



































Y and mitochondrial chromosomes in the heterogeneous stock rat population

Faith Okamoto ¹, Apurva S. Chitre ², Thiago Missfeldt Sanches ¹, Denghui Chen ², Daniel Munro ¹, Allegra T. Aron ^{3,4}, Angela Beeson ⁵, Hannah V. Bimschleger ¹, Maya Eid,⁶ Angel G. Garcia Martinez,⁷ Wenyan Han ⁷, Katie Holl,⁸ Tyler Jackson,⁹ Benjamin B. Johnson ¹, Christopher P. King ¹⁰, Brittany N. Kuhn ¹¹, Alexander C. Lamparelli ¹², Alesa H. Netzley ¹³, Khai-Minh H. Nguyen ¹, Beverly F. Peng ¹, Jordan A. Tripi,¹⁰ Tengfei Wang ⁷, Kendra S. Ziegler ¹, Douglas J. Adams ¹⁴, Amelie Baud ^{15,16}, Lieselot L.G. Carrette ¹, Hao Chen ⁷, Giordano de Guglielmo ¹, Pieter Dorrestein ^{3,17,18,19}, Olivier George ¹, Keita Ishiwari ^{20,21}, Monica M. Jablonski ²², Thomas C. Jhou ⁹, Marsida Kallupi ¹, Rob Knight ^{23,19,24,25}, Paul J. Meyer ¹⁰, Leah C. Solberg Woods ⁵, Oksana Poleskaya ¹, Abraham A. Palmer ^{1,26,*}

¹Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA

²Bioinformatics and System Biology Program, University of California San Diego, La Jolla, CA 92093, USA

³Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

⁴Department of Chemistry and Biochemistry, University of Denver, Denver, CO 80208, USA

⁵Department of Internal Medicine, Molecular Medicine, Wake Forest University School of Medicine, Winston-Salem, NC 27157, USA

⁶Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁷Department of Pharmacology, Addiction Science and Toxicology, University of Tennessee Health Science Center, Memphis, TN 38163, USA

⁸Department of Physiology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

⁹Department of Neurobiology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

¹⁰Department of Psychology, University at Buffalo, Buffalo, NY 14260, USA

¹¹Department of Neuroscience, Medical University of South Carolina, Charleston, SC 29425, USA

¹²Department of Psychology, University of California Los Angeles, Los Angeles, CA 90095, USA

¹³Department of Emergency Medicine, University of Michigan, Ann Arbor, MI 48109, USA

¹⁴Department of Orthopedics, University of Colorado - Anschutz Medical Campus, Aurora, CO 80045, USA

¹⁵Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain

¹⁶Universitat Pompeu Fabra, Barcelona, Spain

¹⁷Collaborative Mass Spectrometry Innovation Center, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

¹⁸Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, USA

¹⁹Center for Microbiome Innovation, University of California San Diego, La Jolla, CA 92093, USA

²⁰Department of Pharmacology and Toxicology, University at Buffalo, Buffalo, NY 14203, USA

²¹Clinical and Research Institute on Addictions, University at Buffalo, Buffalo, NY 14203, USA

²²Department of Ophthalmology and Department of Anatomy and Neurobiology, University of Tennessee Health Science Center, Memphis, TN 38163, USA

²³Department of Pediatrics, University of California San Diego, La Jolla, CA 92093, USA

²⁴Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA

²⁵Department of Bioengineering, University of California San Diego, La Jolla, CA 92093, USA

²⁶Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA

*Corresponding author: Department of Psychiatry, University of California San Diego, La Jolla, CA 92093, USA; Institute for Genomic Medicine, University of California San Diego, La Jolla, CA 92093, USA. Email: aap@ucsd.edu

Genome-wide association studies typically evaluate the autosomes and sometimes the X Chromosome, but seldom consider the Y or mitochondrial (MT) Chromosomes. We genotyped the Y and MT Chromosomes in heterogeneous stock (HS) rats (*Rattus norvegicus*), an outbred population created from 8 inbred strains. We identified 8 distinct Y and 4 distinct MT Chromosomes among the 8 founders. However, only 2 types of each nonrecombinant chromosome were observed in our modern HS rat population (generations 81–97). Despite the relatively large sample size, there were virtually no significant associations for behavioral, physiological, metabolome, or microbiome traits after correcting for multiple comparisons. However, both Y and MT Chromosomes were strongly associated with the expression of a few genes located on those chromosomes, which provided a positive control. Our results suggest that within modern HS rats there are no Y and MT Chromosomes differences that strongly influence behavioral or physiological traits. These results do not address other ancestral Y and MT Chromosomes that do not appear in modern HS rats, nor do they address effects that may exist in other rat populations, or in other species.

Keywords: heterogeneous stock; rat; Y Chromosome; mitochondria; haplotype; low-coverage; PheWAS; RNA-seq

Introduction

Heterogeneous stock (HS) rats (*Rattus norvegicus*) are a well-established outbred population that have been used for genome-wide association studies (GWAS); yet, their Y and mitochondrial (MT) Chromosomes have been largely ignored. The Y Chromosome was poorly assembled in prior versions of the rat genome. This dramatically improved in a recent rat reference genome (mRatBN7.2). In contrast, the MT Chromosome was not updated in mRatBN7.2 (Tutaj et al. 2019; de Jong et al. 2024).

HS rats have been outbred for almost 100 generations. They were created in 1984 by intercrossing eight inbred strains: ACI/N, BN/SsN, BUF/N, F344/N, M520/N, MR/N, WKY/N, and WN/N (Hansen and Spuhler 1984). Modern HS rat genomes are mosaics of those 8 founder haplotypes (Solberg Woods and Mott 2017), which enables precise genetic mapping of complex traits (e.g. Johannesson et al. 2009; Baud et al. 2013; Chitre et al. 2020). However, as Y and MT are nonrecombinant, even in a modern HS rat, they are expected to be inherited in their entirety from a single founder; the Y Chromosome from the father and the MT from the mother.

Some Y and MT haplotyping methods cannot be used in HS rats. For example, we lack complete pedigrees that could have been used to trace expected Y or MT genotypes, as was done in collaborative cross (CC) mice (Broman 2022). We also lack curated lists of informative variants, as in human databases (e.g. Kloss-Brandstätter et al. 2011; Chen et al. 2021).

In humans, Y or MT haplogroups have been tested for association with many phenotypes (e.g. Jamain et al. 2002; Ma et al. 2014; Howe et al. 2017; Cai et al. 2021; Degenhardt et al. 2022). Replication has proved difficult; population structure confounds such work (Hagen et al. 2018). For example, schizophrenia was linked to MT in a Han Chinese (Wang et al. 2013) cohort, but not Spanish (Mosquera-Miguel et al. 2012) or Swedish (Gonçalves et al. 2018) cohorts.

Studies in CC mice found that Y or MT genotype was not associated with sex ratio (Haines et al. 2021), but was associated with the expression of genes located on the Y and MT Chromosomes (Keele et al. 2021). Mouse models designed for isolating genetic effects of Y (e.g. Martincová et al. 2019) and MT (e.g. Welch et al. 2023) found phenotypic associations, even suggesting transgenerational effects of paternal Y Chromosome genotype in daughters (Nelson et al. 2010). However, our review of the literature did not find comparable Y or MT analyses in outbred mice.

