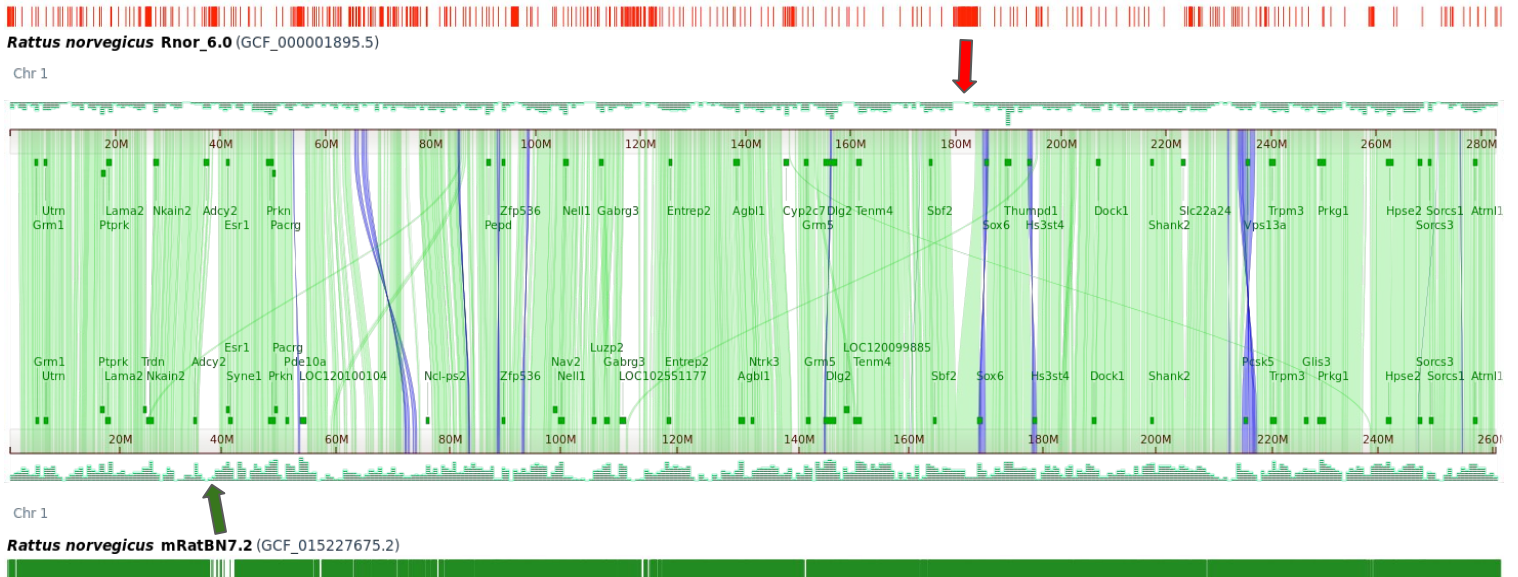
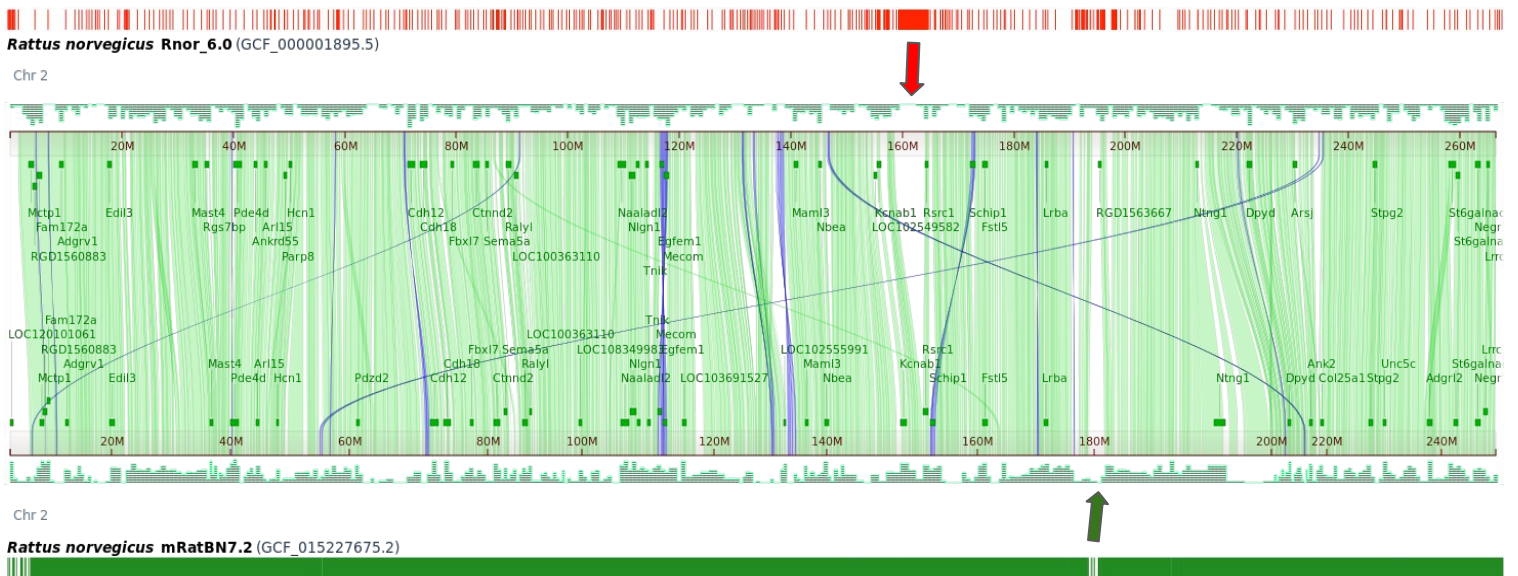
A**B**

