# Characterizing the sequence and expression evolution of the t-haplotype, a model meiotic driver

by

**Réka Kelemen**

June, 2024

*A thesis submitted to the*
*Graduate School*
*of the*
*Institute of Science and Technology Austria*
*in partial fulfillment of the requirements*
*for the degree of*
*Doctor of Philosophy*

Committee in charge:

Jan Maas, Chair
Beatriz Viçoso
Nick Barton
Andrew G. Clark

The thesis of Réka Kelemen, titled *Characterizing the sequence and expression evolution of the t-haplotype, a model meiotic driver*, is approved by:

**Supervisor**: Beatriz Viçoso, ISTA, Klosterneuburg, Austria

Signature: _____

**Committee Member**: Nick Barton, ISTA, Klosterneuburg, Austria

Signature: _____

**Committee Member**: Andrew G. Clark, Cornell University, Ithaca, NY, USA

Signature: _____

**Defense Chair**: Jan Maas, ISTA, Klosterneuburg, Austria

Signature: _____

Signed page is on file

ISTA Thesis, ISSN: 2663-337X

ISBN: 978-3-99078-039-8

I hereby declare that this thesis is my own work and that it does not contain other people's work without this being so stated; this thesis does not contain my previous work without this being stated, and the bibliography contains all the literature that I used in writing the dissertation.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee, and that this thesis has not been submitted for a higher degree to any other university or institution.

I certify that any republication of materials presented in this thesis has been approved by the relevant publishers and co-authors.

Signature: _____

Réka Kelemen
June, 2024

Signed page is on file

# Abstract

Genomes are shaped by natural selection at the level of the organism, as genomic variants that have a beneficial effect on the viability or fecundity of their carriers are on average expected to be passed on to more offspring than less beneficial alleles. However, selection also favors genomic variants that drive their own transmission to the next generation above the mendelian expectation of 50 percent in heterozygotes, even if these self-promoting variants are less beneficial to the organism than other variants at the same locus. Such variants, called meiotic drivers, are found in diverse taxa, and often impose fitness costs on their host organisms. As meiotic drivers often require multiple genes and sequences for transmission ratio distortion, they are often found in regions of low recombination, such as inversions, which prevent their recombination with the non-driving homologous regions. Reduced recombination rates are expected to lead to the accumulation of deleterious mutations, which may affect hundreds of genes trapped in the inversions of meiotic drivers. Although the observed fitness costs of self-promoting haplotypes are thought to possibly reflect sequence degeneration, no study has systematically investigated the level of degeneration on a meiotic driver. Further, the low rates of recombination between driving and non-driving haplotypes have limited the power of traditional genetic studies in uncovering the gene content of meiotic drivers, and made the the identification of the genes causing transmission ratio distortion difficult.

After an introduction to meiotic drivers in Chapter 1, this thesis presents three studies that make use of next generation sequencing data to characterize the sequence and expression evolution of genes on the *t*-haplotype, a large and ancient meiotic driver in house mice that is transmitted to up to 100% of the offspring in males heterozygous for it. Chapter 2 presents a comprehensive assessment of the *t*-haplotype's sequence evolution, which shows signs of sequence degeneration counteracted by occasional recombination with the non-driving homolog over large parts of the meiotic driver, proposing an explanation for its long-term survival. Chapter 3 investigates the sequence and expression evolution of genes on the *t*-haplotype, and finds widespread expression and copy number changes and signs of less efficient purifying selection compared to the genes on the non-driving homolog. Further, this chapter finds candidates for involvment in drive: two positively selected genes on the *t*-haplotype, and the discovery of a *t*-specific gene duplicate, which was gained from another chromosome, and which acquired novel sequence and testis-specific expression on the *t*-haplotype. Finally, Chapter 4 provides unprecedented insights into the gene expression landscape in testes of *t*-carrier mice, using single nucleus sequencing. Cell-resolved RNA-sequencing allows the comparison of expression in spermatids carrying or not carrying the *t*-haplotype as well as the timing of *t*-haplotype-induced expression changes along spermatogenesis. This study shows the timing of previously found drive-associated genes, and uncovers novel candidate genes and biological processes that may underlie the complex biology of transmission ratio distortion of the *t*-haplotype. Chapter 5 synthesizes the findings of the three studies, and discusses them in the context of the current state of meiotic drive research.

# Acknowledgements

First of all, I would like to thank my supervisor, Beatriz Viçoso, who all throughout my PhD was a source of motivation, encouragement, and empathy. She was a role model for being a great and enthusiastic scientist.

I would like to thank members of the Viçoso group for collaborations, discussions and support both scientifically and personally.

I am grateful to my PhD committee members: Nick Barton, Fyodor Kondrashov and Andy Clark, who gave me very useful feedback about my projects and career year after year.

ISTA offered an exceptional working environment with helpful staff and a vibrant scientific community. In particular, the evolutionary biology community at ISTA provided great discussions with nice colleagues and a good seminar series that broadened my knowledge of the field.

My PhD work was supported by the European Research Council under the European Unions Horizon 2020 research and innovation program (grant agreement no. 715257) and by the Austrian Science Foundation (FWF SFB F88-10), grants that were received by Beatriz Viçoso.

Finally, I would like to thank my family and friends, who supported me throughout my PhD.

# About the Author

Réka Kelemen completed a Bachelor of Science degree in Bioinformatics and Computational Biology at Iowa State University, USA and a Master of Science degree in Genome Science and Technology at the University of Tennessee, Knoxville, USA. She joined ISTA in March 2016, where she worked in Beatriz Vicoso's group as a scientific intern for six months, and then as a PhD student on the research project "Characterizing the sequence and expression evolution of the *t*-haplotype, a model meiotic driver". In her projects she collaborated with Anna Lindholm from the University of Zürich, Switzerland. Réka published her results in the journals Genetics and the Proceedings of the Royal Society Biological Sciences. During her PhD studies she presented her work at the ESEB (European Society for Evolutionary Biology) conference in Groningen, the Netherlands, in 2017, and at the Evolution conference in Montpellier, France, in 2018. She contributed to science outreach projects, such as STEB (Selected Topics in Evolutionary Biology) by writing an article for and presenting to high school students about the evolution of sexual reproduction, as well as by hosting them at ISTA and giving them a tour of the Vicoso laboratory.

# List of Collaborators and Publications

### Collaborators

- Beatriz Viçoso (ISTA) supervised all three studies presented in this thesis, and contributed to the analysis part of the study in Chapter 2 and to the writing of Chapters 2 and 3.
- Marwan Elkrewi (ISTA) contributed to the analysis of the study in Chapter 3.
- Anna K. Lindholm (University of Zürich, Switzerland) contributed to the study in Chapter 3 by conducting experiments and providing comments to the manuscript. She provided mouse samples, which served as the basis for the analyses presented in Chapter 4.

### Publications

- R. K. Kelemen and B. Vicoso. Complex history and differentiation patterns of the t-haplotype, a mouse meiotic driver. Genetics, 208(1):365375, 2018.
- R. K. Kelemen, M. Elkrewi, A. K. Lindholm, and B. Vicoso. Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. Proceedings of the Royal Society B, 289(1968):20211985, 2022.
- R. K. Kelemen, A. K. Lindholm, and B. Vicoso. Single nucleus sequencing uncovers candidate poisons and antidotes in testes carrying the t-haplotype, a model meiotic driver. Manuscript in preparation.

# Table of Contents

CHAPTER 1

# Introduction

## 1.1 Meiotic drivers and their population dynamics

Sexually reproducing organisms inherit two genomes from their parents. Through meiosis, the two genomes exchange genetic material, and separate into the gametes of the organism. Some of the gametes will fuse with gametes of another individual, and will pass on their genetic material to the next generation. A variant (allele) at a certain genomic position (locus) will increase in frequency in a population, if it is on average passed on to more progeny than other alleles at that locus [28]. An allele can achieve this by making the organism leave more offspring than other alleles do or by getting into more than the mendelian expectation, 50%, of the offspring, when in heterozygotes (i.e. individuals carrying different alleles). In the first scenario the allele's effect on organismal fitness is naturally selected, and results in a better adapted organism [23]. In the second scenario, natural selection acts on the ability of the allele to bias transmission in heterozygotes, which does not increase organismal fitness, but results in the spread of self-promoting alleles, often called meiotic drivers [67, 60].

Meiotic drivers are maintained in genomes even though they might not be beneficial to the organism – a concept that was first formalized in 1945 by the Swedish scientist Gunnar Östergren [60]. Östergren argued that since B chromosomes, which are additional, non-essential chromosomes in many plant species, promote their own transmission to the gametes, "selection would accumulate such fragments [...] even if they were quite neutral in effect on the plant or even unfavourable" [60]. Later, mathematical modeling showed that meiotic drivers spread in the population as long as their transmission gain outweighs any fitness cost that they impose on their hosts [7, 44, 65, 67]. In order to understand the evolutionary dynamics of such self-promoting alleles and their consequences for the genomes, organisms and populations in which they are present, let us start by considering the expected population dynamics of a newly arisen meiotic driver.

An allele that evolved to promote its own transmission in heterozygotes at the expense of alternative alleles will, if not lost by chance, get more frequent in the population with each generation, until it completely replaces alternative alleles [7, 44, 65, 67]. As the driver sweeps to fixation, the changes in meiosis or gametogenesis that it relied on to bias transmission will be fixed too, possibly explaining the fast turn-over of the genes and genomic regions involved in these processes [39, 72, 74]. The transient nature of meiotic drive also means that it is unlikely to be detected, unless fixed drivers are paired with an alternative allele from another population. Such cryptic drivers have been detected in various yeast, fruit fly, feline and mouse species [57, 61, 59, 17].

## 1.2 Homologous suppressors, enchancers and hybrid incompatibilities

The fast spread of meiotic drivers can be counteracted by certain forces. Populations that inbreed, self-fertilize or reproduce asexually have fewer heterozygotes, and therefore less opportunities for a meiotic driver to bias transmission ratios [41, 1]. In fungi, meiotic drivers tend to be observed in more inbred species [71], in line with the expectation that inbreeding slows their spread to fixation, and gives more time for their discovery [46]. Another way to counteract drive is by evolving a suppressor that acts against its molecular mechanism. For instance, if the allele that is disadvantaged by meiotic drive mutates to be less sensitive to it, it will be transmitted more often than the original sensitive allele, and will outcompete it [35]. Meiotic drivers present on a sex chromosome represent a special case, as heterozygotes bias

the sex ratio of their offspring [48]. As such a driver approaches fixation, the disadvantaged allele is associated with the ever rarer sex, whose individuals have a higher chance to contribute to the next generation than those of the common sex [27]. The increased reproductive success of the carriers of the disadvantaged allele slows the spread of the driver and provides more time for the evolution of suppressors [35, 11, 33].

On the other hand, an enhancer mutation of the driving allele will also spread faster than the original driver, creating an arms race that can lead to fast evolution of both the driver and target alleles [48, 33, 56, 38], as well as of a broaded "battlefield" that includes the biological components of the pathway whose manipulation results in better transmission. Seemingly balanced transmission ratios may be hiding highly altered biological pathways where co-evolved drivers, suppressors and enhancers keep each other in check, as seen for the co-adapted X and Y chromosomes in fruit flies and mice [13, 49]. When a genome containing co-evolved drivers and suppressors is paired with a genome from another population that did not co-evolve with it, the resulting hybrid individual will have incompatibilities within the altered biological pathways. Evidence for this is found in many species, where hybrid sterility loci overlap with cryptic drivers, often on sex chromosomes [78, 59, 17, 64].

## 1.3 Fitness costs and its effects on the rest of the genome

A third force counteracting the fixation of drivers is a fitness cost imposed on the host, which is a common feature of drivers [34, 44]. The cost can be inherent to the mechanism of drive, for example in carriers of male meiotic drivers that often lose half of their gametes, potentially decreasing their reproductive success [77]. Such reduction should be especially pronounced in multiple matings, as in this case gametes from multiple individuals compete for fertilization [77]. The mechanism of drive is often imperfect, and harms also the gametes carrying the driver, further reducing fertility [36, 54]. Alternatively, fitness costs can result from the degeneration of genes in the drive-associated region. Drivers are often multi-locus haplotypes, as drive is achieved by a two-component molecular system, which will be detailed below. It is essential for the success of the driver to keep its alleles of both components on one haplotype, therefore meiotic drivers are often found in regions of suppressed recombination, such as in inversions [25]. As enhancers of drive evolve in neighboring chromosomal regions, they might select for further inversions, further expanding the driving haplotype [52]. Hundreds of genes that are not involved in drive can be trapped on large meiotic drivers, which are expected to degenerate, as in the absence of recombination there will be few opportunities for removal of deleterious mutations. Deleterious alleles may in fact hitch-hike along successful new mutations that enhance drive to become fixed on the driver haplotype [25].

When carriers of a meiotic driver leave on average less offspring than non-carriers due to a driver-associated fitness cost, then any allele in the rest of the genome will spread better in the population if it is not found in a carrier of the meiotic driver. Any allele in the genome that can suppress drive will co-segregate less often with it, and will have higher fitness in the next generations [15]. Indeed, suppressors of drive are often found outside of the driving locus [42, 70]. Other molecular and phenotypic changes can mitigate the drive-associated cost, although this might be harder to detect. For example, multiple matings might be prevented by individuals carrying a meiotic driver through increased territory- or mate-guarding. So far there has not been evidence for this in mice carrying the selfish *t*-haplotype [8], which is highly disadvantaged in polyandrous matings [54]. On a molecular level, degenerated genes might be

compensated for by increased expression from the non-driving chromosome, similarly to dosage compensation in the case of degenerated sex chromosomes [22], although, again, this has not been detected yet in the case of drivers. Another mechanism that reduces the genetic load of a driver is genetic exchange with the standard homolog. Occasional recombination through a double-cross-over inside large inversions might be sufficient to purge the accumulated genetic load, as has been seen for the selfish X chromosome in *Drosophila neotestacea* [62].

## 1.4 Female meiotic drive

The term meiotic drive is used broadly and includes distortion at different points of the transmission process, including meiosis, gametogenesis and even embryogenesis [52], the first two of which will be discussed here. The process of meiosis, where homologous chromosomes separate into the daughter cells, is asymmetrical in females, and can therefore become biased [12]. At the interphase of meiosis the two chromosomes replicate their DNA, and the resulting four chromatids will end up in either the single egg, or in other non-transmissive cells, called the polar bodies. Female meiotic drivers thus typically target the mechanism of spindle attachment to preferentially orient themselves towards the pole that will give rise to the egg [12]. This is often achieved through increases in DNA repeat content in and around centromeres, where spindle attachment is mediated through structures called kinetochores [47, 29]. The surprisingly low conservation of centromeric sequences and kinetochore proteins, and the fast turn-over rate of chromosome karyotypes are thought to be a result of the arms race over preferential segregation into the egg [39, 53, 20].

In some cases, repeat expansions in other parts of the chromosomes can cause female meiotic drive, as seen in maize and mice [66, 21, 2, 76]. The well-studied maize driver, *Ab10*, is an abnormal version of chromosome 10, consisting mostly of two highly repeated motifs that are so enriched for heterochromatin that they appear as two large knobs under the microscope [18]. Instead of associating with kinetochores, ten copies of a novel gene, *Kindr*, which are adjacent to the knob express kinesin proteins that speed up the transport of *Ab10* along the spindle towards the future egg [19]. *Ab10* also increases recombination throughout the genome, possibly to ensure recombination between the centromere and the knob, which spreads the *Ab10* chromatids spatially during meiosis and increases the chances of drive. Since *Kindr* is essential for drive, recombination between Kindr and the repeat arrays would destroy the driving haplotype. However, in the regions of *Ab10* that show homology with the normal chromosome 10 went through a chromosomal inversion, suppressing recombination. Similar repeat arrays have evolved on other chromosomes of maize, which also show meiotic drive, but only when the individual carries the kinesin-expressing *Ab10*. Knob repeats can increase genome size by 20% (500 million basepairs). Despite a drive of 83% in heterozygotes, *Ab10* is found only in about 5-16% of sampled individuals, likely due to significantly reduced male fertility, seed size and number in plants homozygous for *Ab10* [34]. The genetic basis of the fitness cost is unclear, but may be due to deteriorated genes on *Ab10* [34]. On some naturally occuring Ab10 haplotypes with reduced drive, Kindr is epigenetically silenced, possibly due to suppressors of drive. A candidate suppressor is a cluster of degenerated *Kindr* genes on normal chromosome 10, producing high levels of microRNAs [19].

## 1.5 Male meiotic drive and spore killers

In contrast to female meiosis, a male or fungal meiotic cell produces four gametes, each of which has the potential to contribute to the next generation. In such a symmetric process of gametogenesis an allele can distort transmission ratios by interfering with the function or survival of gametes that do not carry it. Differential survival or function is dependent on the differential presence of a factor called responder in the gametes, and the universal presence of a factor called the distorter (or driver) [50]. Several examples of such distorter/responder systems have been described. In the fungus *Podospora anserina* haploid strains mate and create diploid meiotic cells, which inherit their cytoplasm almost exclusively from the female haploid strain. If the female haploid strain carries the spore killer Het-s, the HET-s protein will be passed on in the cytoplasm to the meiotic cell, and all its daughter cells [16]. The het-s locus is transcriptionally silenced until post-meiosis, creating the differential presence of alternative alleles in the spores. In heterozygous matings, spores that inherit a Het-S allele express the HET-S protein intracellularly, which interact with the maternally inherited HET-s protein and destabilize the plasma membrane, ultimately aborting the spore [68]. In this system, as well as in another spore killer of yeast, known as *wtf4*, different forms of the same gene are transcriptionally regulated so that the distorter is expressed pre-meiotically, whereas the responder is expressed most-meiotically to create the differential fates of spores carrying the driver versus spores not carrying it [57].

A conserved feature of animal spermatogenesis is that cells produced from a single progenitor cell remain connected by intercellular bridges [31]. This facilitates the distribution of the distorter products, which can be expressed pre-meiotically, or post-meiotically and shared through intercellular bridges. On the other hand, the cytoplasmic connections of post-meiotic cells safeguard against molecular differences between them, so that responders are limited to DNA sequences or epigenetic marks, or RNA or protein that are expressed post-meiotically and evade sharing. The sequence divergence between many sex chromosomes creates a large target size for a distorter to act on, and X- or Y-linked drivers are easy to detect through their shift in the progeny sex ratio. Consequently, many known drivers acting in spermatogenesis are sex-linked. For instance, in the fruit fly *D. simulans*, a driver on the X chromosome (the "Paris driver", named after its place of discovery) contains a meiotically expressed novel gene duplicate that binds the heterochromatin of the very diverged Y chromosome. In combination with further unidentified distorters, this leads to segregational failures and arrested development of Y-bearing haploid cells. The Paris driver likely arose in East-African fly populations [30], where it swept to fixation along with resistant Y chromosomes that restore normal sex ratios. In the last decade the driving X chromosome has spread into North-African populations, which in turn caused the native resistant Y chromosomes to be replaced by resistant versions within a matter of years [37]. This documented process shows how fast a powerful driver can introgress and repeatedly sweep through populations, erasing variation on the driving and non-driving homolog as well [14].