We identified variants that could be used to determine which founder had contributed the Y and MT to each individual HS rat. This approach is broadly similar to a prior study in DO mice (Chesler et al. 2016). We then tested for associations between Y and MT genotypes and a large collection of phenotypic data that have been collected over almost a decade of studies using HS rats (www.ratgenes.org). These data include behavioral, physiological, metabolome, microbiome and RNA-seq complex traits; in total, 12,055 rats were both haplotyped and phenotyped.

Materials and methods

A Reagent Table is in the Supplementary Files.

Genotype datasets

We used preexisting whole-genome sequencing (WGS) data from males representing each of the 8 founder strains (~40× coverage). Single-nucleotide polymorphisms (SNPs) and indels on the Y and MT Chromosomes were called using GATK, as previously described

(Chen 2022). We identified polymorphic sites in these data which distinguish the different founder Y and MT Chromosomes. We also used preexisting WGS data from 44 male and 44 female outbred HS rats (~33× coverage); SNPs and indels were similarly called using GATK. Short tandem repeats (STRs) on the Y Chromosome were called in all WGS samples with HipSTR (Willems et al. 2017) and filtered with DumpSTR (Mousavi et al. 2020).

We also used preexisting low-coverage (~0.25×) data from 15,120 outbred HS rats. These used double-digest genotyping-by-sequencing (ddGBS; Gileta et al. 2020) or low-coverage WGS (lcWGS; Chen et al. 2023) library preparation. Biallelic SNP genotypes on mRatBN7.2 were imputed by STITCH (Davies et al. 2016), as previously described (Chen et al. 2024). We did not use the variant filters previously described. Instead, we started with all variants produced by STITCH and then used custom filters to avoid excluding variants potentially useful to distinguish founder Y or MT (see “Genotype filters”). Because Y and MT are hemizygous, heterozygous calls are unexpected (Supplementary Fig. S1); when observed, those genotypes were treated as missing. All procedures prior to tissue collection were approved by the relevant Institutional Animals Care and Use Committees.

Genotype filters

Our custom filters removed (1) variants with low INFO score (for low-coverage data), (2) variants with minor allele frequency (MAF) of 0, (3) variants with a high missing rate (>25%), and (4) individual samples with a high missing rate (>50%). We applied all or only some of these filters, always in this order, depending on the analysis. In particular, when visualizing by SNP to determine haplotypes (e.g. in alignments) we skipped the MAF filter to visualize fixed variants, and when plotting statistics (e.g. heterozygosity) by SNP in low-coverage data, we skipped all but the INFO score filter. Supplementary Fig. S2 shows the thresholds used against distributions of these statistics (INFO score, MAF = 0, per-SNP missing rate, per-sample missing rate) for low-coverage samples. We used these filters to create haplotype groups for association analyses.

Unrooted trees

We applied all standard filters to high-coverage genotype data, then used a matrix of Hamming distance (scale of 0–1) pairwise ignoring missingness, i.e. removing variants missing in either sample. We created an unrooted neighbor-joining (NJ) tree (Talevich et al. 2012). These trees were used for understanding HS Y and MT phylogeny, but not for haplotype group-making.

Statistical analysis

We performed a phenome-wide association study (PheWAS) for Y or MT haplotype via mixed linear model-based association (MLMA) analysis (Yang et al. 2014) with GCTA (Yang et al. 2011); see “GWAS phenotype association”. We tested normalized (cpm in edgeR) RNA-seq transcript abundance against Y or MT haplotype via a two-sample Wilcoxon rank-sums (i.e. Mann–Whitney) test (`wilcox.test` in R); see “Gene expression association”. We tested normalized RNA-seq transcript abundance against X Chromosome SNPs via a simple linear model (lm in R); see “Dkc1 expression and X SNPs association”. We used the Benjamini & Hochberg (BH) false discovery rate (FDR) approach (`p.adjust` in R; Benjamini and Hochberg 1995). For a single, binary phenotype (one or two kidneys at birth), we tested for association with MT haplotype using a Fisher’s exact test (`fisher.test` in R).

GWAS phenotype association

We used a genetic relationship matrix (GRM) constructed using PLINK (Chang et al. 2015) to account for autosomal (--chr 1-20) relatedness (Yang et al. 2010), which we expected to be correlated with Y and MT haplotype due to familial structure. After filtration by missingness (--geno 0.1), violations of Hardy-Weinberg equilibrium (--hwe 1e-10; Wigginton et al. 2005), and MAF (--maf 0.005), 5,315,011 SNPs and 15,120 samples remained. We fit a linear model on all raw values and covariates, then inverse-normal transformed the residuals.

The traits used for the PheWAS are shown in Supplementary Table S1, with their sample sizes and minor haplotype frequencies in Supplementary Fig. S3. We encoded Y and MT haplotypes as SNPs: reference-like haplotype (from the same haplogroup as BN) as homozygous reference allele (0), and other haplotype as homozygous alternate allele (2). We ran GCTA's MLMA with these genotypes, the autosomal GRM, and processed phenotypes. We applied BH correction across all GWAS phenotypes, separately for Y and MT. We used FDR < 0.05 to define significance.

Gene expression association

Our previous work mapping cis expression quantitative trait loci (eQTLs) showed a linear mixed model is unnecessary (Munro et al. 2022). Therefore, for computational simplicity, we approached gene expression analysis using methods standard in differential expression (DE) analysis, treating Y and MT haplotype as "conditions", instead of eQTL mapping.

We used RNA-seq data processed using mRatBN7.2, presented as "log2" read count for all 10 tissues which were available from RatGTEx (Supplementary Table S2). The following filtering scheme was applied (separately for Y and MT): (1) samples with a haplotype assignment were retained, (2) for each tissue, genes that had detectable expression in less than 10% of samples were excluded.

We normalized counts using trimmed mean of M values (TMM; Robinson and Oshlack 2010), then used a Mann-Whitney test for DE. This test is robust to violation of a distribution (e.g. negative binomial) in large-sample DE analysis (Li et al. 2022). We again used FDR < 0.05 < 0.05 to define significance for all genes, in all tissues, for both the Y and MT Chromosomes.

A standard eQTL expression normalization method, which involves ranking genes within a sample (Munro et al. 2022), is nonoptimal for highly expressed genes, ranked highly in all samples. Ranking loses raw abundance information by introducing ties between ranks. Thus, we used TMM, which is a standard DE method (Corchete et al. 2020; Zhao et al. 2021).

Dkc1 expression and X SNPs association

For *Dkc1* (ENSRNOG00000055562), we tested for association between its TMM-normalized expression and SNPs on the X Chromosome. SNPs underwent the same filters as the GRM, except that only female rats were used for Hardy-Weinberg equilibrium tests. We used a simple linear model that regressed expression level against SNP genotype (0, 1, 2 copies of alternative allele), and, for male rats, Y haplogroup. We calculated P-values for all SNP associations where at least 5 rats had the minor allele. Associations were performed separately for each sex.