Figure S13. Simulated liftover analysis and comparative genome view. Top red track is the unliftable sites distributed on the Rnor_6.0. Bottom green track is the lifted sites distributed on the mRatBN7.2. The middle is the NCBI comparative genome viewer between chr1 of Rnor_6.0 and mRatBN7.2. The red arrow highlights a region that's not liftable in Rnor_6.0. According to the comparative genome view, the region has no corresponding region in mRatBN7.2 (not just chr1). The green arrow highlights a region in mRatBN7.2 that has no lifted sites from Rnor_6.0. According to the comparative genome view, this region has matching region in Rnor_6.0. **A) Chromosome 1. B) Chromosome 2.**

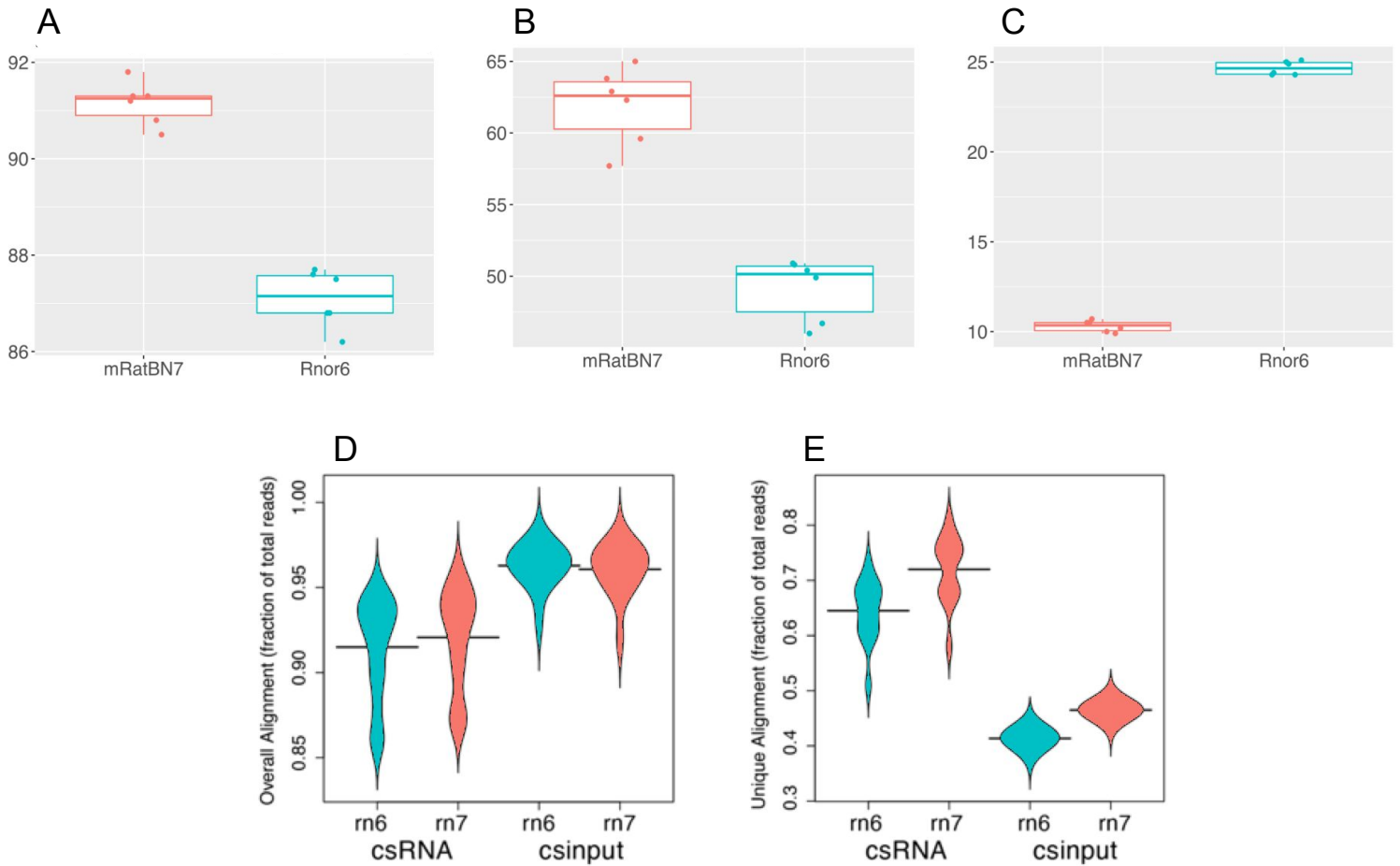


Figure S14. Mapping metrics of single nuclei RNA-seq and small capped RNA-seq data generated using rat brain tissues. Percentage of snRNA-seq reads mapped to **A)** the genome, **B)** the transcriptome, and **C)** an intergenic region of the genome for mRatBN7.2 and Rnor_6.0. **D)** Overall mapping rates of csRNA-seq data to two reference genomes. **E)** Unique alignment of scRNA-seq data to two reference genomes.