Much fewer autosomal drivers have been studied in detail. One of them is *Segregation Distorter (SD)*, a sperm killer found in *Drosophila melanogaster*. The target of the *SD* driver is a 240-basepair repeat near the centromere of the autosomal chromosome 2 [51]. The distorter is a truncated duplicate of a Ran GTPase activating protein also found on this chromosome, that along with an unidentified enhancer prevents gametes with a high copy number of the 240-basepair repeat from undergoing chromatin compaction (an essential step for sperm maturation). The distorter itself is linked to a low-copy-number responder locus, which is resistant to the harmful effects of the distorter, leading to its 95% transmission in

heterozygotes. Chromosomal inversions around the driver region of chromosome 2, as well as the proximity to the centromere suppress recombination with the non-driving chromosome. Despite its strong transmission ratio distortion, *SD* is found consistently at 1-5% frequency worldwide [43, 63], probably due to several counteracting forces. *SD* chromosomes are homozygous sterile, and frequently lethal, while *SD* heterozygotes have reduced fertility [36]. Some non-driving chromosomes contain insensitive responder alleles, while several unlinked suppressors have been found to segregate in the genome of *Drosophila melanogaster*.

## 1.6 The *t*-haplotype, a model meiotic driver

The only known mammalian sperm killer is the *t*-haplotype on the autosomal chromosome 17 of house mice. Its mechanism of action relies on a protein, SMOK (for sperm motility kinase), which is unshared through the cytoplasmic bridges, causing the differential fates of the gametes [73]. Sperm that does not carry the insensitive responder, $Smok^{Tcr}$, is severely impaired in its motility [40, 58, 69], and is transmitted to less than 5% of a heterozygote's offspring. The insensitive responder does not completely restore motility, and all sperm from a *t*-carrier is inferior in motility to sperm from a $+/+$ male, disadvantageing the *t*-haplotype in polyandrous matings [54, 75]. Four distorters have been identified to date [10, 5, 4, 6], but the pathways leading to the drive phenotypes are not fully understood. The *t*-haplotype spans a 40 megabasepair region that contains hundreds of genes on the non-driving homolog (a genomic region called the *t* complex). Since there is no reference sequence for the *t*-haplotype not much is known about the evolutionary fate of ancestral *t* complex genes captured by the *t*-haplotype, or about novel genes that may have arisen on the driver. Four large inversions suppress recombination with the non-driving chromosome 17, making it challenging to narrow down the regions involved in drive using genetic mapping.

While selection for enhancers of drive may solely be responsible for the evolution of such a large driving haplotype that contains so many distorters, the possible presence of suppressors of drive is expected to be an additional selective pressure for recruiting more genes into drive. However, suppressors of drive have not been identified yet, despite the fact that the strength of drive is dependent on the homologous chromosome 17 and the genetic background [32].

Degeneration is also expected to have affected genes over their long history on this non-recombining haplotype, as the *t*-haplotype was shown to pre-date the house mouse subspecies split about 0.5 million years ago [55]. Carroll and colleagues found significant fitness costs of $+/t$ mice compared to $+/+$ mice, such as decreased reproductive output and higher mortality in both sexes and additionally reduced weight in females [8], supporting the accumulation of deleterious mutations. Further, most *t*-haplotypes carry a recessive lethal mutation that eliminates homozygous $t/t$ embryos. Homozygous $t/t$ mice can be made by crossing *t* variants that complement each other's recessive lethals, but homozygous $t/t$ males are invariably sterile. Although embryonic lethals might arise due to degeneration, Charlesworth hypothesized that the lethals evolved to avoid maternal resource investment into sterile sons, and to channel the resources to $+/t$ embryos instead [9].

The *t*-haplotype exists in all three subspecies of house mice usually at a frequency of about 5-15% [3, 45, 24]. Although very diverged from the standard *t* complex, *t*-haplotypes from different parts of the world were found to be very similar, suggestive of a recent sweep of one haplotype worldwide [55]. However, this study sampled variation in a 610-basepair intronic sequence, possibly underrepresenting the variation on other parts of this large driver. Other studies, similarly focusing on a handful of genes, have found genetic exchange with the

non-driving chromosome, mostly in the large fourth, most distal inversion of the *t*-haplotype [26].

The advent of next generation sequencing brought new, more comprehensive toolkits to study genomes and their selfish elements. Sequencing the genomes and transcriptomes of drivers has the potential to uncover features that have been hidden inside the non-recombining, heterochromatic nature of drivers for the last century of their study. This thesis presents three studies of the *t*-haplotype in house mice that use genomic and transcriptomic data to answer open questions about this model meiotic driver. The first study revisits the existing hypotheses about the history and evolution of the *t*-haplotype: whether variant information from the entire *t*-haplotype supports high divergence from the standard chromosome 17, and low diversity among *t*-haplotypes, and in general, how much genetic divergence can be detected. The second study investigates the sequence and expression evolution of ancestral and novel genes on the *t*-haplotype. The third study reconstructs the expression landscape during drive using single nucleus RNA sequencing, a method that allows cell-type-, and even cell-genotype-resolved analysis of expression.

## 1.7 Bibliography


[1] J. A. Ågren and A. G. Clark. Selfish genetic elements. *PLoS genetics*, 14(11):e1007700, 2018.

[2] S. I. Agulnik, A. I. Agulnik, and A. O. Ruvinsky. Meiotic drive in female mice heterozygous for the HSR inserts on chromosome 1. *Genetics Research*, 55(2):97–100, 1990.

[3] K. G. Ardlie and L. M. Silver. Low frequency of t haplotypes in nautral populations of house mice (Mus musculus domesticus). *Evolution*, 52(4):1185–1196, Aug 1998.

[4] H. Bauer, S. Schindler, Y. Charron, J. Willert, B. Kusecek, and B. G. Herrmann. The nucleoside diphosphate kinase gene Nme3 acts as quantitative trait locus promoting non-mendelian inheritance. *PLoS genetics*, 8(3):e1002567, 2012.

[5] H. Bauer, N. Véron, J. Willert, and B. G. Herrmann. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & development*, 21(2):143–147, 2007.

[6] H. Bauer, J. Willert, B. Koschorz, and B. G. Herrmann. The t complex–encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature genetics*, 37(9):969–973, 2005.

[7] D. Bruck. Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Proceedings of the National Academy of Sciences*, 43(1):152–158, 1957.

[8] L. S. Carroll, S. Meagher, L. Morrison, D. J. Penn, and W. K. Potts. Fitness effects of a selfish gene (the Mus t complex) are revealed in an ecological context. *Evolution*, 58(6):1318–1328, 2004.

[9] B. Charlesworth. The evolution of lethals in the t-haplotype system of the mouse. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 258(1352):101–107, 1994.

[10] Y. Charron, J. Willert, B. Lipkowitz, B. Kusecek, B. G. Herrmann, and H. Bauer. Two isoforms of the RAC-specific guanine nucleotide exchange factor TIAM2 act oppositely on transmission ratio distortion by the mouse t-haplotype. *PLoS Genetics*, 15(2):e1007964, 2019.

[11] A. G. Clark. Natural selection and Y-linked polymorphism. *Genetics*, 115(3):569–577, 1987.

[12] F. E. Clark and T. Akera. Unravelling the mystery of female meiotic drive: where we are. *Open Biology*, 11(9):210074, 2021.

[13] C. Courret, P. R. Gérard, D. Ogereau, M. Falque, L. Moreau, and C. Montchamp-Moreau. X-chromosome meiotic drive in Drosophila simulans: a QTL approach reveals the complex polygenic determinism of Paris drive suppression. *Heredity*, 122(6):906–915, 2019.

[14] C. Courret, D. Ogereau, C. Gilbert, A. M. Larracuente, and C. Montchamp-Moreau. The evolutionary history of Drosophila simulans Y chromosomes reveals molecular signatures of resistance to sex ratio meiotic drive. *Molecular Biology and Evolution*, 40(7):msad152, 2023.

[15] J. F. Crow. Why is Mendelian segregation so exact? *Bioessays*, 13(6):305–312, Jun 1991.

[16] H. J. Dalstra, K. Swart, A. J. Debets, S. J. Saupe, and R. F. Hoekstra. Sexual transmission of the Het-S prion leads to meiotic drive in Podospora anserina. *Proceedings of the National Academy of Sciences*, 100(11):6616–6621, 2003.

[17] B. W. Davis, C. M. Seabury, W. A. Brashear, G. Li, M. Roelke-Parker, and W. J. Murphy. Mechanisms underlying mammalian hybrid sterility in two feline interspecies models. *Molecular biology and evolution*, 32(10):2534–2546, 2015.

[18] R. K. Dawe. The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Research*, 30(2-3):205–216, 2022.

[19] R. K. Dawe, E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, D. M. Higgins, J. Ross-Ibarra, J. G. Wallace, L. B. Kanizay, M. Alabady, et al. A kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*, 173(4):839–850, 2018.

[20] F. P.-M. de Villena and C. Sapienza. Female meiosis drives karyotypic evolution in mammals. *Genetics*, 159(3):1179–1189, 2001.

[21] J. P. Didion, A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan, L. Ortiz de Solorzano, J. Britton-Davidian, C. J. Bult, K. J. Campbell, R. Castiglia, et al. R2d2 drives selfish sweeps in the house mouse. *Molecular biology and evolution*, 33(6):1381–1395, 2016.

[22] C. M. Disteche. Dosage compensation of the sex chromosomes. *Annual review of genetics*, 46:537–560, 2012.

[23] T. Dobzhansky. Genetic nature of species differences. *The American Naturalist*, 71(735):404–420, 1937.

[24] B. Dod, C. Litel, P. Makoundou, A. Orth, and P. Boursot. Identification and characterization of t haplotypes in wild mice populations using molecular markers. *Genet Res*, 81(2):103–114, Apr 2003.

[25] K. A. Dyer, B. Charlesworth, and J. Jaenike. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proceedings of the National Academy of Sciences*, 104(5):1587–1592, 2007.

[26] M. A. Erhart, S. Lekgothoane, J. Grenier, and J. H. Nadeau. Pattern of segmental recombination in the distal inversion of mouse t haplotypes. *Mamm Genome*, 13(8):438–444, Aug 2002.

[27] R. Fisher. The Natural Selection, 1930.

[28] R. A. Fisher. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1999.

[29] L. Fishman and A. Saunders. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science*, 322(5907):1559–1562, 2008.

[30] L. Fouvry, D. Ogereau, A. Berger, F. Gavory, and C. Montchamp-Moreau. Sequence analysis of the segmental duplication responsible for Paris sex-ratio drive in Drosophila simulans. *G3: Genes| Genomes| Genetics*, 1(5):401–410, 2011.

[31] M. P. Greenbaum, T. Iwamori, G. M. Buchold, and M. M. Matzuk. Germ cell intercellular bridges. *Cold Spring Harbor perspectives in biology*, 3(8):a005850, 2011.

[32] G. R. Gummere, P. J. McCormick, and D. Bennett. The influence of genetic background and the homologous chromosome 17 on t-haplotype transmission ratio distortion in mice. *Genetics*, 114(1):235–245, 1986.

[33] D. W. Hall. Meiotic drive and sex chromosome cycling. *Evolution*, 58(5):925–931, 2004.

[34] D. W. Hall and R. K. Dawe. Modeling the evolution of female meiotic drive in maize. *G3: Genes, Genomes, Genetics*, 8(1):123–130, 2018.

[35] W. D. Hamilton. Extraordinary sex ratios: A sex-ratio theory for sex linkage and inbreeding has new implications in cytogenetics and entomology. *Science*, 156(3774):477–488, 1967.

[36] D. L. Hartl, Y. Hiraizumi, and J. F. Crow. Evidence for sperm dysfunction as the mechanism of segregation distortion in Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 58(6):2240–2245, 1967.

[37] Q. Helleu, C. Courret, D. Ogereau, K. L. Burnham, N. Chaminade, M. Chakir, S. Aulard, and C. Montchamp-Moreau. Sex-ratio meiotic drive shapes the evolution of the Y chromosome in Drosophila simulans. *Molecular Biology and Evolution*, 36(12):2668–2681, 2019.

[38] Q. Helleu, P. R. Gérard, R. Dubruille, D. Ogereau, B. Prudhomme, B. Loppin, and C. Montchamp-Moreau. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy of Sciences*, 113(15):4110–4115, 2016.

[39] S. Henikoff, K. Ahmad, and H. S. Malik. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 293(5532):1098–1102, 2001.

[40] B. G. Herrmann, B. Koschorz, K. Wertz, K. J. McLaughlin, and A. Kispert. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature*, 402(6758):141–146, 1999.

[41] D. A. Hickey. Selfish DNA: a sexually-transmitted nuclear parasite. *Genetics*, 101(3-4):519–531, 1982.

[42] Y. K. Hihara. Genetic analysis of modifying system of Segregation Distortion in Drosophila melanogaster ii. two modifiers for SD system on the second chromosome of D. melanogaster. *The Japanese journal of genetics*, 49(4):209–222, 1974.

[43] Y. Hiraizumi and K. Nakazima. Deviant sex ratio associated with segregation distortion in Drosophila melanogaster. *Genetics*, 55(4):681, 1967.

[44] Y. Hiraizumi, L. Sandler, and J. F. Crow. Meiotic drive in natural populations of Drosophila melanogaster. iii. populational implications of the segregation-distorter locus. *Evolution*, pages 433–444, 1960.

[45] S. W. Huang, K. G. Ardlie, and H. T. Yu. Frequency and distribution of t-haplotypes in the Southeast Asian house mouse (Mus musculus castaneus) in Taiwan. *Mol Ecol*, 10(9):2349–2354, Sep 2001.

[46] L. D. Hurst. A century of bias in genetics and evolution. *Heredity*, 123(1):33–43, 2019.

[47] A. Iwata-Otsubo, J. M. Dawicki-McKenna, T. Akera, S. J. Falk, L. Chmátal, K. Yang, B. A. Sullivan, R. M. Schultz, M. A. Lampson, and B. E. Black. Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. *Current Biology*, 27(15):2365–2373, 2017.

[48] J. Jaenike. Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics*, 32(1):25–49, 2001.

[49] A. N. Kruger, M. A. Brogley, J. L. Huizinga, J. M. Kidd, D. G. de Rooij, Y.-C. Hu, and J. L. Mueller. A neofunctionalized X-linked ampliconic gene family is essential for male fertility and equal sex ratio in mice. *Current Biology*, 29(21):3699–3706, 2019.

[50] A. N. Kruger and J. L. Mueller. Mechanisms of meiotic drive in symmetric and asymmetric meiosis. *Cellular and Molecular Life Sciences*, 78:3205–3218, 2021.

[51] A. M. Larracuente and D. C. Presgraves. The selfish segregation distorter gene complex of Drosophila melanogaster. *Genetics*, 192(1):33–53, 2012.

[52] A. K. Lindholm, K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, L. Holman, H. Johannesson, U. Knief, H. Kokko, A. M. Larracuente, et al. The ecology and evolutionary dynamics of meiotic drive. *Trends in ecology & evolution*, 31(4):315–326, 2016.

[53] H. S. Malik, D. Vermaak, and S. Henikoff. Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. *Proceedings of the National Academy of Sciences*, 99(3):1449–1454, 2002.

[54] A. Manser, B. König, and A. K. Lindholm. Polyandry blocks gene drive in a wild house mouse population. *Nature communications*, 11(1):5590, 2020.

[55] T. Morita, H. Kubota, K. Murata, M. Nozaki, C. Delarbre, K. Willison, Y. Satta, M. Sakaizumi, N. Takahata, and G. Gachelin. Evolution of the mouse t haplotype: recent and worldwide introgression to Mus musculus. *Proceedings of the National Academy of Sciences*, 89(15):6851–6855, 1992.

[56] K. Nam, K. Munch, A. Hobolth, J. Y. Dutheil, K. R. Veeramah, A. E. Woerner, M. F. Hammer, G. A. G. D. Project, T. Mailund, and M. H. Schierup. Extreme selective sweeps independently targeted the X chromosomes of the great apes. *Proceedings of the National Academy of Sciences*, 112(20):6413–6418, 2015.

[57] N. L. Nuckolls, M. A. Bravo Núñez, M. T. Eickbush, J. M. Young, J. J. Lange, J. S. Yu, G. R. Smith, S. L. Jaspersen, H. S. Malik, and S. E. Zanders. Wtf genes are prolific dual poison-antidote meiotic drivers. *Elife*, 6:e26033, 2017.

[58] P. Olds-Clarke and L. R. Johnson. t haplotypes in the mouse compromise sperm flagellar function. *Developmental biology*, 155(1):14–25, 1993.

[59] H. A. Orr and S. Irving. Segregation distortion in hybrids between the Bogota and USA subspecies of Drosophila pseudoobscura. *Genetics*, 169(2):671–682, 2005.

[60] G. Ostergren. Parasitic nature of extra fragment chromosomes. *Bot. Not.*, 2:157–163, 1945.

[61] N. Phadnis and H. A. Orr. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *Science*, 323(5912):376–379, 2009.

[62] K. E. Pieper and K. A. Dyer. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology*, 29(11):2229–2241, 2016.

[63] D. C. Presgraves, P. R. Gérard, A. Cherukuri, and T. W. Lyttle. Large-scale selective sweep among segregation distorter chromosomes in African populations of Drosophila melanogaster. *PLoS genetics*, 5(5):e1000463, 2009.

[64] D. C. Presgraves and C. D. Meiklejohn. Hybrid sterility, genetic conflict and complex speciation: lessons from the Drosophila simulans clade species. *Frontiers in genetics*, 12:669045, 2021.

[65] T. Prout. Some effects of variations in the segregation ratio and of selection on the frequency of alleles under random mating. *Acta Genetica Et Statistica Medica*, 4(2-3):148–151, 1953.

[66] M. Rhoades. Preferential segregation in maize. *Genetics*, 27(4):395, 1942.

[67] L. Sandler and E. Novitski. Meiotic drive as an evolutionary force. *The American Naturalist*, 91(857):105–110, 1957.