Results

Two major versions of Y are present in modern HS rats

All HS founders have distinct Y Chromosomes (Fig. 1a). BN, ACI, and MR are relatively similar to one another, and are also similar to the reference genome (which is based on BN), while the other five founders form a separate haplogroup.

We separated modern HS rats into two Y groups. We called 5,227 Y SNPs in 7,483 low-coverage samples from male modern HS rats. Coverage was generally low with a few spikes around the center (Supplementary Fig. S4). 4,132 SNPs and 7,471 samples remained after filtration by INFO score, MAF, and missingness (Supplementary Fig. S2a-d). SNPs in the pseudo-autosomal region are likely removed by these filters, as SNPs with high heterozygosity appear mid-filtration as having excess missingness. We grouped samples by whether they had more reference (Y1; 4,732 rats) or alternate (Y2; 2,739 rats) SNP alleles (Fig. 1b). Y1 is slightly more common in the modern male HS rat population (Fig. 1c).

Using STR data for a subset of 44 modern male HS rats that had sufficient coverage to call structural variants, we found ACI to be the most recent common ancestor of modern Y1 rats, while modern Y2 rats are closest to M520 (Supplementary Fig. S5).

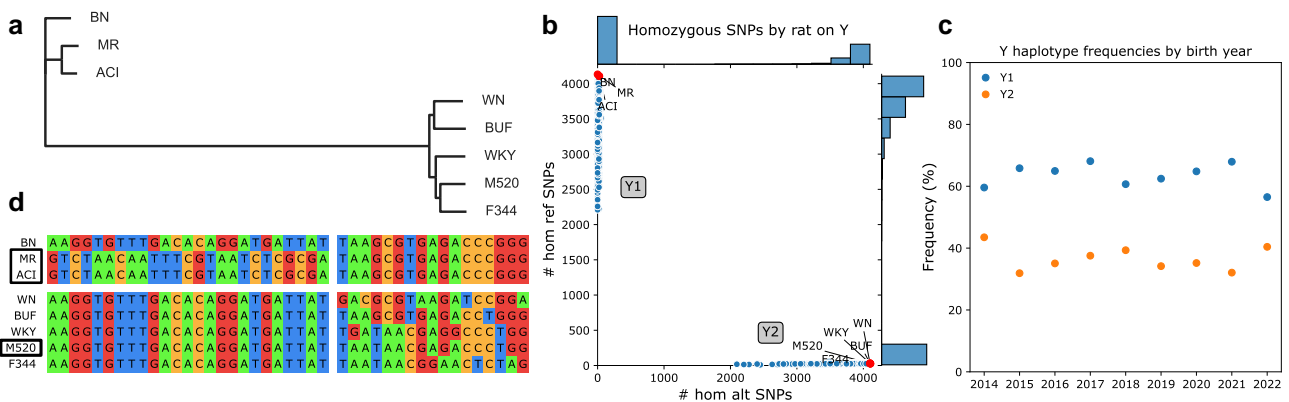


Fig. 1. Y haplogroups present in HS founders and modern HS rats. **a**) NJ, unrooted tree using Y SNPs and indels in HS founders. Branch lengths correspond to genetic distance. **b**) Distribution of alleles by rat among Y SNPs passing filters (Supplementary Fig. S2). Plot shows count of reference alleles on X-axis and count of alternate alleles on Y-axis for each rat. Side plots are histograms of allele counts in modern HS rats (small dots). Missingness in low-coverage modern samples leads to scatter on the axes. Labeled large dots are HS founders. Y1 and Y2 haplogroups are labeled. **c**) Distribution of Y haplogroups in the HS rat population over time. Plot shows birth year on X-axis and haplogroup percentage on Y-axis. **d**) Pseudo-alignment of Y1 (top) or Y2 (bottom) founder haplotypes, at SNPs passing filters (Supplementary Fig. S2), where Y1 (left) or Y2 (right) founders have intragroup SNPs. Boxed founders have haplotypes present in modern HS rats.

We next found the consensus for each Y haplogroup. We use the same filters on the low-coverage samples, except for skipping MAF to retain newly fixed variants. We used these data to determine the consensus SNP genotypes for each haplogroup. The consensus differ by 4,130 SNPs (S6A-C). We matched these consensus to founders in their haplogroup; the results (boxed in Fig. 1d) agree with the founders identified using STR data.

There is negligible variation in SNP genotypes among Y1 rats (Supplementary Fig. S7a). Y2 has more variation; over 80 rats differ at one SNP (Supplementary Fig. S7b), possibly a mutation from the parent haplotype M520. We also identified a subgroup of Y1 with an apparent deletion of *Med14Y* (ENSRNOG0000060437) which low-coverage data cannot confidently call; thus, it was not used for association analyses (Supplementary File S1). Neither DNA-seq nor RNA-seq principal component analysis (PCA) revealed other groupings, except by library preparation method used for low-coverage sequencing (Supplementary Fig. S8a-b).

Two versions of MT are present in modern HS rats

We found four MT haplotypes among the eight HS founders (Fig. 2a). BUF, F344, M520, MR, and WN share SNPs relative to BN, the basis of mRatBN7.2. WKY has a distinct MT haplotype, which was not observed among modern HS rats. The ACI haplotype is barely distinct from BUF, F344, M520, MR, and WN.

MT phylogeny of HS founders has been reported previously (Showmaker et al. 2020). However, their data (Ramdas et al. 2018) swapped WN and WKY. Our data puts WKY by itself, and WN in the large founder block with BUF, F344, M520, and MR. The Rat Genome Database (RGD; Vedi et al. 2023) variant visualizer (parameters: strains = HS founder strains, chromosome = MT, start = 0, end = 16,313) confirms the groups in Fig. 2. Complete MT genome sequencing of inbred substrains related to four of the HS founders (ACI/Eur, BN/NHsdMcwi, F344/NHsd, and WKY/Ncrl) found the same relationships (Schlick et al. 2006).

We separated modern HS rats into MT groups. We called 117 MT SNPs in 15,120 low-coverage samples from modern HS rats, with higher coverage in lcWGS samples and several defined peaks of coverage in ddGBS samples (Supplementary Figs. S9, S10).

77 SNPs and 14,971 samples remained after filtration by INFO score, MAF, and missingness (Supplementary Fig. S2e-h). We grouped samples by whether they had more reference (MT1, 9,287 rats) or alternate (MT2, 5,684 rats) SNP alleles (Fig. 2b). MT1 is somewhat more common in the modern HS rat population (Fig. 2c).

We confirmed these as the only two MT haplotypes present in the modern low-coverage SNP genotypes. All modern HS rat MT match at least one of two SNP haplotypes, ignoring missingness. These two modern MT haplotypes are separated by 77 SNPs (Supplementary Fig. S6d-f). Each modern MT matches an ostensibly extant founder haplotype (Fig. 2d). Neither DNA-seq or RNA-seq PCA revealed further groupings, except by library preparation method (Supplementary Fig. S8c-d).

Y haplogroup is associated with Y gene expression

We investigated the effect of Y haplogroup on various phenotypes. Y haplogroup was not significantly associated with any of the phenotypes examined (Fig. 3a), except for levels of MZ531.3646417_5.08009 (Supplementary Fig. S11), an unannotated metabolite that was measured in the cecum. Y haplogroup was associated with expression of *Ddx3y* and *Dkc1*, both of which are located on the Y Chromosome (Fig. 3b-d, Supplementary Table S3).