Vairant density from 163 rats

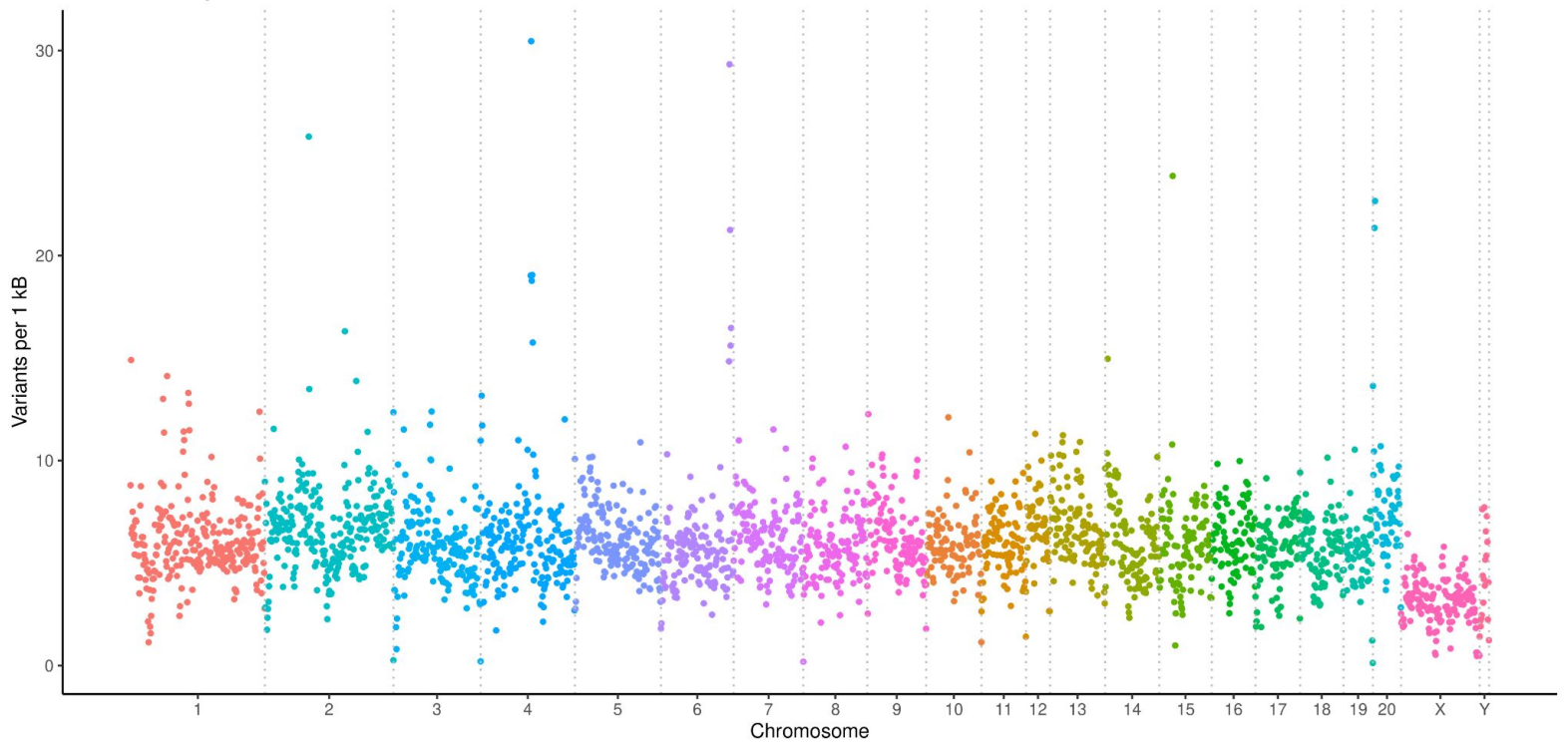


Figure S15. Density of variants across the genome in a collection of 163 rat WGS samples

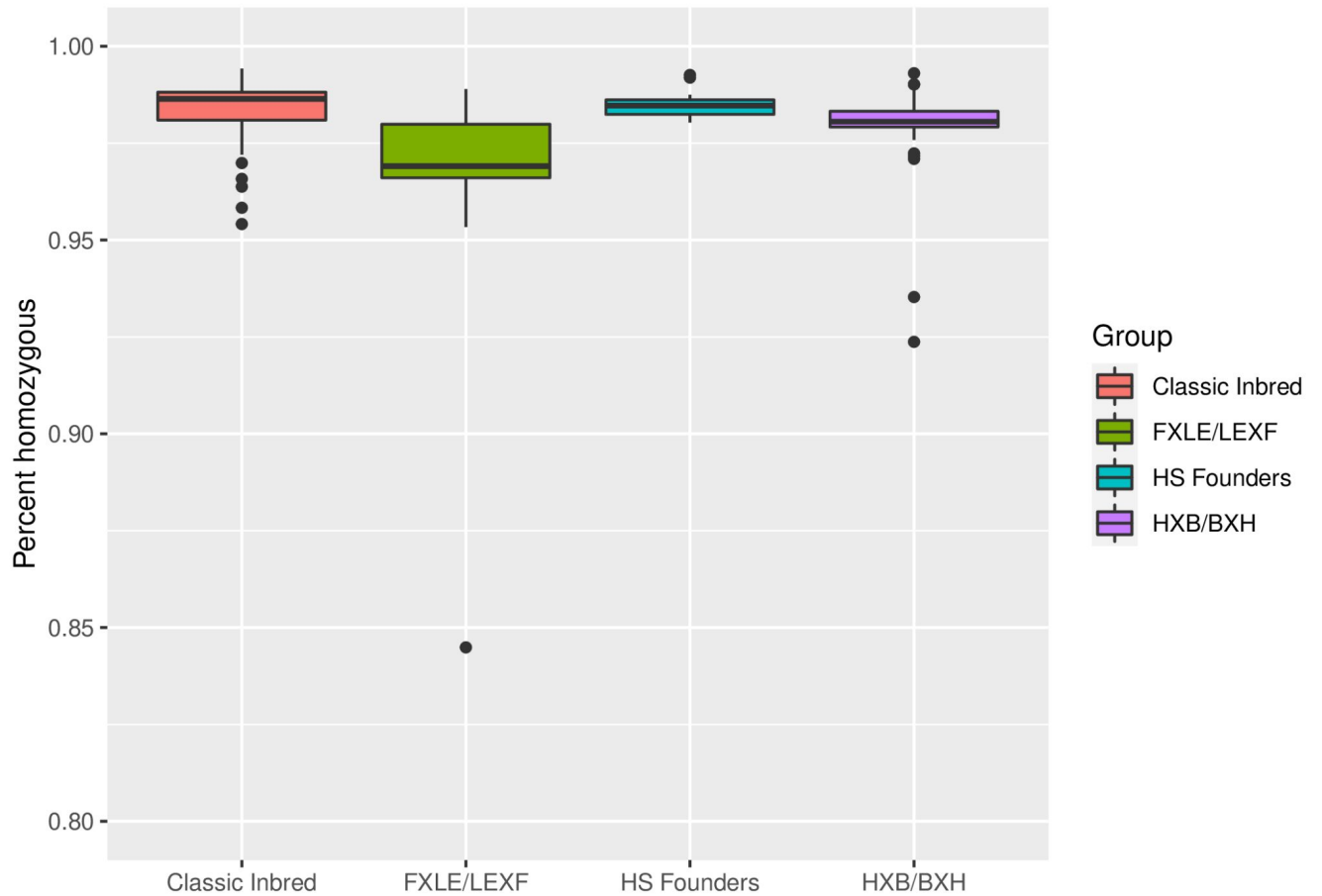


Figure S16. Homozygosity of inbred strains. Variants in most sample were homozygous, confirming the inbred nature of most strains. A few exceptions were noted. For example, 15.6% of the variants from FXLE24 were heterozygous. Additionally, BXH2, which we sequenced two samples, has ~7.7% heterozygous variants.