[68] C. Seuring, J. Greenwald, C. Wasmer, R. Wepf, S. J. Saupe, B. H. Meier, and R. Riek. The mechanism of toxicity in HET-S/HET-s prion incompatibility. *PLoS biology*, 10(12):e1001451, 2012.

[69] A. Sutter and A. K. Lindholm. Meiotic drive changes sperm precedence patterns in house mice: potential for male alternative mating tactics? *BMC Evolutionary Biology*, 16:1–15, 2016.

[70] Y. Tao, L. Araripe, S. B. Kingan, Y. Ke, H. Xiao, and D. L. Hartl. A sex-ratio meiotic drive system in Drosophila simulans. ii: an X-linked distorter. *PLoS biology*, 5(11):e293, 2007.

[71] M. van der Gaag, A. J. Debets, J. Oosterhof, M. Slakhorst, J. A. Thijssen, and R. F. Hoekstra. Spore-killing meiotic drive factors in a natural population of the fungus Podospora anserina. *Genetics*, 156(2):593–605, 2000.

[72] J. Vedanayagam, C.-J. Lin, and E. C. Lai. Rapid evolutionary dynamics of an expanding family of meiotic drive factors and their hpRNA suppressors. *Nature ecology & evolution*, 5(12):1613–1623, 2021.

[73] N. Véron, H. Bauer, A. Y. Weiße, G. Lüder, M. Werber, and B. G. Herrmann. Retention of gene products in syncytial spermatids promotes non-mendelian inheritance as revealed by the t complex responder. *Genes & development*, 23(23):2705–2710, 2009.

[74] M. J. D. White. Models of speciation: New concepts suggest that the classical sympatric and allopatric models are not the only alternatives. *Science*, 159(3819):1065–1070, 1968.

[75] L. Winkler and A. K. Lindholm. A meiotic driver alters sperm form and function in house mice: a possible example of spite. *Chromosome Research*, 30(2):151–164, 2022.

[76] G. Wu, L. Hao, Z. Han, S. Gao, K. E. Latham, F. P.-M. de Villena, and C. Sapienza. Maternal transmission ratio distortion at the mouse Om locus results from meiotic drive at the second meiotic division. *Genetics*, 170(1):327–334, 2005.

[77] S. E. Zanders and R. L. Unckless. Fertility costs of meiotic drivers. *Current Biology*, 29(11):R512–R520, 2019.

[78] L. Zhang, T. Sun, F. Woldesellassie, H. Xiao, and Y. Tao. Sex ratio meiotic drive as a plausible evolutionary mechanism for hybrid male sterility. *PLoS genetics*, 11(3):e1005073, 2015.

CHAPTER 2

# Complex history and differentiation patterns of the *t*-haplotype, a mouse meiotic driver

Réka K. Kelemen and Beatriz Viçoso

*Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

# Complex History and Differentiation Patterns of the *t*-Haplotype, a Mouse Meiotic Driver

**Reka K. Kelemen and Beatriz Vicoso**[1]
Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria



**ABSTRACT** The *t*-haplotype, a mouse meiotic driver found on chromosome 17, has been a model for autosomal segregation distortion for close to a century, but several questions remain regarding its biology and evolutionary history. A recently published set of population genomics resources for wild mice includes several individuals heterozygous for the *t*-haplotype, which we use to characterize this selfish element at the genomic and transcriptomic level. Our results show that large sections of the *t*-haplotype have been replaced by standard homologous sequences, possibly due to occasional events of recombination, and that this complicates the inference of its history. As expected for a long genomic segment of very low recombination, the *t*-haplotype carries an excess of fixed nonsynonymous mutations compared to the standard chromosome. This excess is stronger for regions that have not undergone recent recombination, suggesting that occasional gene flow between the *t* and the standard chromosome may provide a mechanism to regenerate coding sequences that have accumulated deleterious mutations. Finally, we find that *t*-complex genes with altered expression largely overlap with deleted or amplified regions, and that carrying a *t*-haplotype alters the testis expression of genes outside of the *t*-complex, providing new leads into the pathways involved in the biology of this segregation distorter.

**KEYWORDS** meiotic driver; t-haplotype; genome evolution

MEIOTIC drivers (also known as segregation distorters) are selfish alleles or chromosome variants that can transmit themselves to over 50% of the progeny of heterozygous individuals (Burt and Trivers 2009; Lindholm *et al.* 2016), often by killing or inactivating gametes that carry the nondriver allele. This requires the combined action of at least one distorter gene, which attacks gametes, and a responder gene, which protects gametes carrying the driver [reviewed in Lindholm *et al.* (2016)]. Linkage between the distorter and responder genes is required for the survival of the driver, and successful drivers often arise in regions of low recombination (Schwander *et al.* 2014). Conversely, the presence of drivers can select for reduced recombination around the driving and responding loci (Charlesworth and Hartl 1978). Autosomal drivers usually have no detectable phenotypic effects, and much of what is known about them comes primarily from studies of two model systems: Segregation Distorter in *Drosophila melanogaster* (Larracuente and Presgraves 2012) and the *t*-haplotype of the domestic mouse *Mus musculus*.

The *t*-haplotype is a 40-Mb variant of the proximal portion of chromosome 17 (Burt and Trivers 2009; Herrmann and Bauer 2012), which shows suppressed recombination with the standard chromosome due to the accumulation of several inversions (three on the *t*-haplotype and one on the standard chromosome; Artzt *et al.* 1982; Herrmann *et al.* 1986; Burt and Trivers 2009).When present in females, it is transmitted to 50% of the progeny, but $>$ 90% of the progeny of *t*-carrying males inherit it (Chesley and Dunn 1936; Herrmann and Bauer 2012). Despite this strong driving capacity, *t*-haplotypes remain at relatively low frequency (10–25%; Ardlie 1998), partly because individuals carrying two copies of the *t*-haplotype have strongly reduced fertility and viability (Herrmann and Bauer 2012). *t*-haplotypes are found throughout the *M. musculus* species complex (which includes *M. m. domesticus*, *M. m. musculus*, and *M. m. castaneus*), but not in the close outgroup *M. spretus* (Lyon 2003).

The genetics of transmission distortion of the *t*-haplotype are well understood, and several drivers, as well as one responder, have been identified. These lead to morphological


doi: https://doi.org/10.1534/genetics.117.300513
Manuscript received January 2, 2017; accepted for publication November 7, 2017; published Early Online November 14, 2017.
Available freely online through the author-supported open access option.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.117.300513/-/DC1.
[1]Corresponding author: Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria. E-mail: bvicoso@ist.ac.at

defects in spermatozoa that do not carry a *t*-haplotype due to excessive activation of the chromosome 17 gene *Smok* (Bauer *et al.* 2007, 2012; Herrmann and Bauer 2012). By contrast, with one exception (Sugimoto 2014), the loci responsible for the lethality and sterility of homozygous *t*-haplotypes have not been mapped to specific genes. It is further unclear if these are caused by single loci in each *t*-haplotype, or by the accumulation of many deleterious mutations [but see Howell *et al.* (2004) for at least one example of a cryptic lethal mutation]. Selection is ineffective in regions of low recombination and genes located in such regions often accumulate deleterious mutations (Woolfit 2009; Campos *et al.* 2014). Nonrecombining segregation distorters should be particularly affected (Dyer *et al.* 2007), as mutations that arise there can spread if their harmful effect does not outweigh the selective advantage of the linked driver, and new mutations that increase driving efficiency can sweep linked deleterious variants to fixation (Presgraves *et al.* 2009). The extent to which the hundreds of genes on the *t*-haplotype have deteriorated, and whether occasional recombination with the standard chromosome is sufficient to maintain genetic integrity in meiotic drivers over millions of years (Dyer *et al.* 2007; Pieper and Dyer 2016), remain open questions.

Several questions also remain regarding the origin and sequence evolution of this meiotic driver. Sequence divergence between the *t*-haplotype and the standard chromosome 17 led to the conclusion that the first inversion arose over 3 MYA, and inversion 4 ~1.5 MYA (Hammer and Silver 1993). While these are likely overestimates given the current *M. spretus/M. musculus* estimates of divergence (Harr *et al.* 2016), they clearly precede the origin of all the *M. musculus* subspecies in which they are found (White *et al.* 2009), showing that they were present in the ancestral population. However, the *t*-haplotype sequences of the different subspecies show very little differentiation between them, suggesting that a single *t*-haplotype introgressed $< 0.1$ MYA throughout the species group (Morita *et al.* 1992; Hammer and Silver 1993). Much of this early work relied on short sequences and it is unclear if these patterns capture the full history of this driver; further, where this haplotype introgressed from is still unknown. Inversions 3 and 4 have been found to carry more genetic variants than inversion 2, with occasional recombination between different *t*-haplotypes (Dod *et al.* 2003), but also with standard chromosomes (Herrmann *et al.* 1987; Erhart *et al.* 1989, 2002; Hammer *et al.* 1991; Wallace and Erhart 2008) likely playing a role in their differentiation. How this varies throughout each inversion is unclear, something that is potentially problematic, as regions closer to breakpoints generally show a stronger reduction in recombination than the middle of inversions (Wallace and Erhart 2008). It has therefore not been excluded that differences between inversions could represent a sampling bias rather than a real difference in their age, or that estimates of the age of inversions have been biased by secondary recombination events.

Here, we take advantage of a recently published population genomics data set of wild mice (Harr *et al.* 2016), which contains RNA-sequencing (RNA-seq) and genomic data derived from 15 *M. musculus t*-haplotype carriers, 32 noncarriers from the same populations and eight individuals of the closely related species *M. spretus*, to characterize the *t*-haplotype at both the genomic and gene expression level.

## Materials and Methods

### Data source

Harr *et al.* (2016) recently published extensive population genomics resources for three subspecies of *M. musculus*, as well as its close outgroup *M. spretus*. These included 15 individuals heterozygous for *t*-haplotypes [four in *M. m. domesticus*, eight in *M. m. musculus* (excluding mouse CR29 that we suspect to be a partial *t*-haplotype-carrier), and three in *M. m. castaneus*], as well as many noncarriers [see Table 1 of Harr *et al.* (2016)]. For each individual, we downloaded a BAM alignment file with reads mapped to the house mouse reference genome and the respective variant-containing variant call format (VCF) file from http://wwwuser.gwdg.de/~evolbio/evolgen/wildmouse/.

Harr *et al.* (2016) further generated RNA-seq reads for brain, liver, spleen, heart, thyroid, kidney, and testis of the same 16 *M. m. domesticus* specimens that were used for genomic sequencing. The RNA-seq reads were downloaded from the National Center for Biotechnology Information Short Reads Archive (bioproject PRJEB11897).

A detailed protocol of all the steps involved and code used in our analysis is provided in Supplemental Material, File S1, while supplementary figures, tables, and data are provided in File S2.

### Copy number variant (CNV) detection

To avoid biases caused by polymorphic or *t*-specific CNVs, we called CNVs using the software Control-FREEC (Boeva *et al.* 2012) and combined these with the list of CNVs that Harr *et al.* (2016) obtained using the software CNVnator; both methods rely on differences in genomic coverage to detect deletions or duplications. We first ran Control-FREEC on each of the 55 BAM files against the reference genome, using window sizes of 1 and 5 kb. To fully detect *t*-specific CNVs, we then used Control-FREEC to call CNVs between the pooled *t*-carrier mice of each subspecies and four randomly chosen non-*t*-carrier mice controls from the same subspecies (see details in the supplemental methods described in File S1). A genomic region was classified as a CNV if it was detected in at least one sample by either of the two software packages.

### SNP filtering

We downloaded the two multisample VCF files provided by Harr *et al.* (2016). One contained the high-quality variants obtained by GATK's VSQR filtering, while the other contained the unfiltered raw SNPs. We conducted our entire analysis on both variant sets, and used BCFtools (Li 2011) to handle the VCF files.

We carried out our analysis using three different SNP-filtering procedures (detailed in File S1). In filtering procedure 1, we

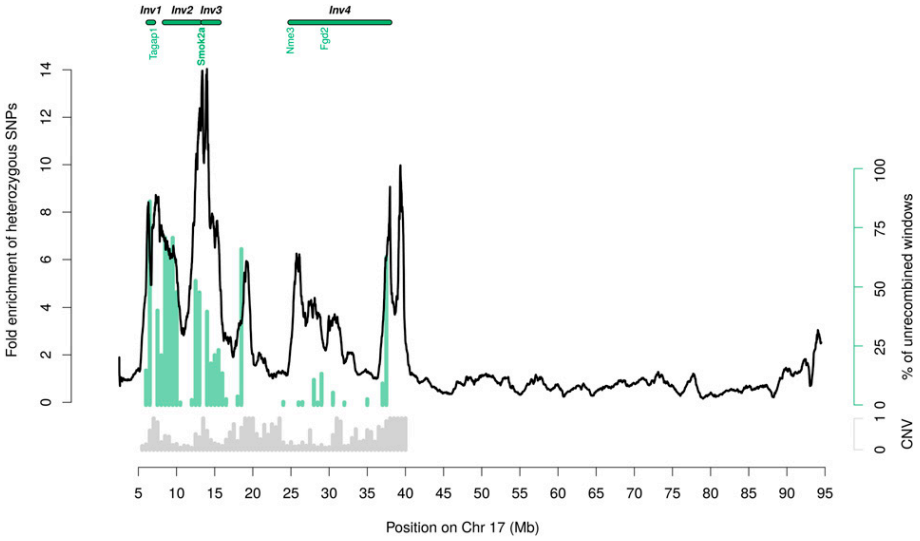



**Figure 1** Heterozygosity levels and phylogenetic topology along the *t*-haplotype of *M. m. domesticus*. The black line shows the heterozygous SNP density of *t*-carrier mice divided by the heterozygous SNP density of control noncarrier mice. The ratio is averaged over 1-Mb windows (sliding by 1-kb). Gray bars below show, for each 0.5-Mb segment of the *t*-complex, the proportion that was identified as a CNV in any of the 55 mice. Green bars indicate in each 0.5-Mb segment the proportion of trees that show all *M. m. domesticus* *t*-haplotypes outside of the *M. musculus* species cluster (see Figure 2 and *Materials and Methods*). We plotted the data for the entire chromosome 17 without masking CNVs. Chr, Chromosome; CNV, copy number variant; Inv, inversion.

used the variants classified as PASS by Harr *et al.* (2016), and removed sites that were not SNPs, such as indels (Figure 1 and Figure S3 and Figure S4 in File S2). We further deleted sites located within CNVs for the phylogenetic and deterioration analyses (Figure 2, Figure 3, Figure 4, and Figure S6, Figure S7, Figure S8, and Figure S9 in File S2).

The raw variants were similarly filtered for CNVs and non-SNP variants, as well as additional criteria:

Filtering procedure 2: a variant was kept only if its total coverage was at least half of the average coverage for the given sample (reported in Table 1 of Harr *et al.* 2016); this yielded Figure S1 in File S2.

Filtering procedure 3: we used the same coverage filtering as in procedure 2, and additionally required that each heterozygous allele be supported by at ≥ 30% of the reads (Figure S2 in File S2).

### *Estimates of heterozygosity in M. m. domesticus*

We extracted variants for each *M. m. domesticus* sample from the multisample VCF file and retained only heterozygous sites. We then computed the average heterozygous SNP density of *t*-carrier mice in 1000-bp regions averaged over sliding 1-Mb windows. We plotted this density curve divided by the average of the corresponding densities computed in all *M. m. domesticus* non-*t*-carrier mice.

### *Extracting "pseudo-t-haplotype" VCF files from heterozygous t-carriers*

Given that *t*-carrier mice are heterozygous for the *t*-haplotype, SNPs found in their VCF files could represent *t*-derived variants or SNPs from their standard chromosome. SNPs that were homozygous in each *t*-carrier mouse were kept for further analysis, as they were likely present on both chromosomes. At heterozygous sites, we discarded all SNPs that were found in at least one noncarrier individual of any of the *M. musculus* subspecies and retained all others as putative *t*-haplotype SNPs. One caveat of this subtraction step is that it excluded any polymorphism that was present on both a *t*-haplotype and a standard chromosome, if it happened to be heterozygous in the *t*-carrier; however, given the low recombination rates between the *t* and the standard chromosome, there should be few shared segregating variants between the standard and *t*-haplotypes, such that these should be a minority. Conversely, rare genetic variants on the standard chromosome of *t*-carriers may be wrongfully retained as *t*-specific if they are not present in any of the noncarriers (but given the high level of *t*-to-standard chromosome divergence relative to genetic diversity in noncarriers, as shown in Figure 1, these should once again represent only a small minority of SNPs).

### *Phylogenetic analysis*

We examined the phylogeny of the 15 *t*-haplotypes from the three different *M. musculus* subspecies, along with the noncarrier mice from each population. SNP profiles from eight individuals from a closely related species, *M. spretus*, served as the outgroup.