Ddx3y is an RNA helicase. In humans, it is involved with neuron development in males (Vakilian et al. 2015). Its male-specificity is sometimes used for determining sex, e.g. in humans (Hoch et al. 2020) and pigs (Teixeira et al. 2019). Consistent with this application, we found that *Ddx3y* was not expressed in female rats (Fig. 3c).

Mutations in *Dkc1*'s human ortholog cause X-linked dyskeratosis congenita (Heiss et al. 1998); many orthologs of this gene are on the X Chromosome (RGD). In mRatBN7.2 *Dkc1* is on an unplaced Y Chromosome contig; in the most recent rat reference genome, GRCh8, it is on the X Chromosome (RGD). Unlike *Ddx3y*, *Dkc1* is expressed in females (Fig. 3d). This suggests *Dkc1* is in the pseudo-autosomal region. However, the rat pseudo-autosomal region, which has lost many genes to autosomes (Maxeiner et al.

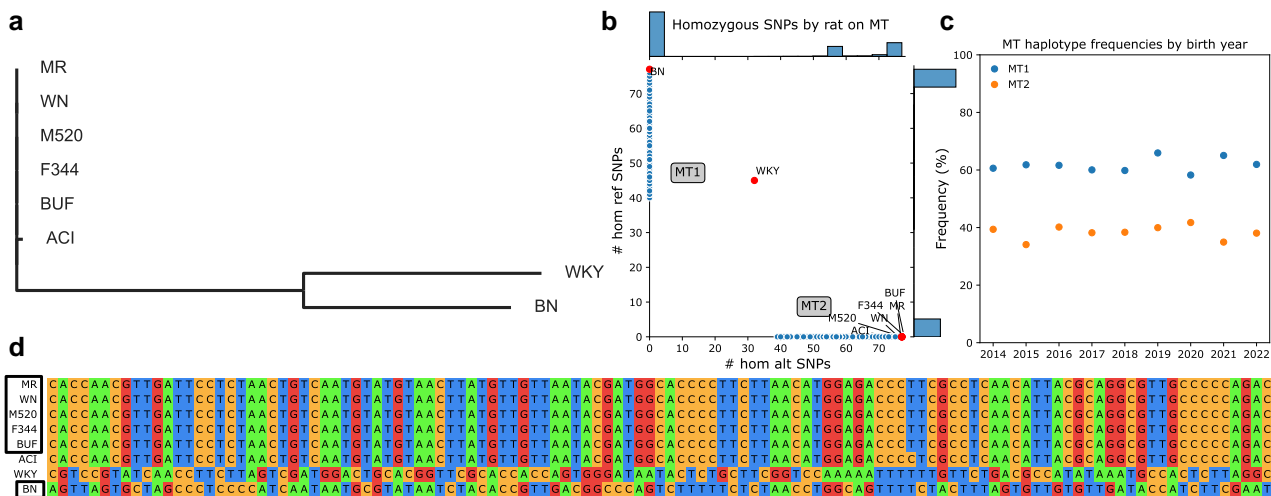


Fig. 2. MT haplotypes present in HS founders and modern HS rats. a) NJ, unrooted tree using MT SNPs and indels in HS founders. Branch lengths correspond to genetic distance. b) Distribution of alleles by rat among MT SNPs passing filters (Supplementary Fig. S2). Plot shows count of reference alleles on X-axis and count of alternate alleles on Y-axis for each rat. Side plots are histograms of allele counts in modern HS rats (small dots). Missingness in low-coverage modern samples leads to scatter on the axes. Labeled large dots are HS founders. MT1 and MT2 haplogroups are labeled. c) Distribution of MT haplotypes in the HS rat population over time. Plot shows birth year on X-axis and haplotype percentage on Y-axis. d) Pseudo-alignment of founder haplotypes, at all MT SNPs called by low-coverage sequencing, colored by nucleotide. Boxed founders have haplotypes present in modern HS rats.

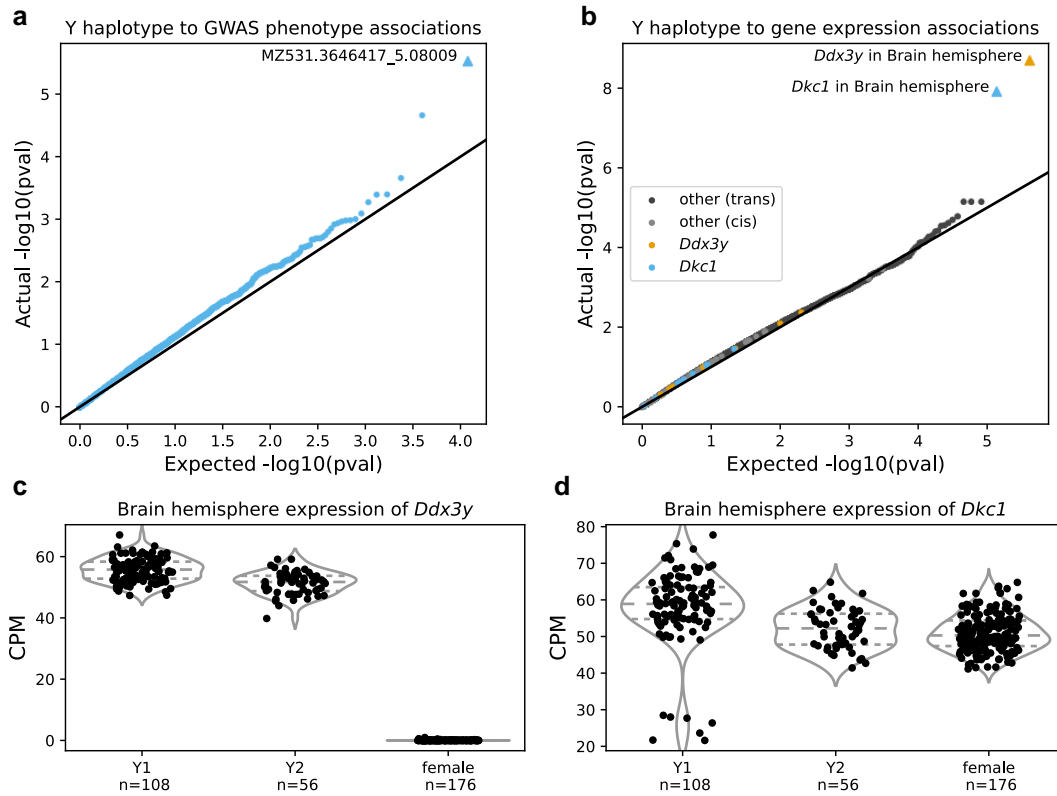


Fig. 3. Results of Y haplogroup association tests. a–b) QQ plots of a) MLMAs between Y haplogroup and GWAS phenotypes and b) Mann–Whitney tests between Y haplogroup and gene expression. Each dot represents a single trait (in b, a single gene in a single tissue, colored by gene). Plot shows actual distribution of unadjusted P -values on Y-axis, against expected distribution (null hypothesis of no association) on X-axis. Significant associations ($FDR < 0.05$) are triangles. c–d) Effect plots of c) *Ddx3y* and d) *Dkc1* expression in the brain, split by Y haplogroup, with females for context. Horizontal lines show quantiles. Plots show each sample's normalized CPM on Y-axis; samples are split into Y haplogroups on X-axis. Q-values in [Supplementary Table S3](#).