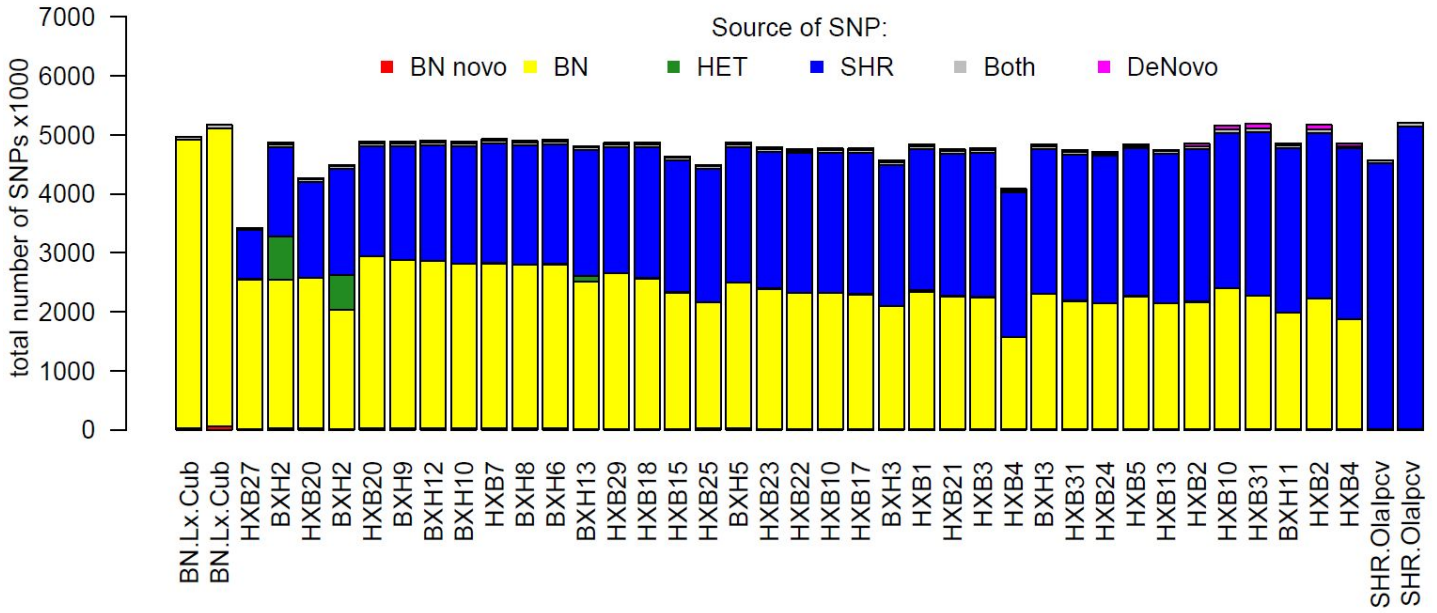


Figure S17. The majority of the variants in HXB/BXH RI panel are inherited from the parental strains. De Novo Mutations on BN.Lx (Red) Mutations originating from BN (Yellow), Heterozygous (Green), Mutations originating from SHR (Blue), SNPs found on both BN and SHR (Grey), De Novo SNPs from neither parental strains (Pink).