We used the pseudo-*t*-haplotype SNP profiles (see previous section) to represent the 15 *t*-haplotypes of the carrier mice. We converted all VCF files to FASTA files using the mouse reference background with the *consensus* function of BCFtools (Li 2011), and concatenated all sequences into a multisample FASTA file. We subsampled this FASTA file into the desired genomic regions using the *faidx* function of SAMtools (Li *et al.* 2009). To compute the maximum likelihood phylogenies of the 15 *t*-haplotypes, and the 40 noncarriers from *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus*, and *M. spretus*, we used the phylogenetic software IQTree (Nguyen *et al.* 2015) with an underlying Hasegawa–Kishono–Yano (HKY) model of DNA substitution. We assessed branch support values using an ultrafast bootstrap approximation UFBoot (Minh *et al.* 2013),

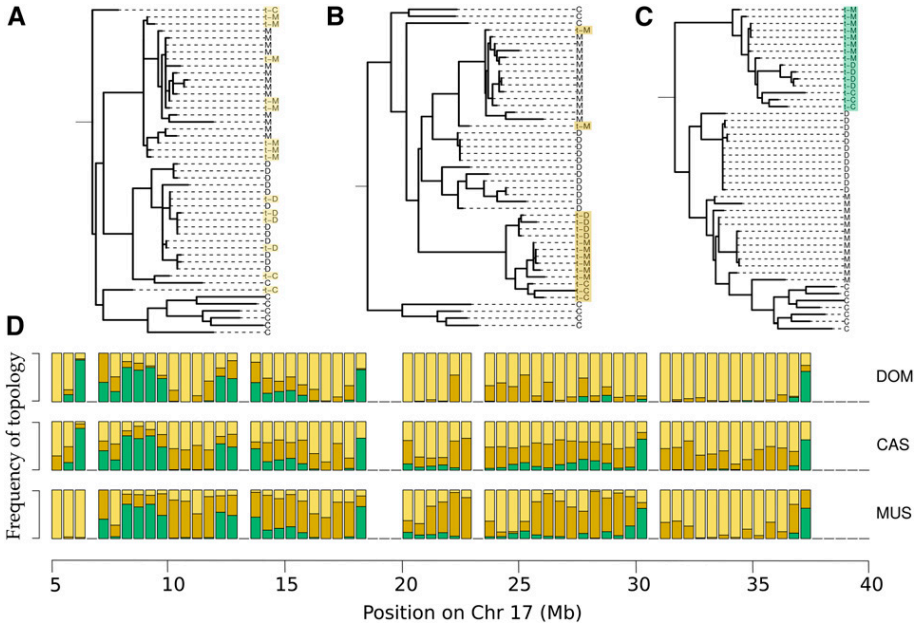



**Figure 2** Extent of recombination along the *t*-haplotype. (A–C) Example trees showing topologies that suggest very recent (A), recent (B), or no (C) recombination between the *t*-haplotype and the standard chromosome. In each tree, D, C, M, and S, respectively, represent noncarriers of *M. m domesticus*, *M. m. castaneus*, *M. m. musculus*, and *M. spretus*, while t-D, t-C, and t-M represent pseudo-*t*-haplotypes of *M. m domesticus*, *M. m. castaneus*, and *M. m. musculus*. Topology A is denoted with yellow, B with orange, and C with green. (D) The proportion of trees (obtained from nonoverlapping 5-kb windows for region 5–40 Mb of chromosome 17) with topologies A, B, and C for the *t*-haplotypes of each of the subspecies (DOM for *M. m. domesticus*, CAS for *M. m. castaneus*, and MUS for *M. m. musculus*). The proportion is shown for each 500-kb nonoverlapping region along the *t*-complex (5–40 Mb on chromosome 17).

which we iterated 1000 times. For computation of the maximum parsimony phylogenies, we used the software MEGA (Kumar *et al.* 2016) with default parameters but keeping only one tree, while for the neighbor-joining method we used FastPhylo with default parameters (Khan *et al.* 2013).

### *Nonsynonymous to synonymous ratio (NS/S) of SNPs at varying frequencies*

SNPs were classified as synonymous or nonsynonymous using the SNPeff software (Cingolani *et al.* 2012). We calculated the frequencies of each synonymous and nonsynonymous SNP in the four *M. m. domesticus* pseudo-*t*-haplotypes, in the 24 non-*t*-carrier *M. m. domesticus* chromosomes, and in the 16 *M. spretus* chromosomes using the genotypes provided in the VCF files. We counted a SNP once when it was found in a heterozygous individual, and twice when it was in a homozygous individual. In the case of the 24 non-*t*-carrier *M. m. domesticus* chromosomes, the four frequency classes were 1–6, 7–12, 13–18, and 19–24, while for the 16 *M. spretus* chromosomes, the frequency classes corresponded to 1–4, 5–8, 9–12, and 13–16. In the case of the four pseudo-*t*-haplotypes, the respective frequency classes were 1, 2, 3, and 4.

### *Gene expression analysis*

We obtained estimates of gene expression for each *M. m. domesticus* RNA-seq sample with Kallisto (Bray *et al.* 2016), using the *M. musculus* GRCm38.p4 coding sequence as reference. The resulting Kallisto transcript quantification was used as input for Sleuth, a software for differential expression analysis, using the gene aggregation feature (Pimentel *et al.* 2016).

### *Data availability*

All data analyzed in this study were previously published (Harr *et al.* 2016). The authors affirm that all analyses performed are fully described within the *Materials and Methods* and *Results* of the manuscript.

## Results

### *Variable levels of divergence along the t-haplotype*

We examined the extent of differentiation between the *t*-haplotype and the standard chromosome 17 of *M. m. domesticus*, for which there were four mice carrying the *t*-haplotype. While all the main figures are based on the high-quality SNPs provided by Harr *et al.* (2016), two alternative SNP-filtering procedures based on (1) coverage, and (2) coverage and allele frequency (see *Materials and Methods*), were used to check that results held independent of the filtering procedure (Figure S1 and Figure S2 in File S2).

We plotted the averaged SNP density of the four *t*-carrier mice along chromosome 17, divided by the respective average for noncarriers (for overlapping sliding windows of 1 Mb, Figure 1). Only heterozygous SNPs were used; in *t*-carriers these SNPs correspond to differences between the *t*-haplotype and the standard chromosome, whereas in control individuals they correspond to general levels of heterozygosity. Noncarrier individuals were used to control for variable genetic diversity rates along the chromosome (Figure S3 in File S2).

Figure 1 shows that *t*-carriers have increased heterozygosity in the region from 5 to 40 Mb of chromosome 17, consistent with the expected location of the *t*-haplotype. The excess of heterozygosity varies, with several regions showing a difference $>$ 10-fold. Many of these also overlap with CNVs identified through differences in coverage between individuals (shown in gray in Figure 1), but these CNVs represent a subset of the high-divergence regions, so divergence in duplicated regions does not account for the increase in heterozygosity. These results hold when only intergenic and

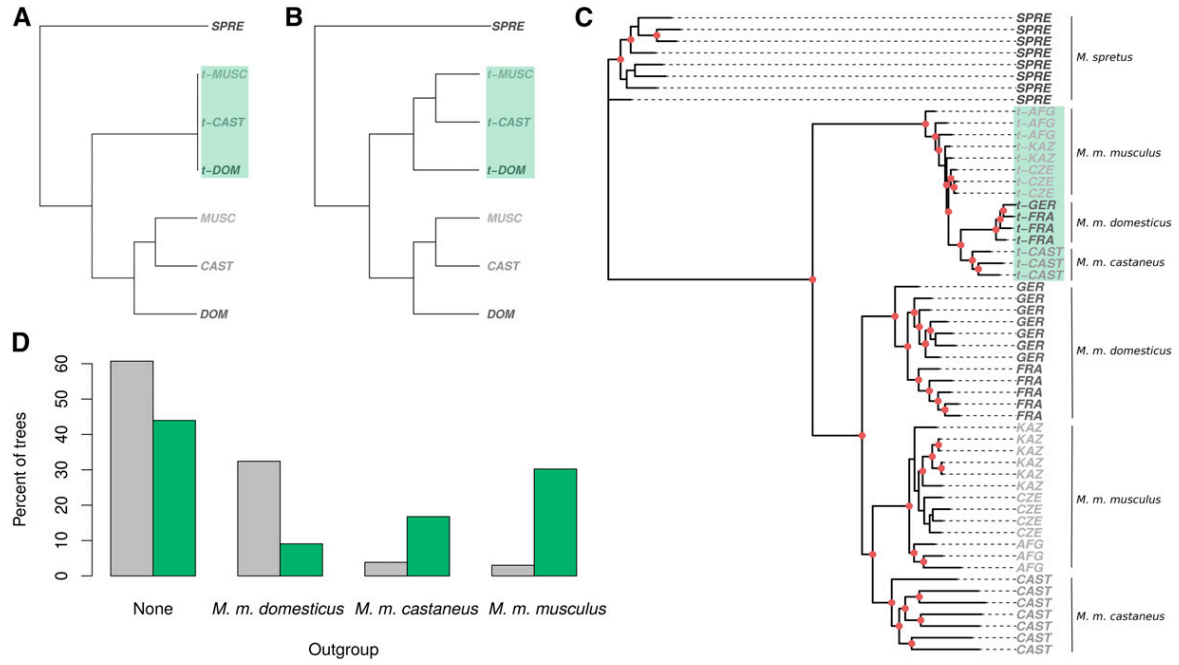



**Figure 3** Recent history of the *t*-haplotype. (A) Model phylogeny under the scenario of recent introgression of a single *t*-haplotype into all *M. musculus* subspecies. (B) Model phylogeny under the hypothesis of independent maintenance of ancestrally present *t*-haplotypes in the different subspecies. CAST, DOM, MUSC, and SPRET represent noncarriers of *M. m. castaneus*, *M. m. domesticus*, *M. m. musculus*, and *M. spretus*, respectively, while t-CAST, t-DOM, and t-MUSC represent *t*-haplotypes of *M. m. castaneus*, *M. m domesticus*, and *M. m. musculus*, respectively. (C) Phylogeny of pseudo-*t*-haplotypes and noncarrier mice from the three *M. musculus* subspecies and the sister species, *M. spretus*. Nodes with bootstrap values > 94% are marked with red dots. Only regions of the *t*-complex where no recombination between the standard chromosomes and the *t*-haplotype could be detected were included (green regions in Figure 2). Sequences starting with "t-" (highlighted with a green background) refer to *t*-haplotypes. AFG, CZE, and KAZ stand for *M. m. musculus* from Afghanistan, the Czech Republic, and Kazakhstan; GER and FRA for *M. m. domesticus* from Germany and France; CAST stands for *M. m. castaneus*; and SPRE for *M. spretus*. (D) Percentage of 5-kb windows without recombination for which the resulting phylogeny yields one subspecies as the outgroup to the others ("none" shows the proportion of windows for which no subspecies was an outgroup). Gray bars represent the phylogeny of non-*t*-carriers, and green bars represent the phylogeny of pseudo-*t*-haplotypes (see *Materials and Methods*).

synonymous SNPs are used (Figure S4B in File S2) and when we compare instead the pseudo-*t*-haplotype SNP density (see *Materials and Methods*) to the SNP density of *M. spretus* (to control for fast-diverging regions, Figure S4, C–F in File S2). We also get a consistent pattern of divergence along the *t*-haplotype when we reproduce Figure 1 using insertions and deletions (Figure S5 in File S2).

We indicated the putative location of the four nonoverlapping inversions (Braidotti and Barlow 1997; Harrison *et al.* 1998; Herrmann *et al.* 1999; Zwart *et al.* 2001; Bauer *et al.* 2007, 2012; Sugimoto 2014) (inversions 1–4 in Figure 1), as well as the position of the genes known to be involved in transmission distortion. As expected, the largest peak of divergence is at the distal end of the second inversion (based on the standard sequence orientation), near the *Smok2A* gene, which is in the vicinity of the previously identified responder gene *Tcr* (Herrmann *et al.* 1999; *Tcr* itself is not present on the standard chromosome). This region is assumed to have been ancestrally recruited to the *t*-haplotype (Hammer and Silver 1993). However, inversion 4, which was hypothesized to have been acquired much later (Hammer and Silver 1993), contains peaks of nearly equally high divergence at ~37 Mb. More generally, levels of divergence differ less between the inversions than they do within them, with putative inversion boundaries often coinciding with major peaks of divergence (Figure 1).

### *Phylogenetic patterns along the t-haplotype suggest widespread recombination with the standard chromosome*

While other factors could create a mosaic pattern of differentiation, the colocalization of high divergence and inversion boundaries suggests that recombination with the standard chromosome in the middle of inversions (Wallace and Erhart 2008) may have eroded the genetic differentiation of the *t*-haplotype, as has been suggested by several smaller-scale studies (Herrmann *et al.* 1987; Erhart *et al.* 1989, 2002; Hammer *et al.* 1991; Wallace and Erhart 2008). Such recombination events can be detected through changes in the phylogenetic topology of the *t*- and standard chromosomes of the *M. musculus* subspecies: segments of the *t*-haplotype that have not recombined since the split of the three subspecies should appear as an outgroup to them, while segments that have undergone recent recombination should cluster within the species group.

After removing SNPs from regions that were classified as CNVs, we created "pseudo-*t*-haplotype" SNP profiles from each of the VCF files of the 15 *t*-carrier mice provided by

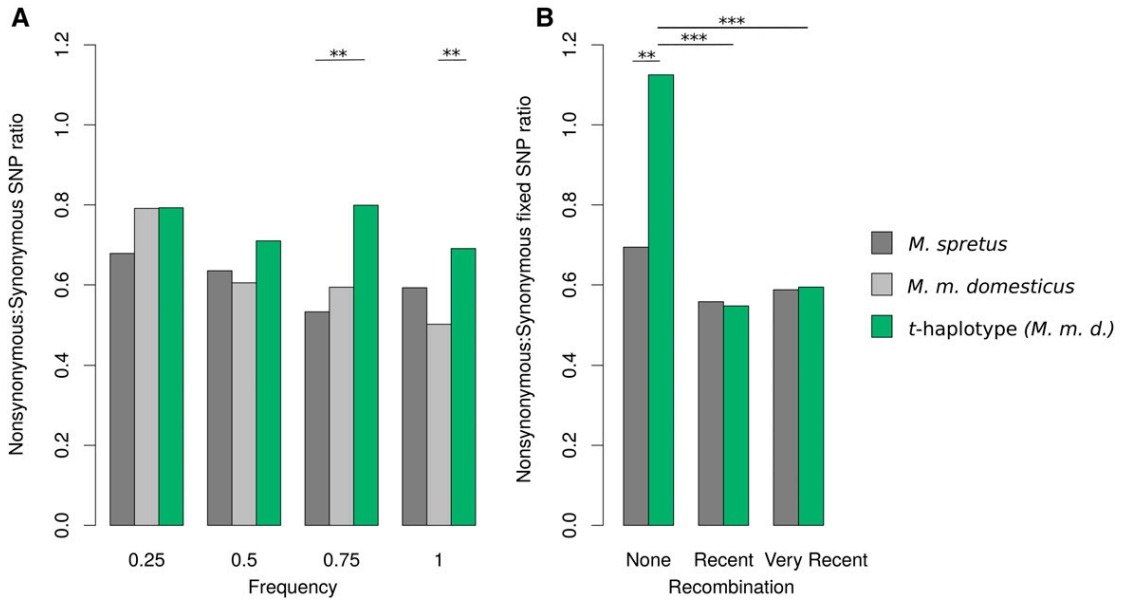



**Figure 4** Nonsynonymous to synonymous (NS/S) SNP ratio in the *t*-complex region. (A) NS/S for SNPs found at different frequencies in the pseudo-*t*-haplotypes (green), in the non-*t*-carrier control *M. m. domesticus* (light gray), and in *M. spretus* individuals (dark gray) for the *t*-complex. (B) NS/S for *t*-haplotype (green) and *M. spretus* (dark gray) SNPs found in regions for which no recent recombination could be detected in *M. m. domesticus* *t*-haplotypes, recent recombination could be detected, and very recent and/or extensive recombination could be detected. The categories were determined based on the phylogenetic topologies shown in Figure 2.

Harr *et al.* (2016). Since these mice are heterozygous for the *t*-haplotype and also carry a standard chromosome 17, we discarded heterozygous SNPs that were also found in any of the noncarrier individuals. Homozygous SNPs were presumed to be both on the standard chromosome and *t*-haplotype, and therefore kept even if also present in noncarriers.

After obtaining *t*-specific SNPs for each of the 15 *t*-carriers, we created *t*-haplotype sequences by replacing these SNPs into the *M. m. domesticus* reference genome. We did the same for each of the noncarrier mice (12 *M. m. domesticus*, 13 *M. m. musculus*, 7 *M. m. castaneus*, and 8 *M. spretus* individuals), for which the list of variants was supplied by Harr *et al.* (2016). To verify the reliability of our pipeline, we applied it to a region of the gene *Tcp-1* for which *t*-haplotype and standard sequences have been published (Morita *et al.* 1992; Figure S6 in File S2). Of the 24 published *t*-specific SNPs, 22 (92%) were recovered on our pseudo-*t*-haplotypes.

Using the HKY nucleotide substitution model of the phylogenetic software IQTree, we estimated the phylogenetic topology of the 15 *t*-haplotypes and 40 noncarriers in non-overlapping 5-kb windows along the *t*-complex (5–40 Mb on chromosome 17). For each subspecies, we observed three distinct tree topologies (Figure 2, A–C): (1) at least one of their *t*-haplotypes was positioned within the subspecies; (2) at least one of their *t*-haplotypes was positioned within the *M. musculus* clade, but all their *t*-haplotypes were outgroups relative to the noncarriers of the subspecies; and (3) all *t*-haplotypes from the subspecies were located outside of the *M. musculus* clade. The first type of windows supports more recent and/or extensive recombination events, and the second type older or smaller-scale recombination events. We also obtained phylogenetic trees for each of these windows using maximum parsimony and neighbor-joining approaches (Figure S7, A and B in File S2, respectively). The resulting tree topologies are consistent for all three methods in 75% (*M. m. castaneus*) to 85% (*M. m. domesticus* and *M. m. musculus*) of the windows, and similarly distributed along the *t*-haplotype for all the methods.

There is a good correspondence between the peaks of high divergence and regions where most windows show no evidence of recent recombination between the *t* and standard chromosomes (green bars in Figure 1). The phylogenetic patterns along the *t*-haplotype therefore support the view of four ancient inversions, of which large sections have been replaced by genetic material from the standard chromosome, likely through occasional events of recombination between the two.

### *The phylogeny of the t-haplotype does not mirror that of the standard chromosome, but does not support a single recent introgression*

A previous phylogenetic analysis suggested that a single *t*-haplotype introgressed into all *M. musculus* subspecies $< 0.8$ MYA (Morita *et al.* 1992). Figure 3A shows the expected phylogeny under this scenario: all *t*-haplotypes are highly diverged from the standard chromosomes, but very similar to each other, and the *t*-haplotype tree is polytomic. Figure 3B is the expected phylogeny if *t*-haplotypes were present in the three *M. musculus* subspecies before these split, and have been maintained in each independently. In

this case, the phylogenetic topology of *t*-haplotypes reflects the history of the *M. musculus* species complex.

To test these two models, we estimated the phylogeny of the 15 *t*-haplotypes and 40 noncarrier mice using only the 364 5-kb windows for which no recombination could be detected using any subspecies and phylogenetic method (to exclude signals caused by recent genetic exchange with the standard chromosomes). Figure 3C shows that the resulting phylogeny is not fully consistent with a very recent sweep of a single *t*-haplotype across the three subspecies: *t*-haplotypes have diverged sufficiently for *M. m. castaneus* and *M. m. domesticus* *t*-haplotypes to cluster by subspecies, while *M. m. musculus* *t*-haplotypes are outside of the *M. m. castaneus/M. m. domesticus* cluster. Since polytomies can mistakenly yield highly supported resolved trees (White *et al.* 2009), we tested whether the branch leading to the *M. m. castaneus/M. m. domesticus* cluster was significantly different from zero (Almeida *et al.* 2011). We took the maximum likelihood tree shown in Figure 3C, manually collapsed this branch, and ran the IQ-tree "tree topology test" on the original and the polytomic trees. This yielded much higher support for the original tree [$P = 0$, Shimodaira–Hasegawa test (Shimodaira and Hasegawa 1999)].