2020), does not have well-established boundaries (Raudsepp and Chowdhary 2016). We did not find strong evidence of linkage between SNPs on the X Chromosome and *Dkc1* expression ([Supplementary Fig. S12](#)), which would have supported its localization to the pseudo-autosomal region. The subgroup of Y1 rats with low *Dkc1* expression is more readily explained by a related presence–absence variant ([File S1](#)).

MT haplotype is associated with gene expression

We investigated the effect of MT haplotype on all available phenotypes; none of the results were significant ([Fig. 4a](#)), despite sufficient sample size and minor haplotype frequency for good power ([Supplementary Fig. S3](#)). In addition, we separately tested for association with having one or two kidneys at birth, as [Showmaker et al. \(2020\)](#) hypothesized an effect from BN (MT1) vs. ACI (similar to MT2) MT genotype. MT1 rats have a higher rate of being born with a single kidney ([Supplementary Table S5](#)) but a one-sided Fisher's exact test against MT haplotype was insignificant ($P = 0.14$). However, MT haplotype was associated with the expression of several MT genes ([Fig. 4b–f](#), [Supplementary Fig. S11](#), [Table S4](#)).

Complex I is the first enzyme in the electron transport chain. In 7 of 10 tissues tested, its *Mt-nd3* subunit is up-regulated in MT2 relative to MT1. Every other MT-encoded subunit (*Mt-nd1*, *Mt-nd2*, *Mt-nd4*, *Mt-nd4l*, *Mt-nd5*, *Mt-nd6*) is down-regulated in MT2, resulting in different subunit ratios. Also, both MT-encoded ribosomal RNAs have significant DE, which are possible artifacts of imperfect poly-A tail selection.

Discussion

We performed a large-scale study to identify phenotypes influenced by the nonrecombinant Y and MT Chromosomes in 12,055 HS rats. The 8 founders of the HS population had two major Y Chromosome ([Fig. 1a](#)) and three major MT Chromosome ([Fig. 2a](#)) haplotype groups. In modern HS rats, we observed two Y haplogroups (Y1 and Y2; [Fig. 1b](#)), with minimal intragroup variation. Similarly, in modern HS rats we observed two MT haplotypes (MT1 and MT2; [Fig. 2b](#)), with no genotyped intragroup variation.

We assigned 12,055 rats who were both phenotyped (n phenotyped = 12,116) and haplotyped (n haplotyped = 15,042) to Y1 or Y2 (for males) and MT1 or MT2 haplotypes and then sought to identify associations with an array of behavioral and physiological phenotypes. Remarkably, there were virtually no significant associations ([Figs. 3a, 4a](#), [Supplementary Table S5](#)). For both Y and MT, we identified cis-located genes with DE between haplogroups ([Figs. 3b, 4b](#)). While these eQTLs do not cause detectable changes in the behavioral and physiological traits we studied, they provide an important positive control, demonstrating that we can accurately call Y and MT haplotypes. Overall, our results show that previous genetics studies in HS rats which did not examine the Y and MT Chromosomes, did not overlook important genetic effects.

A strength of our study is the fact that the genetic structure of HS rats makes them well suited for studying Y and MT. Whereas human studies can be confounded by correlations between MT and nuclear genotype ([Hagen et al. 2018](#)), the HS breeding strategy ([Solberg Woods and Mott 2017](#)) and our use of MLMA for PheWAS