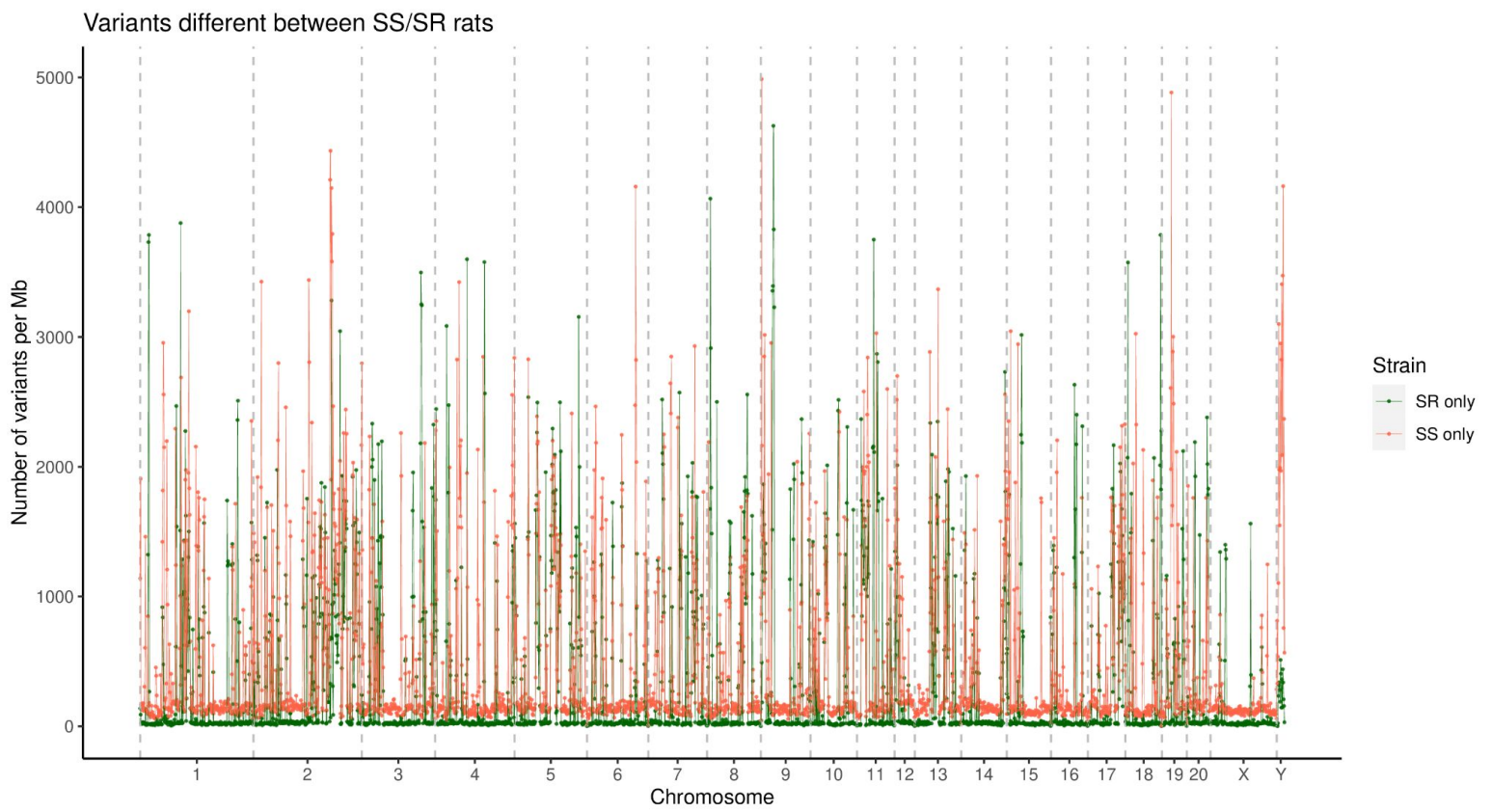


Figure S18. Distribution of variants different between SS and SR rats. Total number of variants unique to either SS or SR rats across the genome. The SS strain contains more unique variants than SR.