We ran two more controls to check that the clustering of the *M. m. castaneus* and *M. m. domesticus* *t*-haplotypes was not an artifact of the data or analysis (Figure S8 in File S2). First, we reestimated the phylogeny using more stringent pseudo-*t*-haplotype SNP profiles, which included only SNPs that were not found in any of the noncarriers of any subspecies, even if they were homozygous in *t*-carriers (Figure S8B in File S2). The *M. m. musculus* *t*-haplotypes remained outside of the *M. m. domesticus/M. m. castaneus* cluster. Second, we applied our SNP subtraction pipeline to the rest of chromosome 17 (50–90 Mb), to check that our procedure was removing enough SNPs from *t*-carriers to prevent them from clustering by subspecies simply due to residual variants (Figure S8A in File S2). This yielded an unresolved species tree for the *t*-carriers.

The data were also inconsistent with a simple model of maintenance of an ancestral *t*-haplotype in the three subspecies: while the species tree obtained for the noncarriers reflected the presumed history of the species complex (White *et al.* 2009), with *M. m. castaneus* and *M. m. musculus* clustering as sister species, *M. m. musculus* was an outgroup to the other two for the *t*-haplotype. This discrepancy was fairly consistent: 83% of the resolved 5-kb windows clustered *M. m. musculus* and *M. m. castaneus* for the noncarrier individuals, whereas 54% of such windows placed *M. m. musculus* as an outgroup for the *t*-haplotype (Figure 3D). Maximum parsimony and neighbor-joining approaches also supported primarily the *M. m. domesticus/M. m. castaneus* *t*-haplotype sister relationship (Figure S9, A and B in File S2, respectively). This confirms that *t*-haplotypes were still exchanged between the subspecies during early speciation (Morita *et al.* 1992), but suggests that some genetic flow persisted for longer between *M. m. domesticus* and *M. m. castaneus*.

While these patterns generally hold using our alternative SNP-filtering procedures (Figure S1 and Figure S2 in File S2), the allele frequency filtering yields support for both the *M. m. castaneus/M. m. domesticus* and the *M. m. musculus/M. m. domesticus* sister relationships. This seems to be driven by the loss of many heterozygous SNPs in low-coverage individuals due to deviations in allele frequency from 50%; when only a coverage filter is applied, the results once again support primarily the *M. m. castaneus/M. m. domesticus* clustering of *t*-haplotypes.

### *Recombination with the standard chromosome counteracts the genetic deterioration of the t-haplotype*

We estimated the ratio of nonsynonymous to synonymous SNPs (NS/S) of the pseudo-*t*-haplotype SNP profiles and compared it to the respective ratio for noncarrier *M. m. domesticus* individuals. When all SNPs are considered, NS/S is similar (0.74 for the *t*-haplotypes and 0.69 for noncarriers); however, most nonsynonymous SNPs found on the standard chromosomes are segregating at low frequency, as expected if they are overall deleterious, whereas many are fixed or at high frequency on the *t*-haplotype. We therefore reestimated NS/S for different SNP frequency classes (Figure 4A) among the four *t*-haplotypes, 24 control *M. m. domesticus* and 16 *M. spretus* chromosomes, for the *t*-complex region. As expected, both *M. m. domesticus* and *M. spretus* show a decreased NS/S ratio for high-frequency SNPs. Pseudo-*t*-haplotypes harbor an excess of nonsynonymous SNPs for all frequency classes. This difference is more pronounced for mutations that are shared by 50–100% of the chromosomes ($P = 0.07$ and $P = 0.007$ for frequency classes 0.75 and 1, respectively, between *M. m. domesticus* and the *t*-haplotypes, and $P = 0.006$ and $P = 0.06$ for the corresponding comparisons between *M. spretus* and the *t*-haplotype using Yates-corrected $\chi^2$ test), consistent with the idea that the *t*-haplotype has accumulated an excess of deleterious variants.

It was recently suggested that occasional gene flow between a meiotic drive system of *Drosophila* and the standard chromosome was sufficient to purge deleterious mutations from the driver (Pieper and Dyer 2016), and that this may contribute to its long-term viability. We similarly hypothesized that occasional recombination between the *t*-haplotype and the standard chromosome may contribute to the regeneration of coding sequences, so fixed SNPs in nonrecombined regions should have higher NS/S overall than regions that have recently recombined with the standard chromosome. We assigned SNPs to nonrecombined, recently or very recently recombined regions of the *t*-complex based on the *M. m. domesticus* phylogenetic topologies shown in Figure 2. For each category, we computed NS/S for SNPs found on the *t*-haplotypes and on *M. spretus* (as a control for differences in selective pressure along the chromosome). Figure 4B shows that while *M. spretus* harbors no difference in NS/S between the three types of regions, *t*-haplotypes have higher NS/S in the nonrecombined regions than in either class of recombined regions ($P < 0.001$, Yates-corrected $\chi^2$ tests) and

the corresponding regions in *M. spretus* ($P = 0.008$). The two classes of recombined regions are not significantly different from each other or from the corresponding regions in *M. spretus*.

The decreased NS/S in the most recently/extensively recombined regions was observable using both of our alternative filtering procedures (Figure S1 and Figure S2 in File S2), but did not yield significant differences for the third filtering procedure (based on coverage and allele-specific frequency, Figure S2 in File S2), again due to the removal of many heterozygous SNPs in low-coverage individuals.

#### *CNVs drive expression divergence in t-haplotype carriers*

While several chromosome 17 genes were found to differ in expression between *t*-haplotype carriers and noncarriers (Lader *et al.* 1989; Ha *et al.* 1991; Braidotti and Barlow 1997; Zwart *et al.* 2001), the overall effect of carrying a degenerating *t*-haplotype on genome-wide patterns of gene expression has not yet been assessed. We took advantage of the availability of RNA-seq data for several tissues derived from the same *M. m. domesticus* invididuals (Harr *et al.* 2016) to contrast gene expression levels (in Transcripts Per Million, TPM) between the four *t*-carrier mice and all noncarriers from France and Germany (Table S1 in File S2).

Figure 5 shows the extent to which expression has changed along the *t*-haplotype in brain, liver, and testis (other tissues are shown in Figure S10 in File S2), using a sliding window of 20 genes. Although different genes are differentially expressed in each tissue (Table S1 in File S2 and supplemental data in File S2), the general patterns of divergence are similar for all tissues, with large peaks of expression divergence at ~5 and 39 Mb. Both of these regions overlap with CNVs that were detected when the coverage of *t*-haplotype carriers and noncarriers was compared (gray bars in Figure 5).

Duplications and deletions are known to affect gene expression and have been detected for several differentially expressed genes on the *t*-haplotype (Braidotti and Barlow 1997; Zwart *et al.* 2001). Similarly, gene amplification was recently found to be essential for R2D2, another mouse meiotic driver (Didion *et al.* 2016; Morgan *et al.* 2016), and meiotic drivers on the mouse sex chromosomes have been postulated to lead to the extensive gene amplification that is observed on both the X and Y chromosomes (Soh *et al.* 2014). We therefore tested whether genes that overlapped with *t*-specific CNVs had diverged more in expression than the rest of the *t*-haplotype. The boxplots in Figure 5 show that this is indeed the case ($P < 10^{-6}$ in all three tissues), with a median change of $> 30\%$ for CNV-overlapping genes *vs.* 10% for other genes.

Genes located in windows that have recently recombined with the standard chromosome were expected to show lower levels of gene expression divergence. We classified genes into each recombination class if $\geq 80\%$ of the windows they overlapped with were of that class. While genes in recently recombined regions had a lower median percentage change than unrecombined windows in all tissues (Figure 5 and Figure S10 in File S2), the difference was generally not significant. This is likely due to the fact that highly diverged regions often overlap with CNVs (Figure 1) and were excluded from our phylogenetic analysis, such that only a few genes were left in unrecombined regions.

Finally, while our differential expression analysis in the testis recovered mainly chromosome 17 genes that were previously known to differ in expression in *t*-carriers (Lader *et al.* 1989; Ha *et al.* 1991; Braidotti and Barlow 1997; Zwart *et al.* 2001), one of the genes with the lowest *q*-value, Ppp1cb, is located on chromosome 5; despite not being on the *t*-haplotype, it shows a consistent 10-fold overexpression in *t*-carriers (Table S1 in File S2). Protein phosphatase 1 proteins are known to be essential for spermatogenesis (Silva *et al.* 2014). One of the active forms of PP1 has been shown to repress sperm motility in the epididymis, making Ppp1cb a promising candidate for involvement in drive and/or response to the driver (Vijayaraghavan *et al.* 1996). Another 2 out of 12 differentially expressed genes (Dr1 and Scamp2) are located on other chromosomes, emphasizing that regulatory changes on the *t*-haplotype can affect its biology through changes in the expression of genes located on other chromosomes.

## Discussion

Despite having been studied for close to a century, reduced recombination rates on the *t*-haplotype have limited the power of traditional genetic studies for this selfish element, and next-generation sequencing approaches offer a promising alternative to complement this body of work.

The variable levels of divergence along the *t*-haplotype complicate the inference of the history of the four inversions. While large sections of inversion 4 have lower divergence levels than the other inversions, as expected if it was acquired later (Hammer and Silver 1993), a peak of very high divergence is found at ~37–40 Mb. Two hypotheses could account for this: (1) inversion 4 may be of similar age as inversion 2, but much of its differentiation may have been lost through recombination with the standard chromosome, and (2) this region may have particularly high rates of divergence. Although we do find evidence of recombination over much of inversion 4, the region of highest divergence contains clusters of olfactory, immune, and pheromone genes, all of which tend to be highly polymorphic and fast evolving (Figure S3 in File S2). We control for these by normalizing by the noncarrier heterozygosity, by checking that the pattern holds when only neutral SNPs are used (Figure S4B in File S2), and by comparing the SNP density of the pseudo-*t*-haplotype to the SNP density of *M. spretus* (Figure S4, C–F in File S2). However, the reduced recombination rates on the *t*-haplotype may have led to the fixation of many ancestral neutral polymorphisms, and consequently increased rates of neutral divergence specifically on the *t*-haplotype. These results are therefore not incompatible with a younger age of inversion

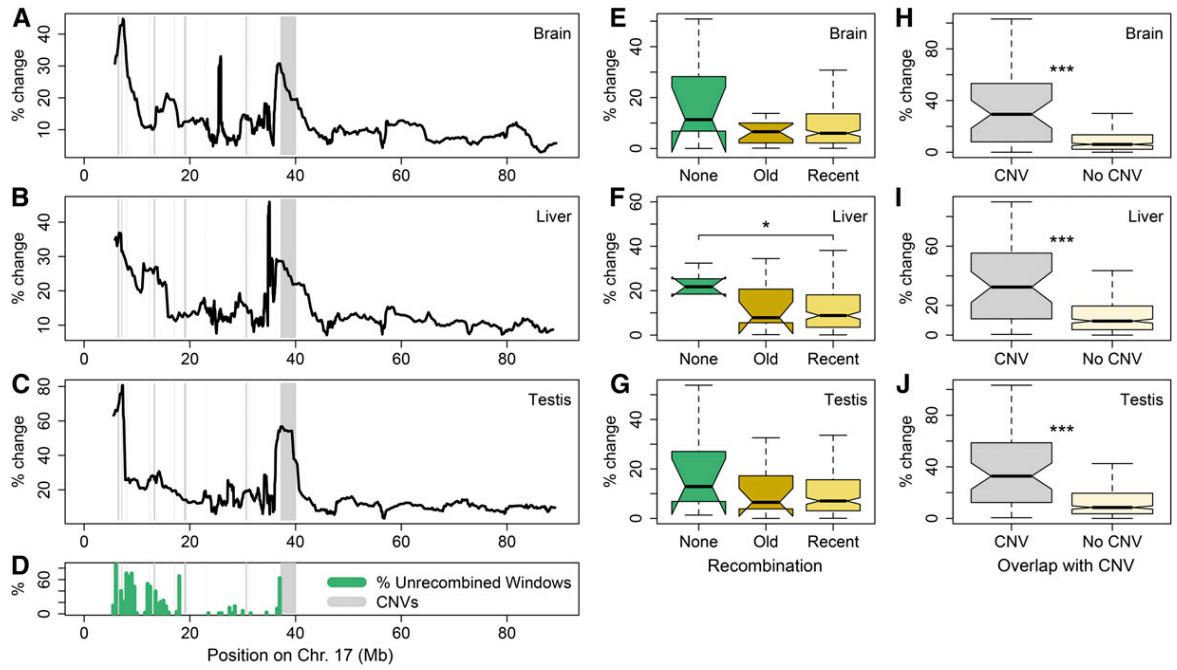



**Figure 5** Divergence of gene expression between *t*-carriers and noncarriers. (A–C) Percentage difference between the average gene expression of *t*-carrier and noncarriers (estimated as: | average_t-carrier − average_noncarrier | / average_noncarrier), plotted using a sliding window of 20 genes (using all genes with expression values > 10 in noncarriers). Expression divergence is shown for (A) the brain, (B) the liver, and (C) the testis. Regions that contain *t*-specific copy number variants (CNVs) (obtained by comparing the coverage of *t*-carriers to noncarriers, see *Materials and Methods*) are marked by gray rectangles. (D) The percentage of 5-kb windows for which no recombination was detected on *M. m. domesticus* t-haplotypes. (E–G) Boxplots showing the percentage difference in expression of *t*-carriers relative to that of noncarriers for genes that overlap with at ≥ 80% 5-kb windows for which no recombination was detected (green), some/old recombination was detected (orange), and recent/extensive recombination was detected (yellow), in (E) the brain, (F) the liver, and (G) the testis. (H–J) Boxplots showing the percentage difference in expression of *t*-carriers relative to that of noncarriers for genes overlapping or not overlapping a CNV, in (H) the brain, (I) the liver, and (J) the testis.

4, and emphasize that care should be taken when interpreting data obtained from small genomic regions.

Genetic exchange between the *t*-haplotype and the standard homolog was supported both by the phylogenetic topology and by the colocalization of some of the most diverged regions with putative inversion boundaries. Although several studies have found evidence for small-scale gene conversion in the fourth and largest inversion (Herrmann *et al.* 1987; Erhart *et al.* 1989, 2002; Hammer *et al.* 1991; Wallace and Erhart 2008), the extent of recombination that we observe here is unexpected, as repressed recombination is thought to be a hallmark of successful segregation distorters. However, it is in-line with Pieper and Dyer (2016), who suggested that occasional recombination events could provide a mechanism to counteract the accumulation of deleterious mutations on meiotic drivers due to Hill–Robertson effects. Consistent with this, the excess of fixed nonsynonymous SNPs on the *t*-haplotype is reduced in regions for which we detect recent recombination. Another effect of occasional gene flow with the standard homolog may be the maintenance of optimal gene expression levels on the *t*-haplotype. Although several genes showed altered expression in *t*-carriers (12 out of 463 genes in the testis, fewer in other tissues), the vast majority did not, and it is likely that such conserved expression results at least in part from genetic homogenization due to recombination.

Finally, our phylogenetic analysis uncovered variation between *t*-haplotypes sampled from the three *M. musculus* subspecies, and a phylogenetic topology that disagrees with that of the standard subspecies tree, but also seems inconsistent with a very recent introgression of a single *t*-haplotype. Some caveats should be taken into account when interpreting these data. First, we use pseudo-*t*-haplotypes, which may contain some residual SNPs from the standard chromosome, and it will be important to confirm these results using sequences derived from homozygous *t*-carriers. Second, gene conversion from the standard chromosome to the *t*-haplotype could result in *t*-haplotypes becoming quickly differentiated after introgressing. Finally, the evolutionary history of the *M. musculus* subspecies complex is itself challenging to disentangle, as the three subspecies are estimated to have diverged less than half a million years ago, and because there is a varying rate of gene flow between them and across genomic regions (Geraldes *et al.* 2008). Despite this variance, 39% of the genome supports the *M. m. musculus/M. m. castaneus* sister species relationship (White *et al.* 2009), which is most likely the primary phylogenetic history (in agreement with our findings for noncarriers). Contrary to this, the *M. m. castaneus* and *M. m. domesticus* *t*-haplotypes are most closely related in 53% of the windows. However, the fact that the other two topologies are also supported by 16% (for the

*M. m. musculus/M. m. castaneus* cluster) and 30% (*M. m. domesticus/M. m. castaneus* cluster) of windows suggests that, similar to what occurred on the standard chromosome, gene flow between the *t*-haplotypes of the different subspecies may have shaped the phylogenetic topology of this large meiotic driver, as expected if *t*-haplotypes were being regularly exchanged between subspecies during early speciation.

### *Conclusions*

Our global analysis of the sequence and expression patterns of the *t*-haplotype confirmed its ancient origin, the involvement of large parts of chromosome 17, and revealed an excess of nonsynonymous mutations consistent with the genetic deterioration that is expected in the absence of recombination. Surprisingly, this was counteracted by occasional recombination with the standard chromosome over a large proportion of the *t*-complex, providing an explanation for its long-term survival. Finally, the fact that most of the change in gene expression is driven by the accumulation of CNVs, but that regulatory changes on the *t*-haplotype can also affect the expression of genes elsewhere, provides new insights into the biology of the *t*-haplotype, and opens new avenues of exploration for this model segregation distorter.

## Acknowledgments

We are grateful to Dominik Schrempf for assistance with the phylogenetic analyses, to Brian Charlesworth for comments on the manuscript, and to the Vicoso laboratory for many lively discussions. This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number 715257).

## Literature Cited

Almeida, F. C., N. P. Giannini, R. DeSalle, and N. B. Simmons, 2011 Evolutionary relationships of the old world fruit bats (Chiroptera, Pteropodidae): another star phylogeny? BMC Evol. Biol. 11: 281.

Ardlie, K. G., 1998 Putting the brake on drive: meiotic drive of t haplotypes in natural populations of mice. Trends Genet. 14: 189–193.

Artzt, K., H. S. Shin, and D. Bennett, 1982 Gene mapping within the T/t complex of the mouse. II. Anomalous position of the H-2 complex in t haplotypes. Cell 28: 471–476.

Bauer, H., N. Véron, J. Willert, and B. G. Herrmann, 2007 The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. Genes Dev. 21: 143–147.

Bauer, H., S. Schindler, Y. Charron, J. Willert, B. Kusecek *et al.*, 2012 The nucleoside diphosphate kinase gene Nme3 acts as quantitative trait locus promoting non-Mendelian inheritance. *PLoS Genet.* 8: e1002567.

Boeva, V., T. Popova, K. Bleakley, P. Chiche, J. Cappo *et al.*, 2012 Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28: 423–425.