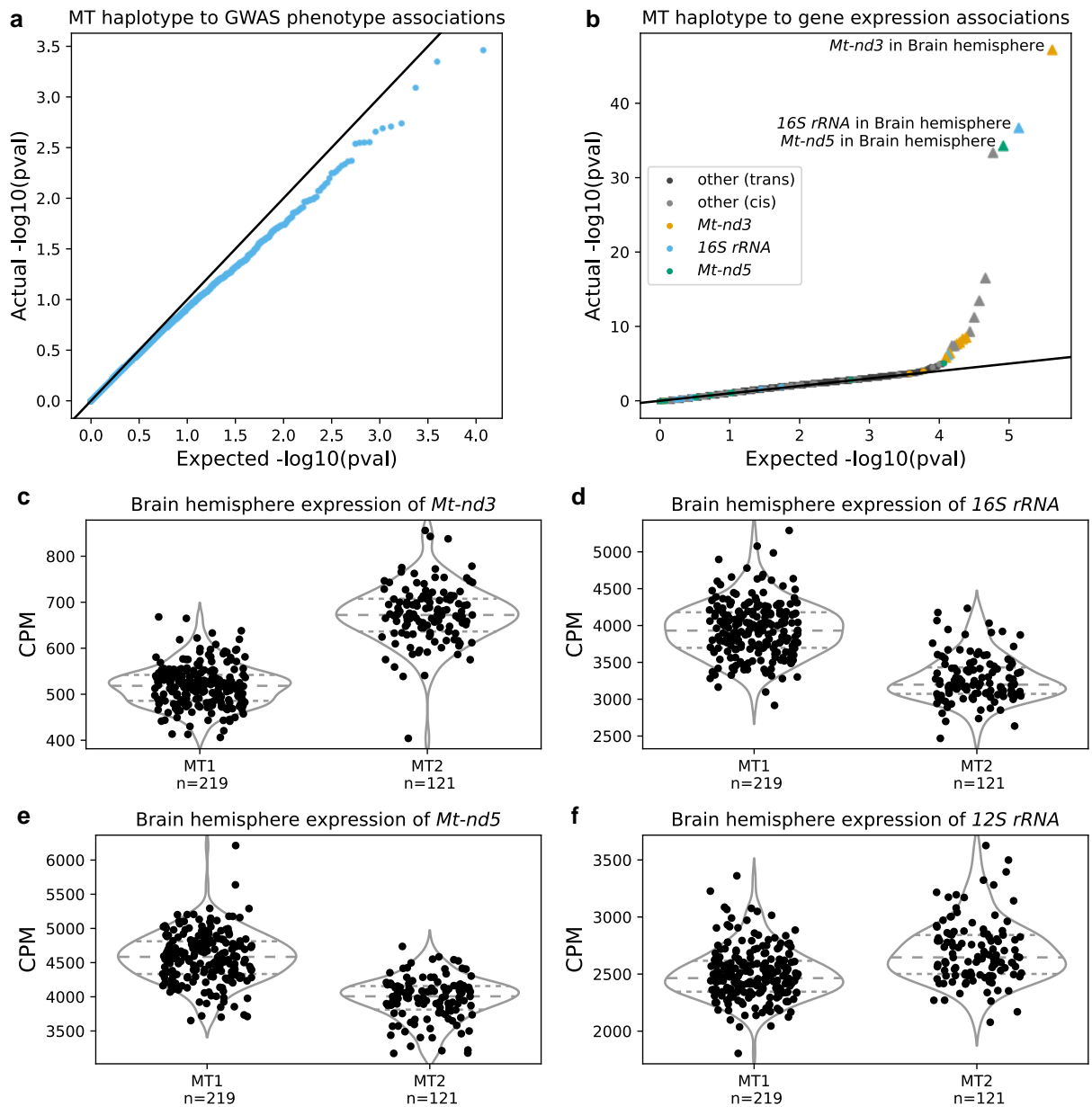


Fig. 4. Results of MT haplotype association tests. a–b) QQ plots of a) MLMAs between MT haplotype and GWAS phenotypes and b) Mann–Whitney tests between MT haplotype and gene expression. Each dot represents a single trait (in b, a single gene in a single tissue, colored by gene). Plot shows actual distribution of unadjusted P-values on Y-axis, against expected distribution (null hypothesis of no association) on X-axis. Significant associations (FDR < 0.05) are triangles. c–f) Representative effect plots for significant associations, of c) *Mt-nd3*, d) *16S rRNA*, e) *Mt-nd5*, and f) *12S rRNA* expression in the brain, split by MT haplotype. Plots show each sample’s normalized CPM on Y-axis; samples are split into MT haplotypes on X-axis. Effect plots for all significant associations with MT haplotype are shown in [Supplementary Fig. S13](#). Q-values in [Supplementary Table S4](#).

avoided these problems. In addition, all of the observed Y or MT haplotypes are very common (Figs. 1c, 2c), unlike the situation in DO mice (Chesler et al. 2016) or humans (Howe et al. 2017), providing good power to detect associations.

Our results indicate that only a few of the founder Y and MT Chromosomes have persisted into modern HS rats. There are possibly further subtle Y and MT variations which our data could not distinguish. The lost haplotypes could reflect genetic drift or inadvertent selection due to differences in fitness or fecundity; our data can not distinguish between these two possibilities. Thus, it is possible that some of the unobserved Y and MT Chromosomes

would have shown phenotypic consequences had they been present among the modern HS rats that we studied. For example, rats with comparable nuclear genotypes and a BN or WKY MT differ on various physiological traits (Sathiaseelan et al. 2023). However, as WKY’s MT is not observed in modern HS rats, we were unable to test for its effect on physiological traits.

In summary, we describe Y and MT haplotype structure in modern HS rats, and present results from well-powered association analyses with various phenotypes. Haplotypes are inherited from specific HS founders and cause differential expression of several genes of biological importance, including Complex I

subunits and genes with orthologs to human sex-linked disorders. Methods described here may be extended to other rat populations for further investigation of Y and MT.

Data availability

HS rats are available at <https://ratgenes.org/cores/core-b/>. All data required to reproduce these analyses, and raw results (with unadjusted *P*-values) from association tests, are in UC San Diego Library Digital Collections with DOI [10.6075/J0VX0GQQ](https://doi.org/10.6075/J0VX0GQQ). (Filename key in [File S2](#).) Code to reproduce these analyses is in Zenodo with DOI [10.5281/zenodo.11493119](https://doi.org/10.5281/zenodo.11493119), and in [GitHub](#). Low-coverage raw reads are in the NCBI SRA with accession PRJNA1022514. Low-coverage autosomal genotypes used for the GRM are in UCSD Library with DOI [10.6075/J00G3KBX](https://doi.org/10.6075/J00G3KBX). Low-coverage genotyping was conducted at the IGM Genomics Center.

[Supplemental material](#) available at G3 online.

Acknowledgements

The authors thank Robert Vogel for advice in statistical analyses, Gregory Keele and Gary Churchill for useful insights, and Ryan Eveloff for unpublished STR data. We used the [Triton Shared Computing Cluster](#) from the San Diego Supercomputer Center.

Funding

This work was supported by the National Institute on Drug Abuse (P50DA037844) and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (R01AR070879).

Conflicts of interest

RK is a scientific advisory board member, and consultant for BiomeSense, Inc., has equity and receives income. He is a scientific advisory board member and has equity in GenCirq. He is a consultant for DayTwo, and receives income. He has equity in and acts as a consultant for Cybele. He is a co-founder of Biota, Inc., and has equity. He is a cofounder of Micronoma, and has equity and is a scientific advisory board member. The terms of these arrangements have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