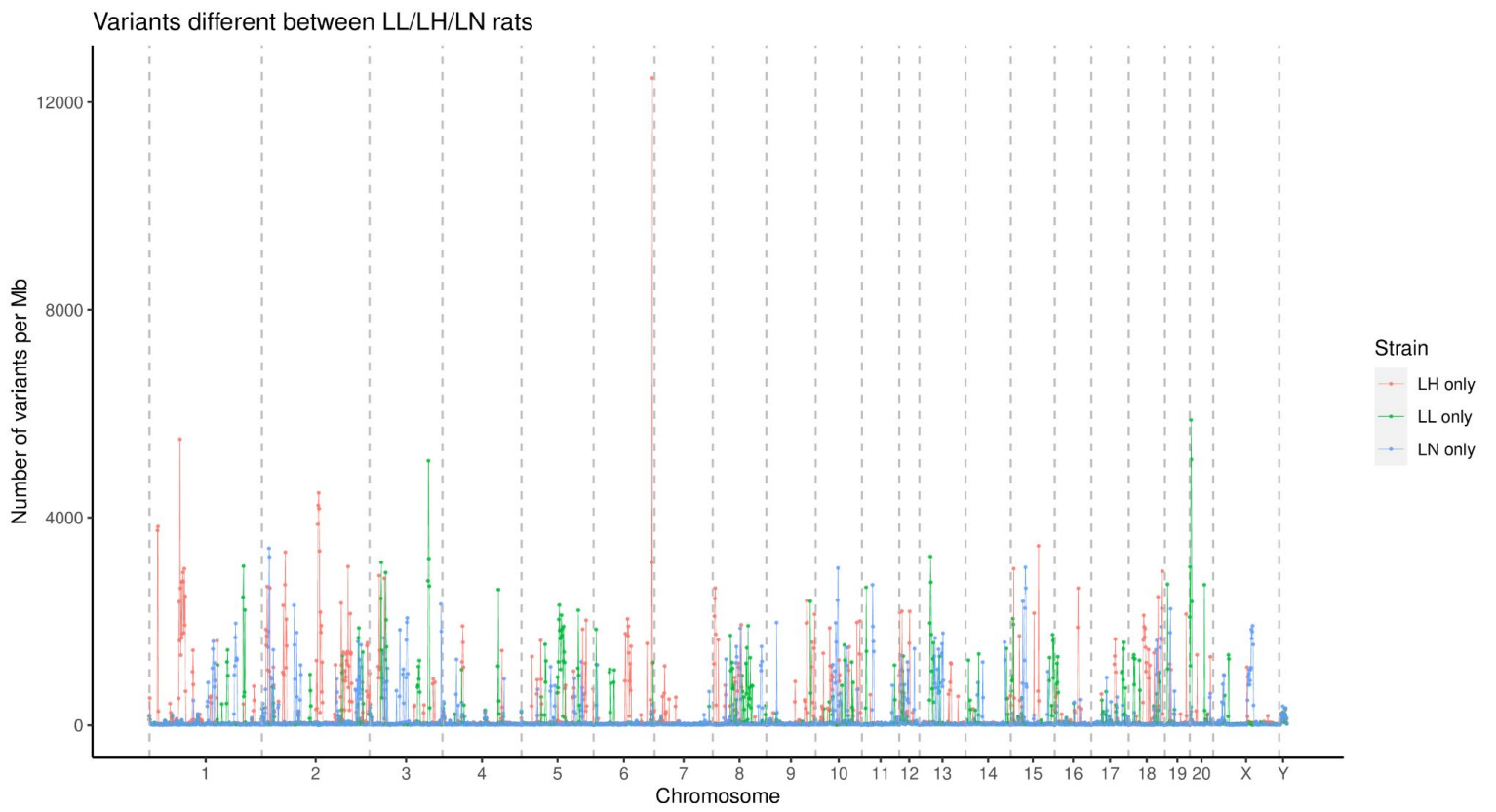


Figure S19 Distribution of variants unique to LL/LH/LN rats. Strain-specific