Braidotti, G., and D. P. Barlow, 1997 Identification of a male meiosis-specific gene, Tcte2, which is differentially spliced in species that form sterile hybrids with laboratory mice and deleted in t chromosomes showing meiotic drive. Dev. Biol. 186: 85–99.

Bray, N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34: 525–527.

Burt, A., and R. Trivers, 2009 *Genes in Conflict*. Harvard University Press, Cambridge, MA.

Campos, J. L., D. L. Halligan, P. R. Haddrill, and B. Charlesworth, 2014 The relation between recombination rate and patterns of molecular evolution and variation in Drosophila melanogaster. Mol. Biol. Evol. 31: 1010–1028.

Charlesworth, B., and D. L. Hartl, 1978 Population dynamics of the segregation distorter polymorphism of DROSOPHILA MELANOGASTER. Genetics 89: 171–192.

Chesley, P., and L. C. Dunn, 1936 The inheritance of taillessness (Anury) in the house mouse. Genetics 21: 525–536.

Cingolani, P., A. Platts, Le L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) 6: 80–92.

Didion, J. P., A. P. Morgan, L. Yadgary, T. A. Bell, R. C. McMullan *et al.*, 2016 R2d2 drives selfish sweeps in the house mouse. Mol. Biol. Evol. 33: 1381–1395.

Dod, B., C. Litel, P. Makoundou, A. Orth, and P. Boursot, 2003 Identification and characterization of t haplotypes in wild mice populations using molecular markers. Genet. Res. 81: 103–114.

Dyer, K. A., B. Charlesworth, and J. Jaenike, 2007 Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. Proc. Natl. Acad. Sci. USA 104: 1587–1592.

Erhart, M. A., S. J. Phillips, F. Bonhomme, P. Boursot, E. K. Wakeland *et al.*, 1989 Haplotypes that are mosaic for wild-type and t complex-specific alleles in wild mice. Genetics 123: 405–415.

Erhart, M. A., S. Lekgothoane, J. Grenier, and J. H. Nadeau, 2002 Pattern of segmental recombination in the distal inversion of mouse t haplotypes. Mamm. Genome 13: 438–444.

Geraldes, A., P. Basset, B. Gibson, K. L. Smith, B. Harr *et al.*, 2008 Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. Mol. Ecol. 17: 5349–5363.

Ha, H., C. A. Howard, Y. I. Yeom, K. Abe, H. Uehara *et al.*, 1991 Several testis-expressed genes in the mouse t-complex have expression differences between wild-type and t-mutant mice. Dev. Genet. 12: 318–332.

Hammer, M. F., and L. M. Silver, 1993 Phylogenetic analysis of the alpha-globin pseudogene-4 (Hba-ps4) locus in the house mouse species complex reveals a stepwise evolution of t haplotypes. Mol. Biol. Evol. 10: 971–1001.

Hammer, M. F., S. Bliss, and L. M. Silver, 1991 Genetic exchange across a paracentric inversion of the mouse t complex. Genetics 128: 799–812.

Harr, B., E. Karakoc, R. Neme, M. Teschke, C. Pfeifle *et al.*, 2016 Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. Sci. Data 3: 160075.

Harrison, A., P. Olds-Clarke, and S. M. King, 1998 Identification of the t complex-encoded cytoplasmic dynein light chain tctex1 in inner arm I1 supports the involvement of flagellar dyneins in meiotic drive. J. Cell Biol. 140: 1137–1147.

Herrmann, B., M. Bućan, P. E. Mains, A. M. Frischauf, L. M. Silver *et al.*, 1986 Genetic analysis of the proximal portion of the mouse t complex: evidence for a second inversion within t haplotypes. Cell 44: 469–476.

Herrmann, B. G., and H. Bauer, 2012 The mouse t-haplotype: a selfish chromosome-genetics, molecular mechanism, and evolution,

pp. 297–314 in *Evolution of the House Mouse*, edited by M. Macholán, S. J. E. Baird, P. Munclinger, and J. Piálek. Cambridge University Press, Cambridge.

Herrmann, B. G., D. P. Barlow, and H. Lehrach, 1987 A large inverted duplication allows homologous recombination between chromosomes heterozygous for the proximal t complex inversion. Cell 48: 813–825.

Herrmann, B. G., B. Koschorz, K. Wertz, K. J. McLaughlin, and A. Kispert, 1999 A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. Nature 402: 141–146.

Howell, G. R., R. A. Bergstrom, R. J. Munroe, J. Masse, and J. C. Schimenti, 2004 Identification of a cryptic lethal mutation in the mouse t(w73) haplotype. Genet. Res. 84: 153–159.

Khan, M. A., I. Elias, E. Sjölund, K. Nylander, R. V. Guimera *et al.*, 2013 Fastphylo: fast tools for phylogenetics. BMC Bioinformatics 14: 334.

Kumar, S., G. Stecher, and K. Tamura, 2016 MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33: 1870–1874.

Lader, E., H.-S. Ha, M. O'Neill, K. Artzt, and D. Bennett, 1989 tctex-1: a candidate gene family for a mouse t complex sterility locus. Cell 58: 969–979.

Larracuente, A. M., and D. C. Presgraves, 2012 The selfish segregation distorter gene complex of Drosophila melanogaster. Genetics 192: 33–53.

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078–2079.

Lindholm, A. K., K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier *et al.*, 2016 The ecology and evolutionary dynamics of meiotic drive. Trends Ecol. Evol. (Amst.) 31: 315–326.

Lyon, M. F., 2003 Transmission ratio distortion in mice. Annu. Rev. Genet. 37: 393–408.

Minh, B. Q., M. A. T. Nguyen, and A. von Haeseler, 2013 Ultrafast approximation for phylogenetic bootstrap. Mol. Biol. Evol. 30: 1188–1195.

Morgan, A. P., J. M. Holt, R. C. McMullan, T. A. Bell, A. M.-F. Clayshulte *et al.*, 2016 The evolutionary fates of a large segmental duplication in mouse. Genetics 204: 267–285.

Morita, T., H. Kubota, K. Murata, M. Nozaki, C. Delarbre *et al.*, 1992 Evolution of the mouse t haplotype: recent and worldwide introgression to Mus musculus. Proc. Natl. Acad. Sci. USA 89: 6851–6855.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh, 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32: 268–274.

Pieper, K. E., and K. A. Dyer, 2016 Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. J. Evol. Biol. 29: 2229–2241.

Pimentel, H. J., N. Bray, S. Puente, P. Melsted, and L. Pachter, 2016 Differential analysis of RNA-Seq incorporating quantification uncertainty. bioRxiv DOI: 10.1101/058164.

Presgraves, D. C., P. R. Gérard, A. Cherukuri, and T. W. Lyttle, 2009 Large-scale selective sweep among segregation distorter chromosomes in African populations of Drosophila melanogaster. *PLoS Genet.* 5: e1000463.

Schwander, T., R. Libbrecht, and L. Keller, 2014 Supergenes and complex phenotypes. Curr. Biol. 24: R288–R294.

Shimodaira, H., and M. Hasegawa, 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16: 1114–1116.

Silva, J. V., M. J. Freitas, and M. Fardilha, 2014 Phosphoprotein phosphatase 1 complexes in spermatogenesis. Curr. Mol. Pharmacol. 7: 136–146.

Soh, Y. Q. S., J. Alföldi, T. Pyntikova, L. G. Brown, T. Graves *et al.*, 2014 Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. Cell 159: 800–813.

Sugimoto, M., 2014 Developmental genetics of the mouse t-complex. Genes Genet. Syst. 89: 109–120.

Vijayaraghavan, S., D. T. Stephens, K. Trautman, G. D. Smith, B. Khatra *et al.*, 1996 Sperm motility development in the epididymis is associated with decreased glycogen synthase kinase-3 and protein phosphatase 1 activity. Biol. Reprod. 54: 709–718.

Wallace, L. T., and M. A. Erhart, 2008 Recombination within mouse t haplotypes has replaced significant segments of t-specific DNA. Mamm. Genome 19: 263–271.

White, M. A., C. Ané, C. N. Dewey, B. R. Larget, and B. A. Payseur, 2009 Fine-scale phylogenetic discordance across the house mouse genome. *PLoS Genet.* 5: e1000729.

Woolfit, M., 2009 Effective population size and the rate and pattern of nucleotide substitutions. Biol. Lett. 5: 417–420.

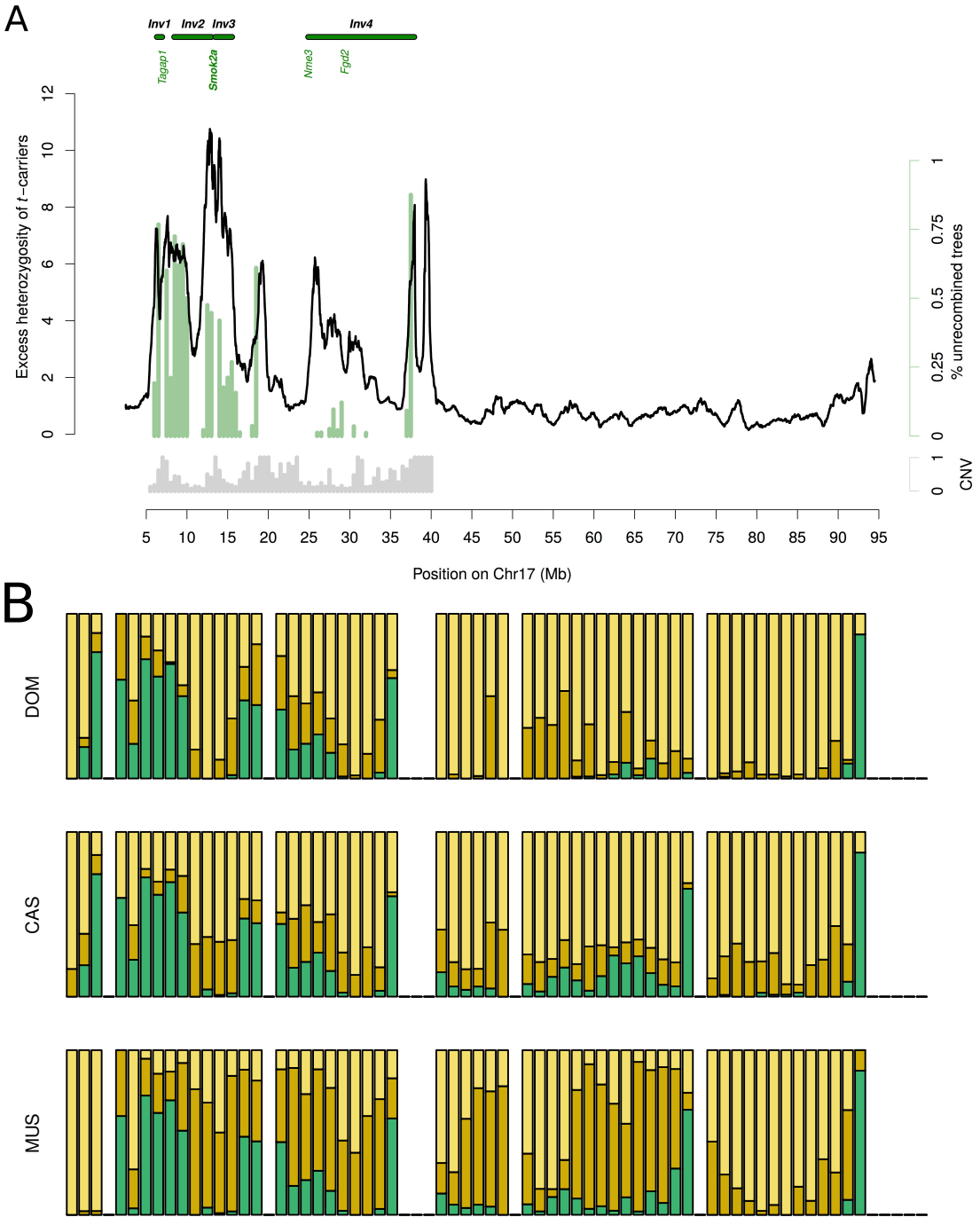
Zwart, R., S. Verhaagh, J. de Jong, M. Lyon, and D. P. Barlow, 2001 Genetic analysis of the organic cation transporter genes Orct2/Slc22a2 and Orct3/Slc22a3 reduces the critical region for the t haplotype mutant tw73 to 200 kb. Mamm. Genome 12: 734–740.

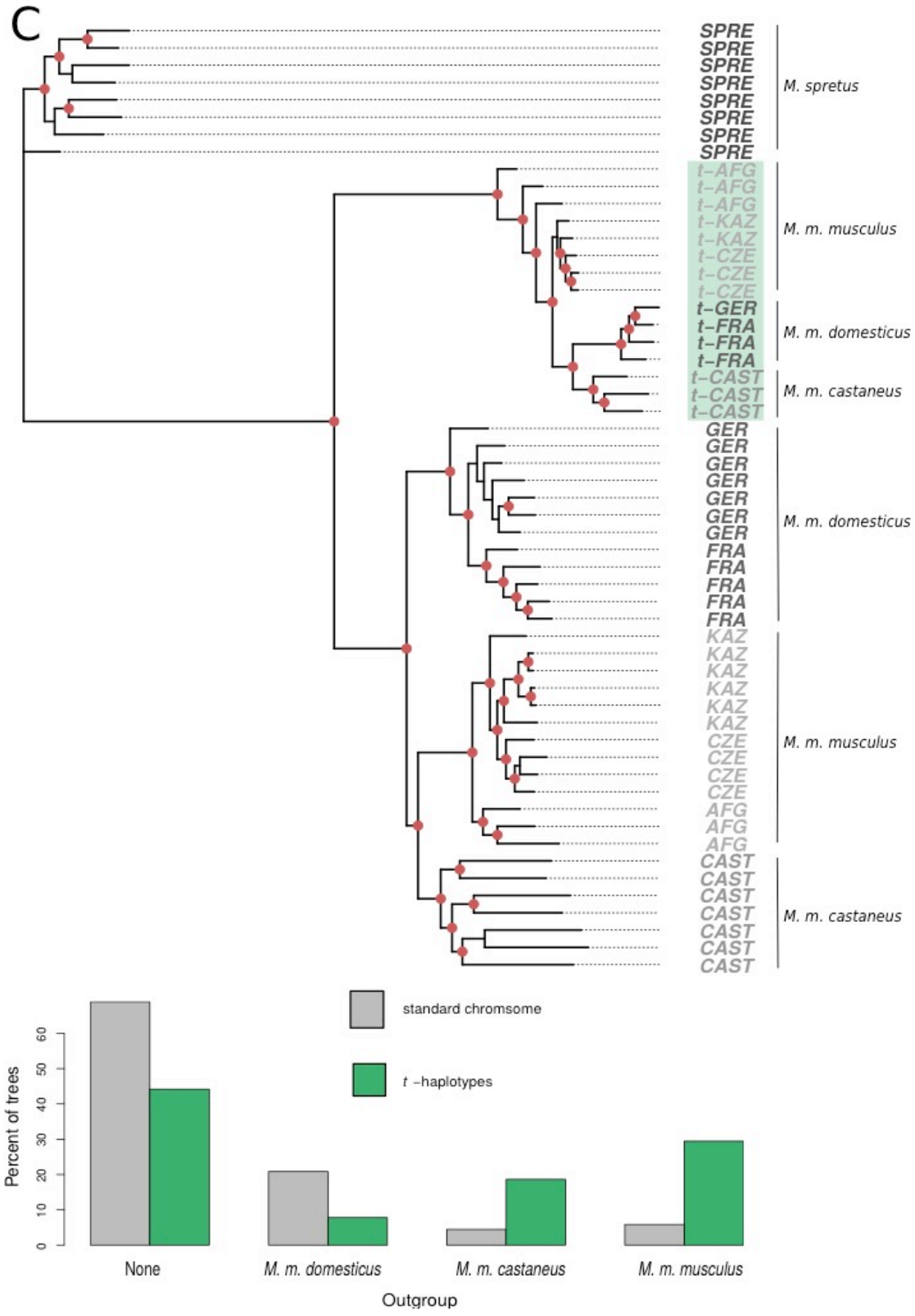
*Communicating editor: B. Payseur*

**Supplementary File 2**

This document contains:

**Figure S1 (continues in next two pages): Figures 1-4 using coverage-filtered RAW SNPs (filtering procedure 2).** A reanalysis using Harr et al.'s published raw SNP dataset, filtered based on coverage (see Materials and Methods). Panels A, B, C and D reproduce Figures 1, 2, 3 and 4, respectively.