Literature cited

- Baud A, Hermsen R, Guryev V, Stridh P, Graham D, McBride MW, Foroud T, Calderari S, Diez M, Ockinger J, et al. 2013. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nat Genet.* 45(7):767–775. doi:[10.1038/ng.2644](https://doi.org/10.1038/ng.2644)
- Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol).* 57(1):289–300. doi:[10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)
- Broman KW. 2022. A generic hidden Markov model for multiparent populations. *G3 Genes Genom Genet.* 12(2):jkab396. doi:[10.1093/g3journal/jkab396](https://doi.org/10.1093/g3journal/jkab396)
- Cai N, Gomez-Duran A, Yonova-Doing E, Kundu K, Burgess AI, Golder ZJ, Calabrese C, Bonder MJ, Camacho M, Lawson RA, et al. 2021. Mitochondrial DNA variants modulate N-formylmethionine, proteostasis and risk of late-onset human diseases. *Nat Med.* 27(9):1564–1575. doi:[10.1038/s41591-021-01441-3](https://doi.org/10.1038/s41591-021-01441-3)
- Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 4(1):7. doi:[10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8)
- Chen D. 2022. Palmer lab high coverage WGS genotyping pipeline. doi:[10.5281/zenodo.6584834](https://doi.org/10.5281/zenodo.6584834)
- Chen D, Chitre A, Cheng R, Peng B, Polesskaya O, Palmer A. 2023. Palmer lab heterogeneous stock rats genotyping pipeline. doi:[10.5281/zenodo.10002191](https://doi.org/10.5281/zenodo.10002191)
- Chen D, Chitre AS, Nguyen KMH, Cohen K, Peng B, Ziegler KS, Okamoto F, Lin B, Johnson BB, Sanches TM. 2024. A cost-effective, high-throughput, highly accurate genotyping method for outbred populations. doi:[10.1101/2024.07.17.603984](https://doi.org/10.1101/2024.07.17.603984)
- Chen H, Lu Y, Lu D, Xu S. 2021. Y-LineageTracker: a high-throughput analysis framework for Y-chromosomal next-generation sequencing data. *BMC Bioinformatics.* 22:114. doi:[10.1186/s12859-021-04057-z](https://doi.org/10.1186/s12859-021-04057-z)
- Chesler EJ, Gatti DM, Morgan AP, Strobel M, Trepanier L, Oberbeck D, McWeeney S, Hitzemann R, Ferris M, McMullan R. 2016. Diversity outbred mice at 21: maintaining allelic variation in the face of selection. *G3 Genes Genom Genet.* 6:3893–3902. doi:[10.1534/g3.116.035527](https://doi.org/10.1534/g3.116.035527)
- Chitre AS, Polesskaya O, Holl K, Gao J, Cheng R, Bimschleger H, Garcia Martinez A, George T, Gileta AF, Han W, et al. 2020. Genome-wide association study in 3,173 outbred rats identifies multiple loci for body weight, adiposity, and fasting glucose. *Obesity.* 28:1964–1973. doi:[10.1002/oby.22927](https://doi.org/10.1002/oby.22927)
- Corchete LA, Rojas EA, Alonso-López D, De Las Rivas J, Gutiérrez NC, Burguillo FJ. 2020. Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis. *Sci Rep.* 10:19737. doi:[10.1038/s41598-020-76881-x](https://doi.org/10.1038/s41598-020-76881-x)
- Davies RW, Flint J, Myers S, Mott R. 2016. Rapid genotype imputation from sequence without reference panels. *Nat Genet.* 48:965–969. doi:[10.1038/ng.3594](https://doi.org/10.1038/ng.3594)
- Degenhardt F, Ellinghaus D, Juzenas S, Lerga-Jaso J, Wendorff M, Maya-Miles D, Uellendahl-Werth F, ElAbd H, Rühlemann MC, Arora J, et al. 2022. Detailed stratified GWAS analysis for severe COVID-19 in four European populations. *Hum Mol Genet.* 31:3945–3966. doi:[10.1093/hmg/ddac158](https://doi.org/10.1093/hmg/ddac158)
- de Jong TV, Pan Y, Rastas P, Munro D, Tutaj M, Akil H, Benner C, Chen D, Chitre AS, Chow W, et al. 2024. A revamped rat reference genome improves the discovery of genetic diversity in laboratory rats. *Cell Genom.* 4:100527. doi:[10.1016/j.xgen.2024.100527](https://doi.org/10.1016/j.xgen.2024.100527)
- Gileta AF, Gao J, Chitre AS, Bimschleger HV, Gopalakrishnan S, Palmer AA. 2020. Adapting genotyping-by-sequencing and variant calling for heterogeneous stock rats. *G3: Genes Genom Genet.* 10:2195–2205. doi:[10.1534/g3.120.401325](https://doi.org/10.1534/g3.120.401325)
- Gonçalves VF, Giamberardino SN, Crowley JJ, Vawter MP, Saxena R, Bulik CM, Yilmaz Z, Hultman CM, Sklar P, Kennedy JL, et al. 2018. Examining the role of common and rare mitochondrial variants in schizophrenia. *PLoS One.* 13:e0191153. doi:[10.1371/journal.pone.0191153](https://doi.org/10.1371/journal.pone.0191153)
- Hagen CM, Gonçalves VF, Hedley PL, Bybjerg-Grauholm J, Bækvad-Hansen M, Hansen CS, Kanters JK, Nielsen J, Mors O, Demur AB, et al. 2018. Schizophrenia-associated mt-DNA SNPs exhibit highly variable haplogroup affiliation and nuclear ancestry: Bi-genomic dependence raises major concerns for link to disease. *PLoS One.* 13:e0208828. doi:[10.1371/journal.pone.0208828](https://doi.org/10.1371/journal.pone.0208828)
- Haines BA, Barradale F, Dumont BL. 2021. Patterns and mechanisms of sex ratio distortion in the collaborative cross mouse mapping population. *Genetics.* 219:iyab136. doi:[10.1093/genetics/iyab136](https://doi.org/10.1093/genetics/iyab136)