variants tend to cluster by strain.

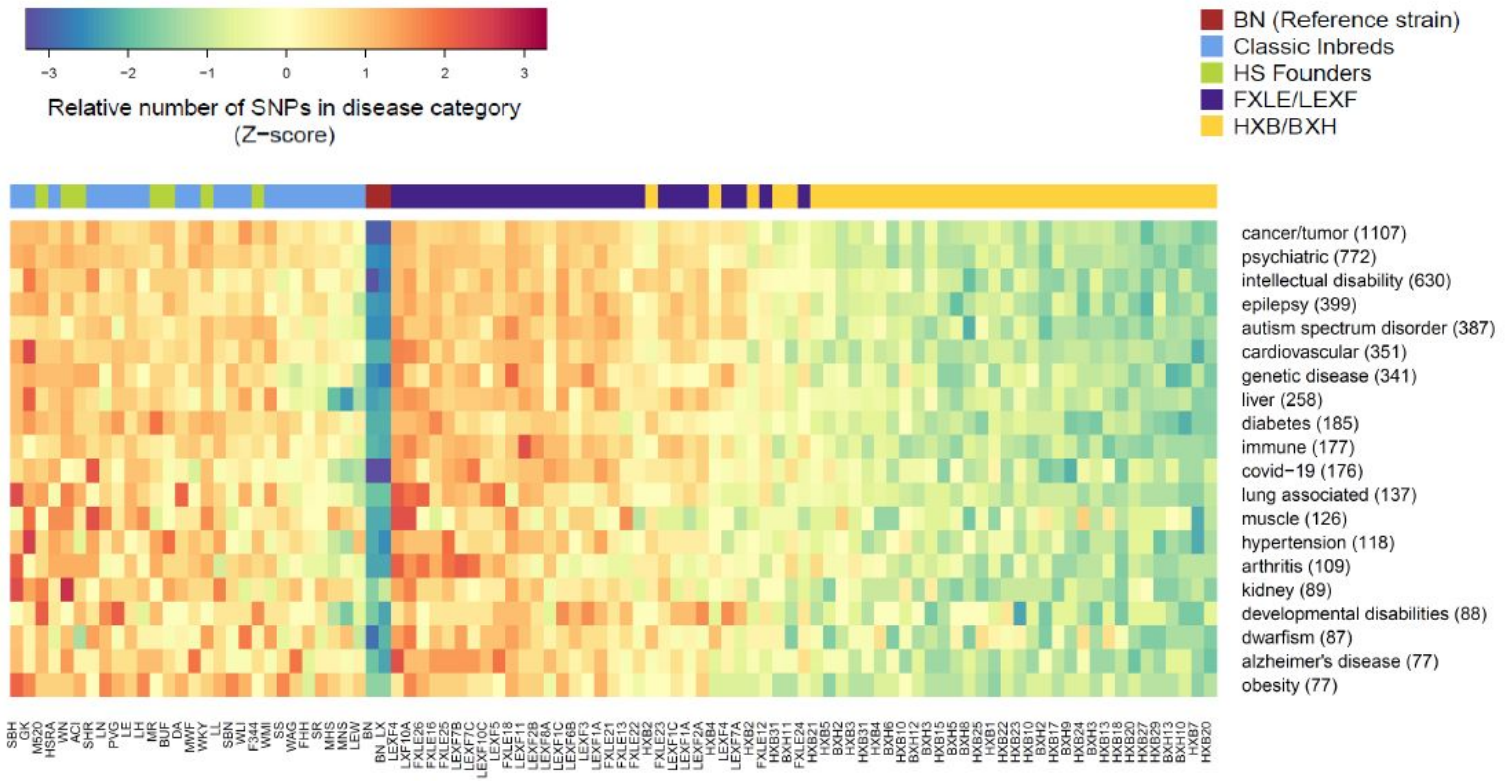
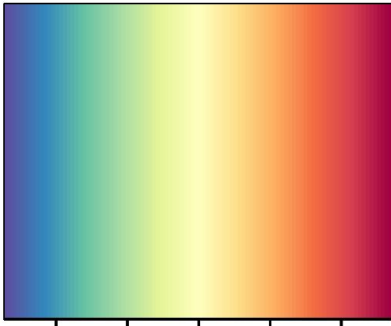