C
SPRE
SPRE
SPRE
SPRE
SPRE
SPRE
SPRE
M. spretus
t–AFG
t–AFG
t–AFG
t–KAZ
t–KAZ
t–CZE
t–CZE
t–CZE
M. m. musculus
t–GER
t–FRA
t–FRA
t–FRA
M. m. domesticus
t–CAST
t–CAST
t–CAST
M. m. castaneus
GER
GER
GER
GER
GER
GER
GER
FRA
FRA
FRA
FRA
FRA
M. m. domesticus
KAZ
KAZ
KAZ
KAZ
KAZ
KAZ
CZE
CZE
CZE
CZE
AFG
AFG
AFG
M. m. musculus
CAST
CAST
CAST
CAST
CAST
CAST
CAST
M. m. castaneus
standard chromsome
t –haplotypes
Percent of trees
0
10
20
30
40
50
60
None
M. m. domesticus
M. m. castaneus
M. m. musculus
Outgroup

D

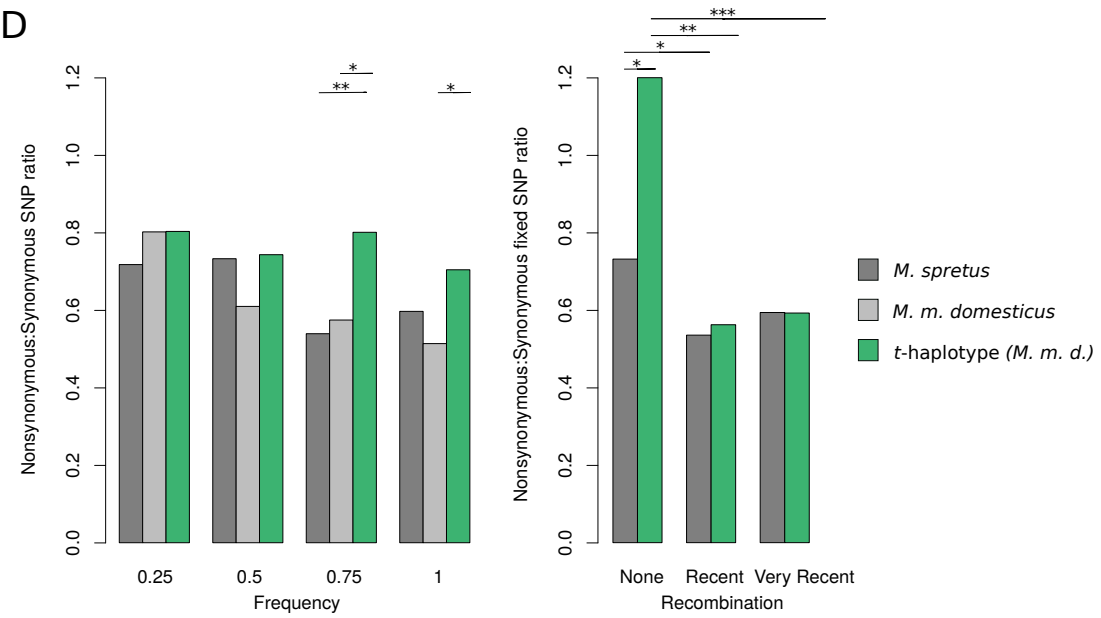
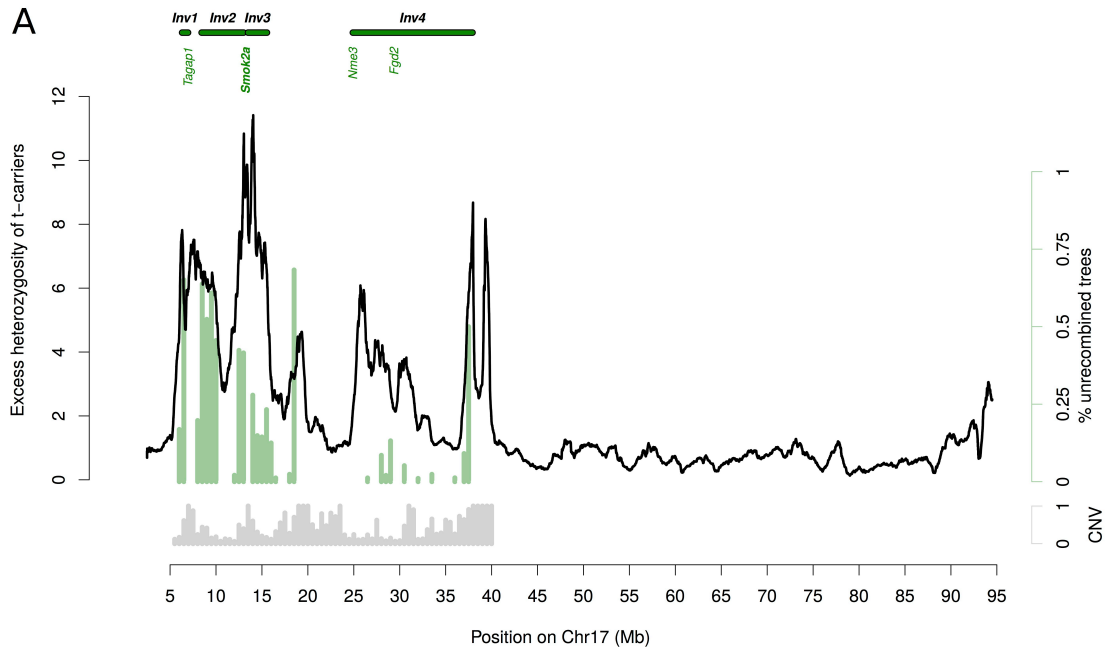
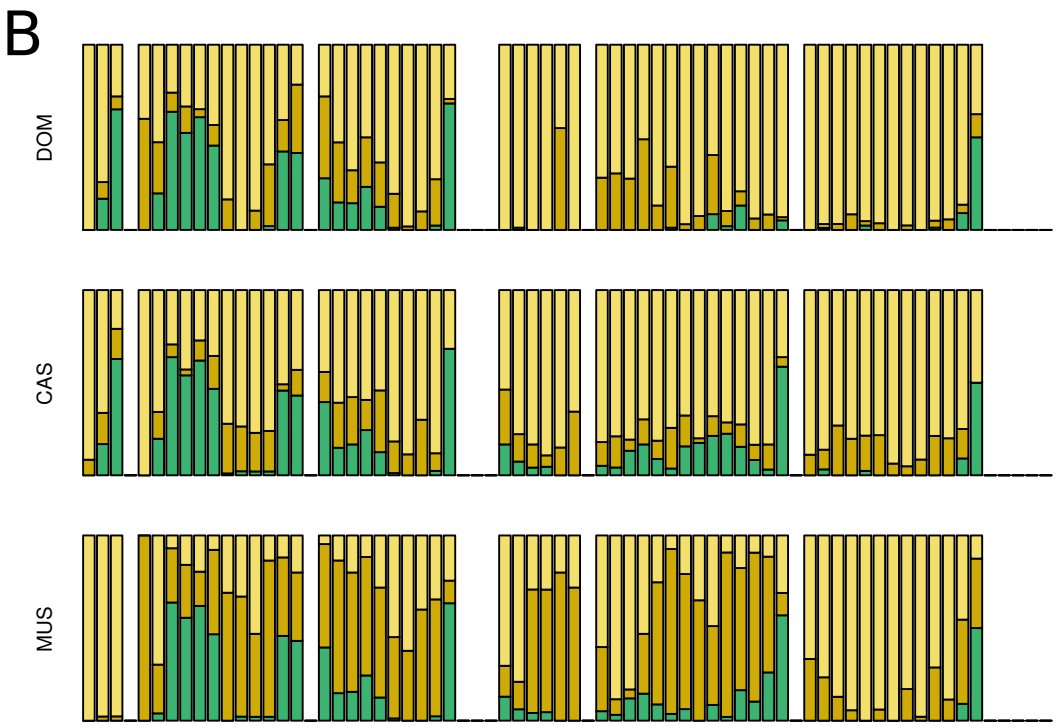
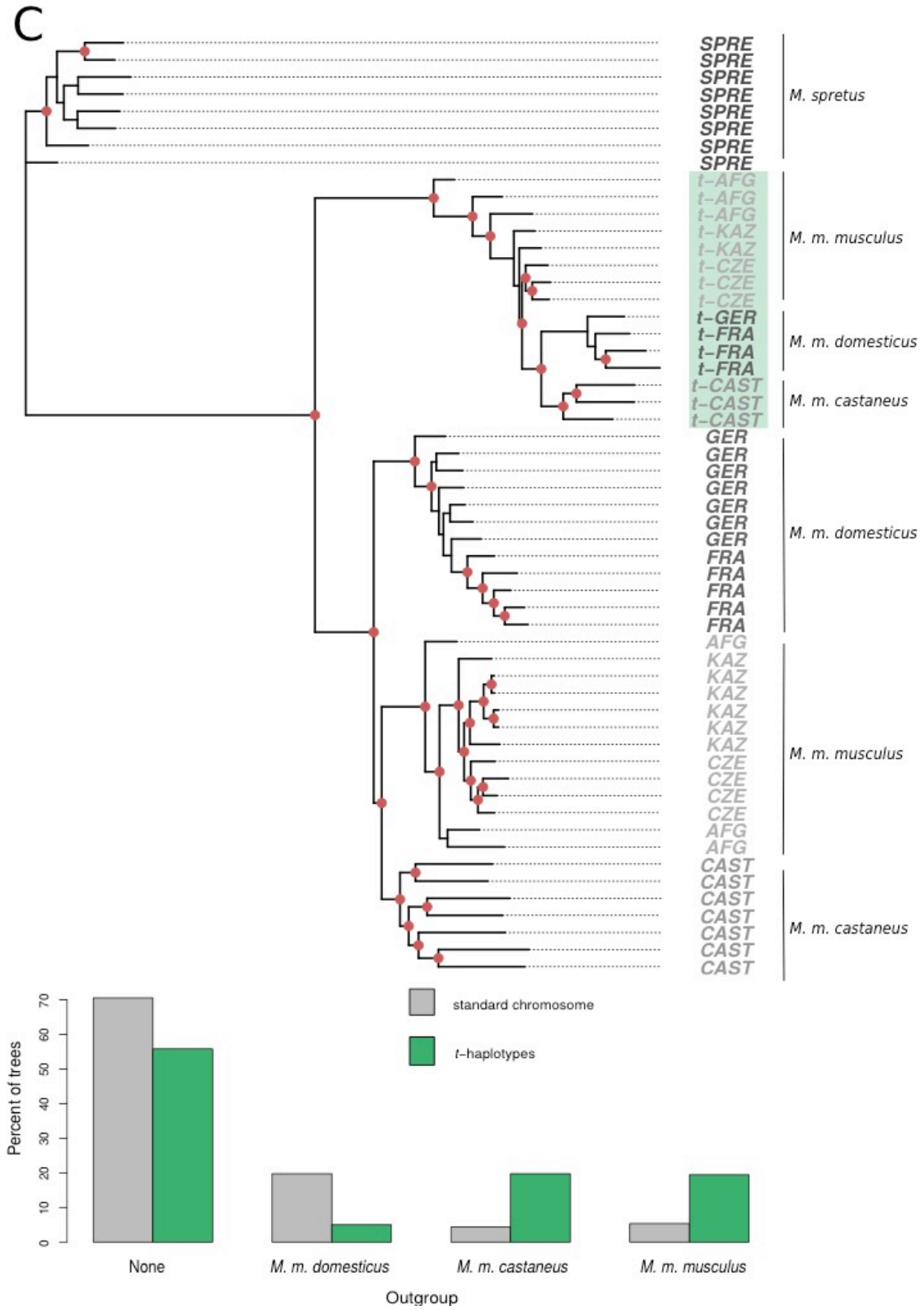

Nonsynonymous:Synonymous SNP ratio
0.0
0.2
0.4
0.6
0.8
1.0
1.2
*
**
*
0.25
0.5
0.75
1
Frequency
Nonsynonymous:Synonymous fixed SNP ratio
0.0
0.2
0.4
0.6
0.8
1.0
1.2
***
**
*
*
None
Recent
Very Recent
Recombination
M. spretus
M. m. domesticus
t-haplotype (M. m. d.)

**Figure S2 (continues in the next two pages): Figures 1-4 using coverage- and allele-frequency-filtered RAW SNPs (filtering procedure 3).** A reanalysis using Harr et al.'s published raw SNP dataset, filtered based on coverage and allele-frequency (see Materials and Methods). Panels A, B, C and D reproduce Figures 1, 2, 3 and 4, respectively.




C
SPRE
SPRE
SPRE
SPRE
SPRE
SPRE
SPRE
SPRE
M. spretus
t–AFG
t–AFG
t–AFG
t–KAZ
t–KAZ
t–CZE
t–CZE
t–CZE
M. m. musculus
t–GER
t–FRA
t–FRA
t–FRA
M. m. domesticus
t–CAST
t–CAST
t–CAST
M. m. castaneus
GER
GER
GER
GER
GER
GER
GER
FRA
FRA
FRA
FRA
FRA
M. m. domesticus
AFG
KAZ
KAZ
KAZ
KAZ
KAZ
KAZ
CZE
CZE
CZE
CZE
AFG
AFG
M. m. musculus
CAST
CAST
CAST
CAST
CAST
CAST
CAST
M. m. castaneus
standard chromosome
t–haplotypes
Percent of trees
0
10
20
30
40
50
60
70
None
M. m. domesticus
M. m. castaneus
M. m. musculus
Outgroup

D

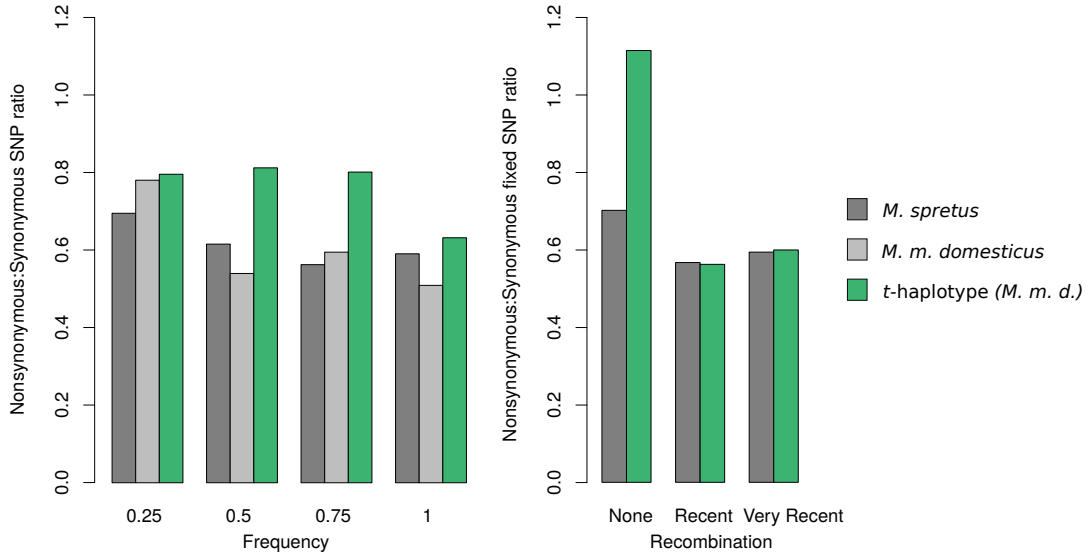

Nonsynonymous:Synonymous SNP ratio
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0.25
0.5
0.75
1
Frequency
Nonsynonymous:Synonymous fixed SNP ratio
None
Recent
Very Recent
Recombination
M. spretus
M. m. domesticus
t-haplotype (M. m. d.)

**Figure S3: SNP density in non-carrier *M. m. domesticus* individuals along chromosome 17.** The average number of heterozygous SNPs per kilobase per individual is shown in windows of 1 Mb, sliding in steps of 1 Kb. All 12 non-carrier *M. m. domesticus* were used. Blue, purple and gray lines on the bottom of the plot mark genomic regions with large clusters of olfactory, immune and pheromone genes.

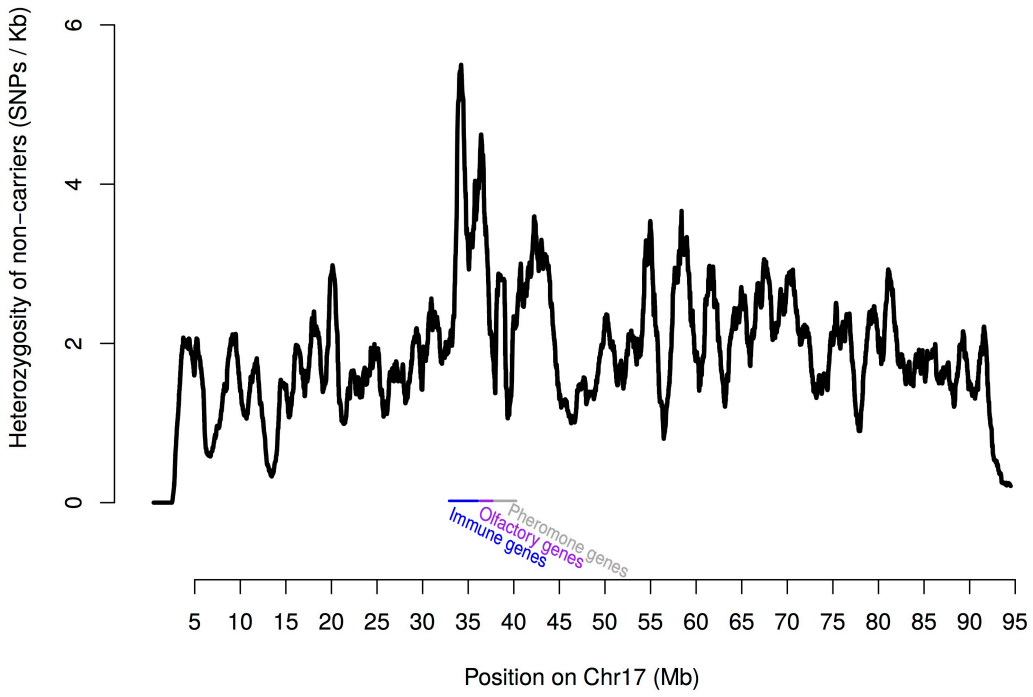


**Figure S4: Divergence of the *t*-haplotype normalized by *M. m. domesticus* or *M. spretus* and using all, or only neutral SNPs.** (A) Heterozygosity levels along the t-haplotype of *M. m. domesticus* estimated from all SNPs, and normalized by non-carrier heterozygosity (same as Figure 1). (B) Neutral heterozygosity levels along the t-haplotype of *M. m. domesticus*. The same procedure was followed as for Figure 1, but we only included intergenic or synonymous SNPs. (C) SNP density (per Kb) of pseudo-t-haplotypes normalized by that of M. spretus individuals. (D) Same as (C), but using only synonymous and intergenic SNPs. (E) SNP density (per Kb) of pseudo-t-haplotypes, *M. m. domesticus* non-carriers and *M. spretus* individuals. (F) Same as (E) but using only synonymous and intergenic SNPs. In all of the panels we used data for the entire chromosome 17 without masking CNVs.

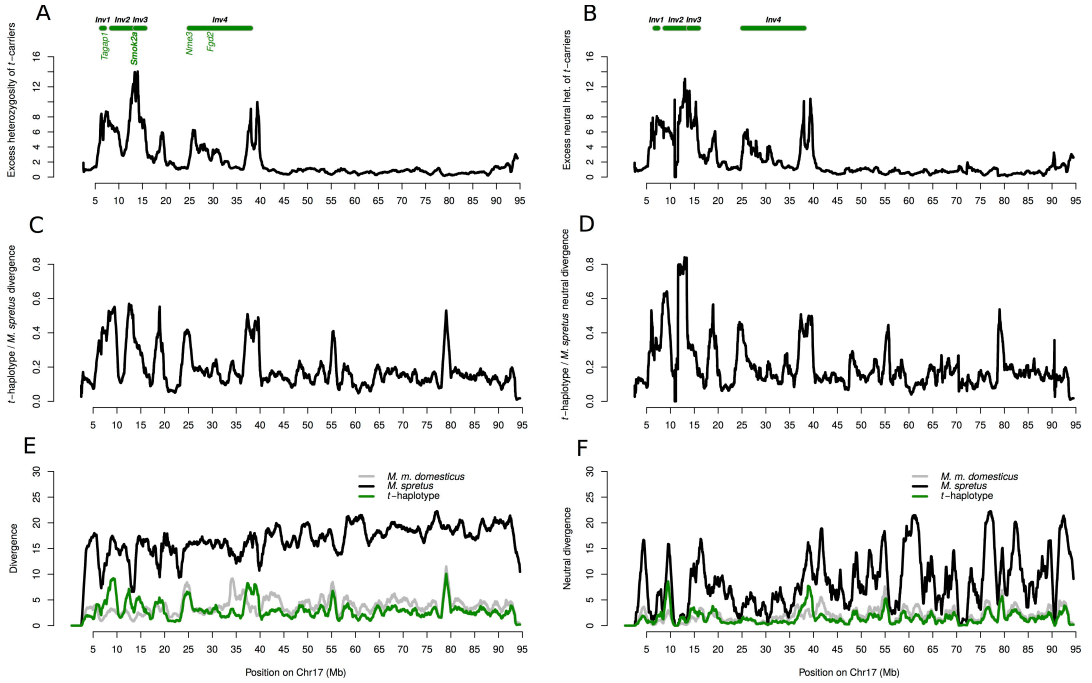


**Figure S5: Divergence of the *t*-haplotype measured in abundance of insertions and deletions (INDELs).** We plotted the ratio of INDEL density on chromosome 17 found in *t*-carriers compared to that found in non-carriers in *M. m. domesticus*. The densities were computed in windows of 1Kb, and the ratios were averaged over 1 Mb (sliding by 1Kb). We used the INDELs published in the raw SNP dataset of Harr et al. (2016). We plotted data for the entire chromosome 17 without masking CNVs.