- Hansen C, Spuhler K. 1984. Development of the national institutes of health genetically heterogeneous rat stock. *Alcohol Clin Exp Res*. 8:477–479. doi:[10.1111/j.1530-0277.1984.tb05706.x](https://doi.org/10.1111/j.1530-0277.1984.tb05706.x)
- Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, Dokal I. 1998. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet*. 19:32–38. doi:[10.1038/ng0598-32](https://doi.org/10.1038/ng0598-32)
- Hoch D, Novakovic B, Cvitic S, Saffery R, Desoye G, Majali-Martinez A. 2020. Sex matters: XIST and DDX3Y gene expression as a tool to determine fetal sex in human first trimester placenta. *Placenta*. 97:68–70. doi:[10.1016/j.placenta.2020.06.016](https://doi.org/10.1016/j.placenta.2020.06.016)
- Howe LJ, Erzurumluoglu AM, Davey Smith G, Rodriguez S, Stergiakouli E. 2017. Y chromosome, mitochondrial DNA and childhood behavioural traits. *Sci Rep*. 7:11655. doi:[10.1038/s41598-017-10871-4](https://doi.org/10.1038/s41598-017-10871-4)
- Jamain S, Quach H, Quintana-Murci L, Betancur C, Philippe A, Gillberg C, Sponheim E, Skjeldal OH, Fellous M, Leboyer M, et al. 2002. Y chromosome haplogroups in autistic subjects. *Mol Psychiatry*. 7:217–219. doi:[10.1038/sj.mp.4000968](https://doi.org/10.1038/sj.mp.4000968)
- Johannesson M, Lopez-Aumatell R, Stridh P, Diez M, Tuncel J, Blázquez G, Martínez-Membrives E, Cañete T, Vicens-Costa E, Graham D, et al. 2009. A resource for the simultaneous high-resolution mapping of multiple quantitative trait loci in rats: the NIH heterogeneous stock. *Genome Res*. 19:150–158. doi:[10.1101/gr.081497.108](https://doi.org/10.1101/gr.081497.108)
- Keele GR, Zhang T, Pham DT, Vincent M, Bell TA, Hock P, Shaw GD, Paulo JA, Munger SC, Pardo-Manuel de Villena F, et al. 2021. Regulation of protein abundance in genetically diverse mouse populations. *Cell Genom*. 1:100003. doi:[10.1016/j.xgen.2021.100003](https://doi.org/10.1016/j.xgen.2021.100003)
- Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat*. 32:25–32. doi:[10.1002/humu.21382](https://doi.org/10.1002/humu.21382)
- Li Y, Ge X, Peng F, Li W, Li JJ. 2022. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome Biol*. 23:79. doi:[10.1186/s13059-022-02648-4](https://doi.org/10.1186/s13059-022-02648-4)
- Ma J, Coarfa C, Qin X, Bonnen PE, Milosavljevic A, Versalovic J, Aagaard K. 2014. mtDNA haplogroup and single nucleotide polymorphisms structure human microbiome communities. *BMC Genomics*. 15:257. doi:[10.1186/1471-2164-15-257](https://doi.org/10.1186/1471-2164-15-257)
- Martincová I, Ďureje Ľ, Kreisinger J, Macholán M, Piálek J. 2019. Phenotypic effects of the Y chromosome are variable and structured in hybrids among house mouse recombinant lines. *Ecol Evol*. 9:6124–6137. doi:[10.1002/ece3.5196](https://doi.org/10.1002/ece3.5196)
- Maxeiner S, Benseler F, Krasteva-Christ G, Brose N, Südhof TC. 2020. Evolution of the autism-associated neuroligin-4 gene reveals broad erosion of pseudoautosomal regions in rodents. *Mol Biol Evol*. 37:1243–1258. doi:[10.1093/molbev/msaa014](https://doi.org/10.1093/molbev/msaa014)
- Mosquera-Miguel A, Torrell H, Abasolo N, Arrojo M, Paz E, Ramos-Ríos R, Agra S, Páramo M, Brenlla J, Martínez S, et al. 2012. No evidence that major mtDNA European haplogroups confer risk to schizophrenia. *Am J Med Genet Part B Neuropsychiatric Genet*. 159B:414–421. doi:[10.1002/ajmg.b.32044](https://doi.org/10.1002/ajmg.b.32044)
- Mousavi N, Margoliash J, Pusarla N, Saini S, Yanicky R, Gymrek M. 2020. TRTools: a toolkit for genome-wide analysis of tandem repeats. *Bioinformatics*. 37:731–733. doi:[10.1093/bioinformatics/btaa736](https://doi.org/10.1093/bioinformatics/btaa736)
- Munro D, Wang T, Chitre AS, Poleskaya O, Ehsan N, Gao J, Gusev A, Woods LS, Saba L, Chen H, et al. 2022. The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. *Nucleic Acids Res*. 50:10882–10895. doi:[10.1093/nar/gkac912](https://doi.org/10.1093/nar/gkac912)
- Nelson VR, Spiezio SH, Nadeau JH. 2010. Transgenerational genetic effects of the paternal Y chromosome on daughters' phenotypes. *Epigenomics*. 2:513–521. doi:[10.2217/epi.10.26](https://doi.org/10.2217/epi.10.26)
- Ramdas S, Ozel AB, Li J, Solberg Woods L. Figshare. 2018.
- Raudsepp T, Chowdhary BP. 2016. The eutherian pseudoautosomal region. *Cytogenet Genome Res*. 147:81–94. doi:[10.1159/000443157](https://doi.org/10.1159/000443157)
- Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 11:R25. doi:[10.1186/gb-2010-11-3-r25](https://doi.org/10.1186/gb-2010-11-3-r25)
- Sathiaseelan R, Ahn B, Stout M, Logan S, Wanagat J, Nguyen H, Hord N, Vandiver A, Selvarani R, Ranjit R, et al. 2023. A genetically heterogeneous rat model with divergent mitochondrial genomes. *J Gerontol Ser A*. 78:771–779. doi:[10.1093/gerona/glad056](https://doi.org/10.1093/gerona/glad056)
- Schlick NE, Jensen-Seaman MI, Orlebeke K, Kwitek AE, Jacob HJ, Lazar J. 2006. Sequence analysis of the complete mitochondrial DNA in 10 commonly used inbred rat strains. *Am J Physiol Cell Physiol*. 291:C1183–C1192. doi:[10.1152/ajpcell.00234.2006](https://doi.org/10.1152/ajpcell.00234.2006)
- Showmaker KC, Cobb MB, Johnson AC, Yang W, Garrett MR. 2020. Whole genome sequencing and novel candidate genes for CAKUT and altered nephrogenesis in the HSRA rat. *Physiol Genomics*. 52:56–70. doi:[10.1152/physiolgenomics.00112.2019](https://doi.org/10.1152/physiolgenomics.00112.2019)
- Solberg Woods LC, Mott R. 2017. Heterogeneous stock populations for analysis of complex traits. *Methods Mol Biol (Clifton, N.J.)*. 1488:31–44. doi:[10.1007/978-1-4939-6427-7_2](https://doi.org/10.1007/978-1-4939-6427-7_2)
- Talevich E, Invergo BM, Cock PJ, Chapman BA. 2012. Bio.Phylo: a unified toolkit for processing, analyzing and visualizing phylogenetic trees in Biopython. *BMC Bioinformatics*. 13:209. doi:[10.1186/1471-2105-13-209](https://doi.org/10.1186/1471-2105-13-209)
- Teixeira SA, Ibelli AMG, Cantão ME, de Oliveira HC, Ledur MC, Peixoto JdO, Marques DBD, Costa KA, Coutinho LL, Guimarães SEF. 2019. Sex determination using RNA-sequencing analyses in early prenatal pig development. *Genes*. 10:1010. doi:[10.3390/genes10121010](https://doi.org/10.3390/genes10121010)
- Tutaj M, Smith JR, Bolton ER. 2019. Rat genome assemblies, annotation, and variant repository. In: Hayman GT, Smith JR, Dwinell MR, Shimoyama M, editors, *Rat Genomics*. New York (NY): Springer (Methods in Molecular Biology) p. 43–70. doi:[10.1007/978-1-4939-9581-3_2](https://doi.org/10.1007/978-1-4939-9581-3_2)
- Vakilian H, Mirzaei M, Sharifi Tabar M, Pooyan P, Habibi Rezaee L, Parker L, Haynes PA, Gourabi H, Baharvand H, Salekdeh GH. 2015. DDX3Y, a male-specific region of Y chromosome gene, may modulate neuronal differentiation. *J Proteome Res*. 14:3474–3483. doi:[10.1021/acs.jproteome.5b00512](https://doi.org/10.1021/acs.jproteome.5b00512)
- Vedi M, Smith JR, Thomas Hayman G, Tutaj M, Brodie KC, De Pons JL, Demos WM, Gibson AC, Kaldunski ML, Lamers L, et al. 2023. 2022 updates to the rat genome database: a findable, accessible, interoperable, and reusable (FAIR) resource. *Genetics*. 224:iyad042. doi:[10.1093/genetics/iyad042](https://doi.org/10.1093/genetics/iyad042)
- Wang G, Zhang Y, Zhang Y, Dong Y, Lv Z, Sun M, Wu D, Wu Y. 2013. Mitochondrial haplogroups and hypervariable region polymorphisms in schizophrenia: a case-control study. *Psychiatry Res*. 209:279–283. doi:[10.1016/j.psychres.2013.01.001](https://doi.org/10.1016/j.psychres.2013.01.001)
- Welch DR, Larson MA, Vivian CJ, Vivian JL. 2023. Generating mitochondrial-nuclear exchange (MNX) mice to identify mitochondrial determinants of cancer metastasis. In: Kasid UN, Clarke R, editors, *Cancer Systems and Integrative Biology*. New York (NY): Springer US (Methods in Molecular Biology) p. 43–59. doi:[10.1007/978-1-0716-3163-8_4](https://doi.org/10.1007/978-1-0716-3163-8_4)
- Wigginton JE, Cutler DJ, Abecasis GR. 2005. A note on exact tests of hardy-weinberg equilibrium. *Am J Hum Genet*. 76:887–893. doi:[10.1086/429864](https://doi.org/10.1086/429864)
- Willems T, Zielinski D, Yuan J, Gordon A, Gymrek M, Erlich Y. 2017. Genome-wide profiling of heritable and de novo STR variations. *Nat Methods*. 14:590–592. doi:[10.1038/nmeth.4267](https://doi.org/10.1038/nmeth.4267)
- Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW. 2010.

- Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 42:565–569. doi:[10.1038/ng.608](https://doi.org/10.1038/ng.608)
- Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 88:76–82. doi:[10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011)
- Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet.* 46:100–106. doi:[10.1038/ng.2876](https://doi.org/10.1038/ng.2876)
- Zhao Y, Li MC, Konaté MM, Chen L, Das B, Karlovich C, Williams PM, Evrard YA, Doroshow JH, McShane LM. 2021. TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI patient-derived models repository. *J Transl Med.* 19:269. doi:[10.1186/s12967-021-02936-w](https://doi.org/10.1186/s12967-021-02936-w)

Editor: L. McIntyre