Figure S20. Overview of disease ontology annotation of variants in RatCollection. Heat maps shows relative number of SNPs in proximity of genes associated per disease category within the reference, classic inbreds, HS progenitors and recombinant inbred panels.

Color Key



-2 -1 0 1 2

Relative number of SNPs in disease category
(Z-score)

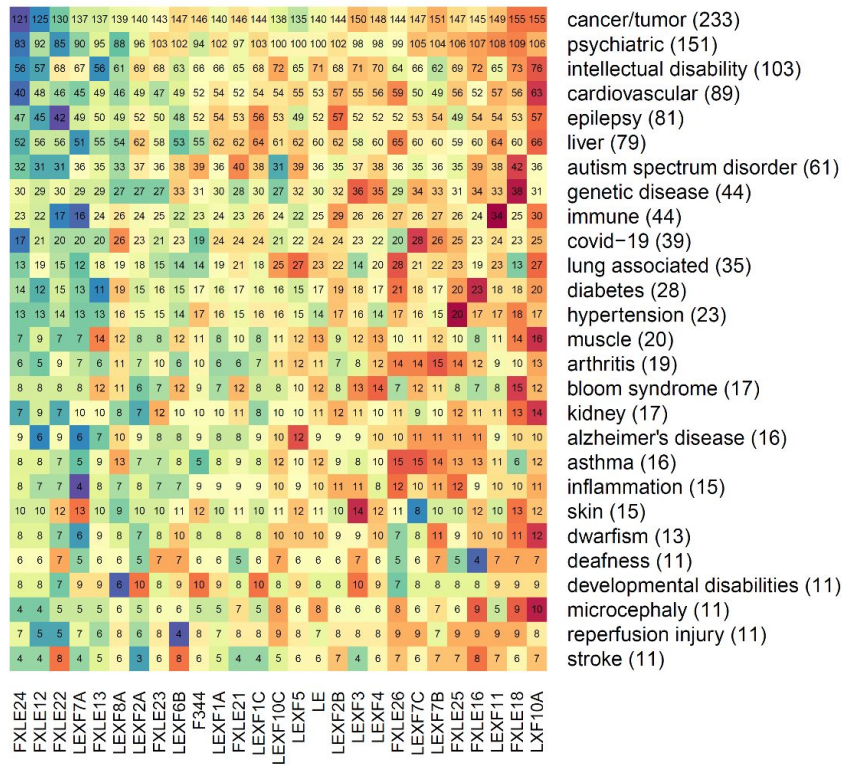


Figure S22. Disease ontology annotation of variants in the LEXF/FXLE RI panel. Numbers within the grid show the absolute number of SNPs with high impact on genes within the disease annotation. Numbers following the disease name shows the total number of annotated genes with at least 1 high impact variant within the panel.