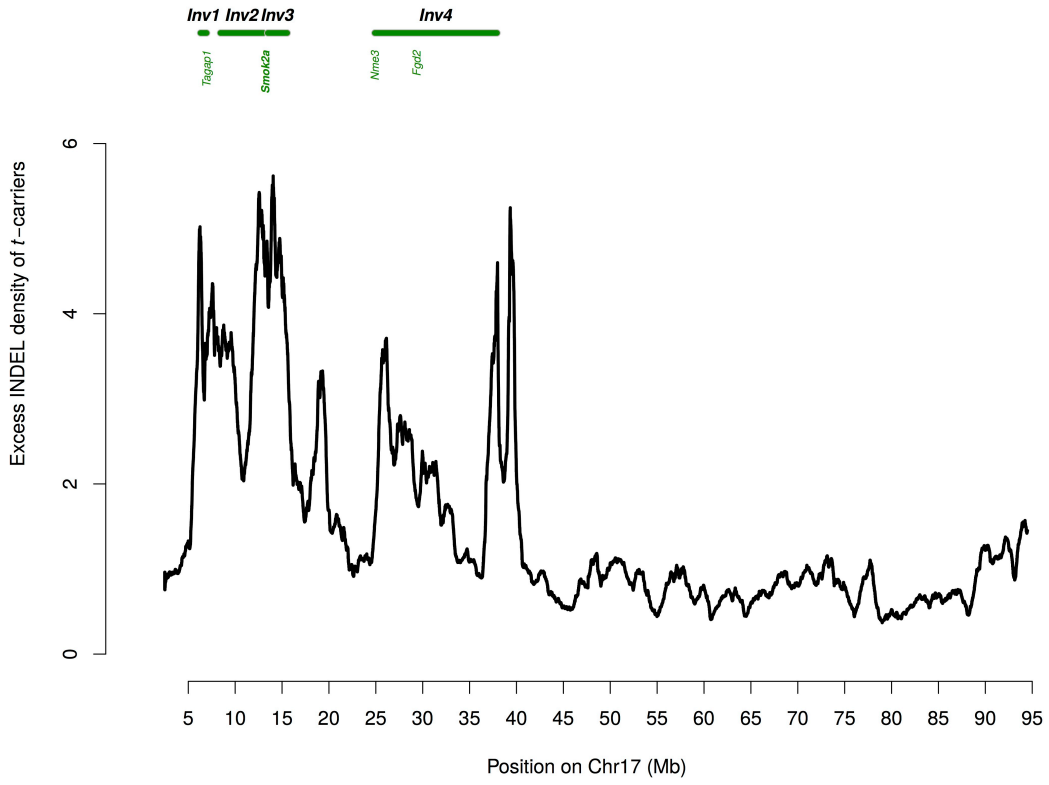


**Figure S6: Alignment of the *tcp-1* region (exons 8-10) for standard chromosomes and *t*-haplotypes of *M. m. musculus*, *M. m. domesticus*, *M. m. castaneus* and *M. spretus*.** Only sites that contain SNPs are shown. Sequences surrounded by red rectangles were generated by Morita et al. (1992) and retrieved from the NCBI nucleotide database (accessions X61222.1, X61212.1, X61219.1, X61217.1, X61215.1 and X61214.1). Other sequences were obtained from the variants provided in Harr et al. (2016). Arrows represent sites that were found in the original *t*-haplotype sequences but are not present on any of the standard *M. musculus* chromosomes. Black arrows represent the subset of these SNPs that are also detected on at least one pseudo-*t*-haplotype, red arrows mark SNPs that are not found in the pseudo-*t*-haplotypes.

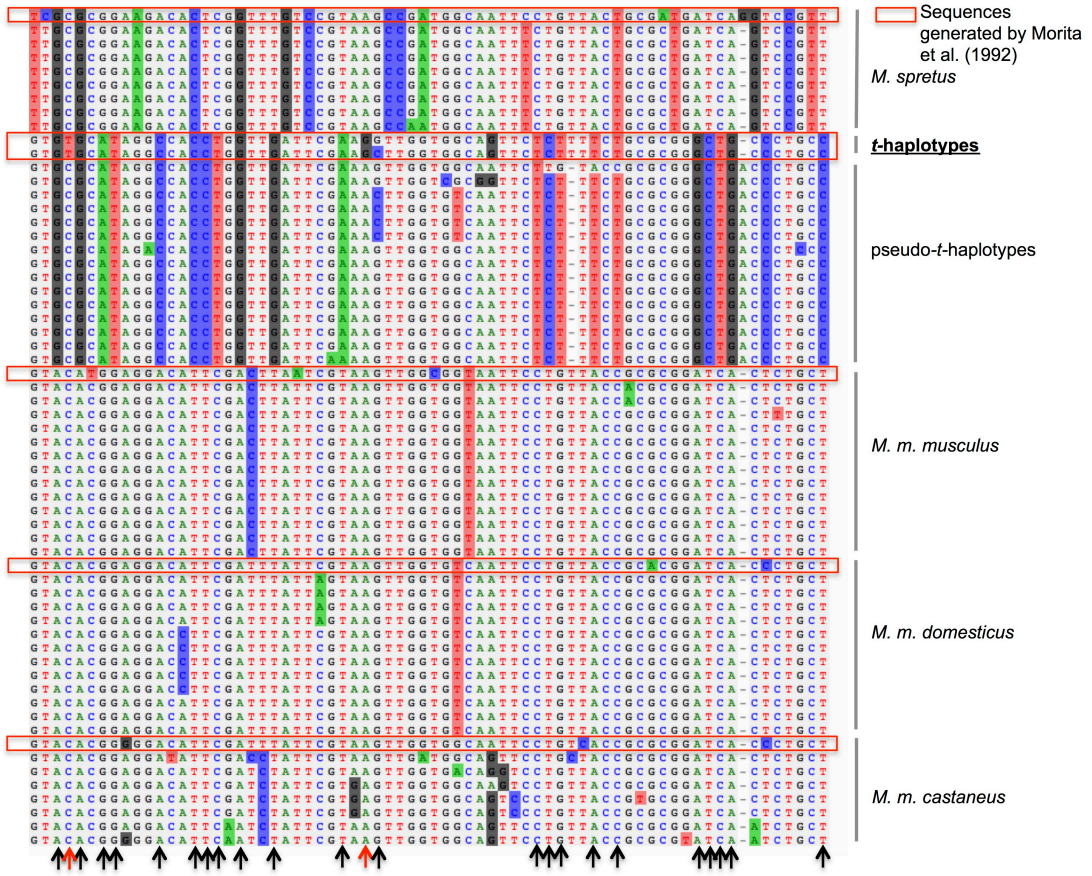


**Figure S7: Phylogenetic topology along the t-haplotypes of the three subspecies using maximum parsimony (A) and neighbor joining (B) algorithm.** The colors in the three bars represent the phylogenetic position of the "pseudo"-t-haplotypes from each of the three subspecies. The upper bar shows the results for *M. m. domesticus*, the middle bar for *M. m. musculus* and the lower bar for *M. m. castaneus*. Each segment in the bars corresponds to a 0.5 Mb window from 5-40 Mb on chromosome 17. The proportion of yellow, orange and green represent the proportion of the 5 Kb trees in the 0.5 Mb windows that show a certain topology. Yellow indicates that at least one t-haplotype is within its own subspecies, orange means that t-haplotypes are clustered within the *M. musculus* species complex but not within their respective subspecies, and green shows windows for which all t-haplotypes cluster outside of the *M. musculus* clade. DOM stands for *M. m. domesticus*, CAS for *M. m. castaneus* and MUS for *M. m. musculus*.

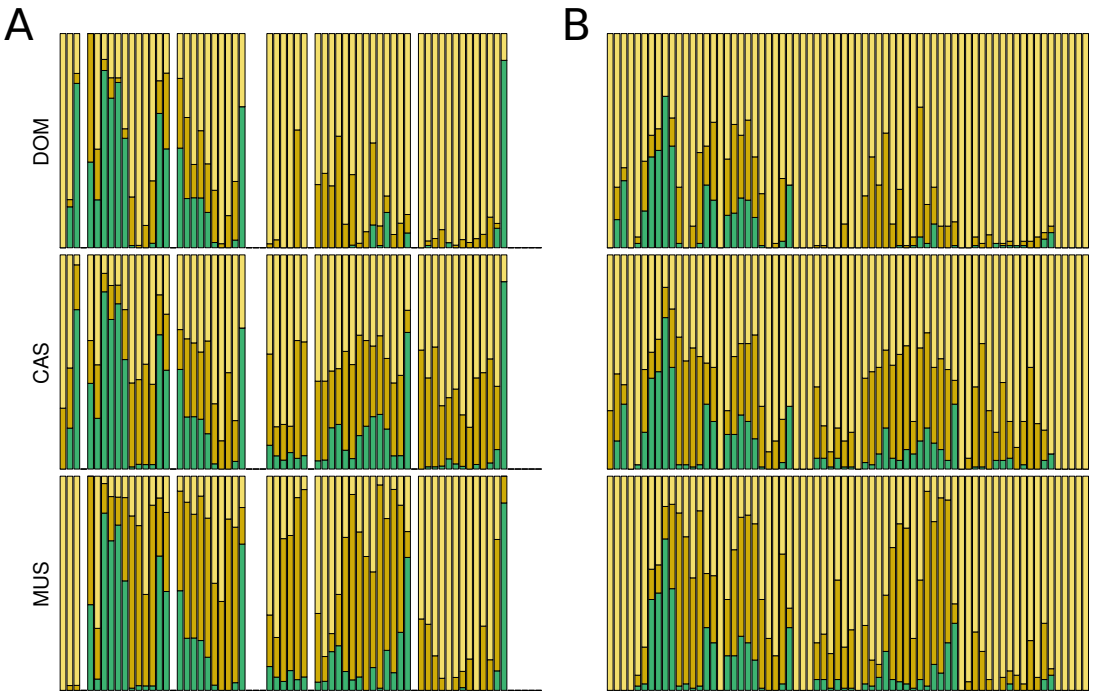


**Figure S8: The topology of the *t*-haplotype phylogeny is robust to the removal of duplicated regions and to reconstruction based only on SNPs private to *t*-carriers.** (A) Tree based on Chr. 17 region 50-90 Mb, which is outside of the *t* complex, using only SNPs that are private to *t*-carriers. (B) Tree based on all diverged regions (dark green bars in Figure 2 and Figure S7) for pseudo-*t*-haplotypes created using only SNPs that are private to *t*-carriers (independent of their homo/heterozygosity status). Red represents *M. m. castaneus*, orange *M. m. musculus*, yellow *M. m. domesticus*, and black *M. spretus*; sequences starting with "t-" refer to *t*-haplotypes. AFG, CZE and KAZ stand for *M. m. musculus* from Afghanistan, the Czech Republic and Kazakhstan, respectively, GER and FRA for *M. m. domesticus* from Germany and France, respectively, CAST stands for *M. m.* castaneus, and SPRE for *M. spretus*. (C) Percentage of 5 Kb regions in chromosome 17 region 50-90 Mb that support one subspecies being the outgroup to the other two. Grey bars represent the phylogeny of non-*t*-carriers, and green bars represent the phylogeny of *t*-carriers after retaining only private SNPs. (D) Percentage of 5 Kb regions in the non-recombined regions of the t complex that support one subspecies being the outgroup to the other two. Grey bars represent the phylogeny of non-*t*-carriers, and green bars represent the phylogeny of *t*-carriers after retaining only private SNPs.

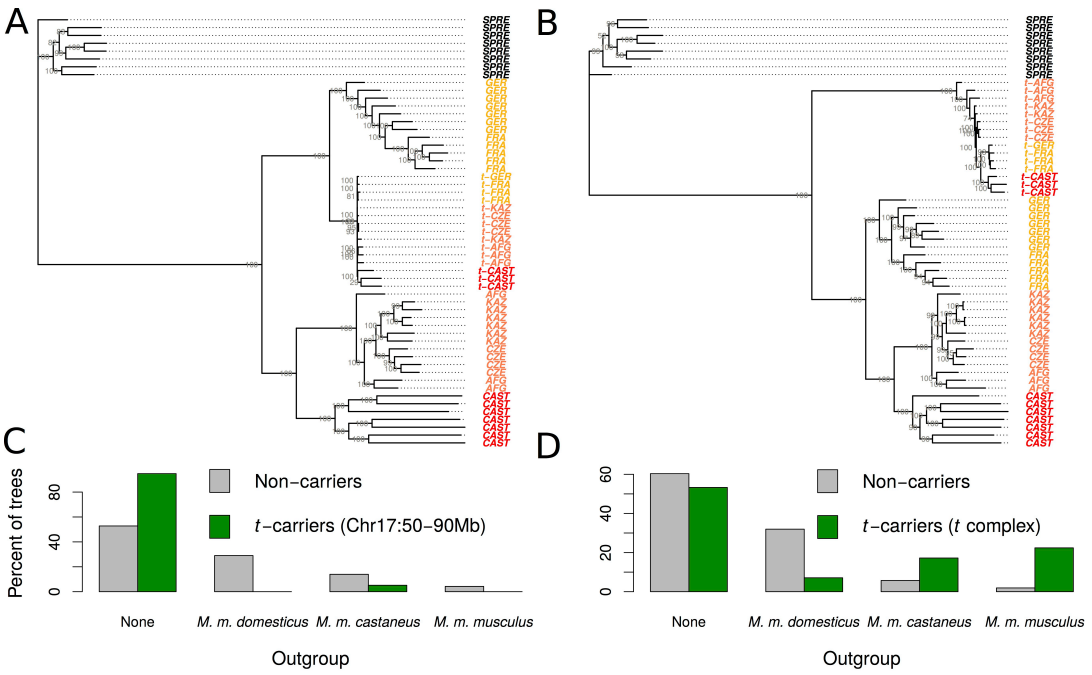


**Figure S9: Percentage of trees supporting each possible arrangement of subspecies in the *t*-haplotype and non-carrier phylogeny using the maximum parsimony (A) and neighbor joining (B) algorithm.** Percentage of trees based on 5 Kb windows of the region without recombination that show one subspecies as the outgroup to the others. Grey bars represent the phylogeny of non-*t*-carriers, and green bars represent the phylogeny of pseudo-*t*-haplotypes.

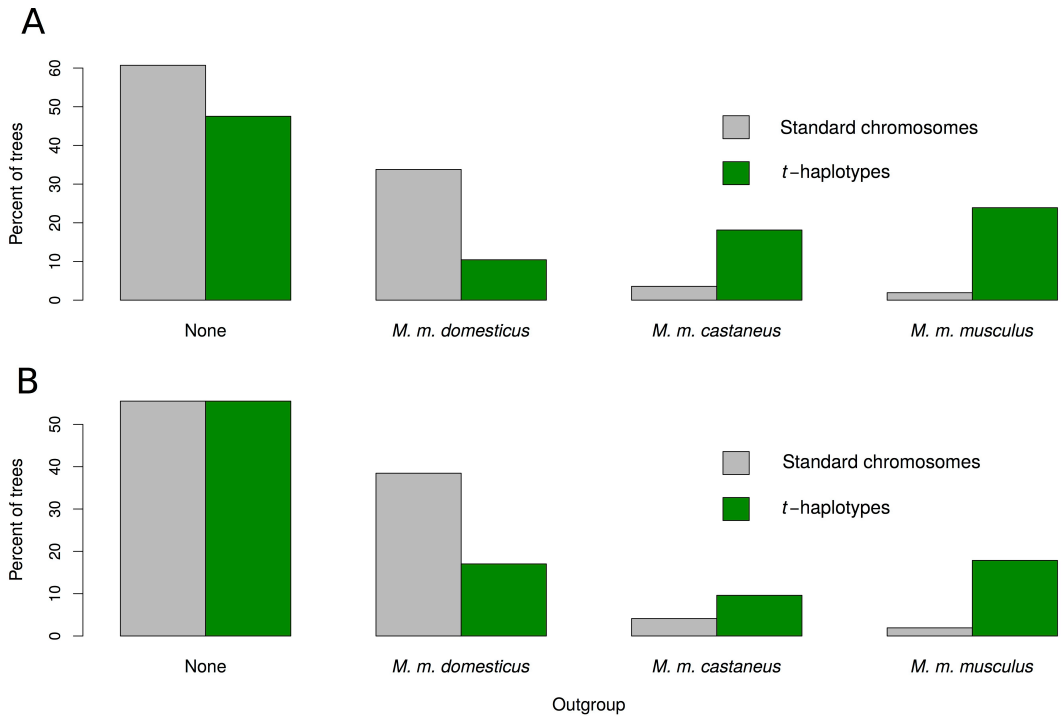


**Figure S10: Divergence of gene expression between *t*-carriers and non carriers.** (A-E) Percentage difference between the average gene expression of *t*-carrier and non-carriers (estimated as: | average_t-carrier - average_non-carrier | / average_non-carrier), plotted using a sliding window of 20 genes (using all genes with TPM>10 in non-carriers). Expression divergence is shown for (A) heart, (B) kidney, (C) muscle, (D) spleen, (E) thyroid. Regions that contain *t*-specific copy number variants (obtained by comparing the coverage of *t*-carriers to non-carriers, see Materials and Methods) are marked by grey rectangles. (F-J) Boxplots showing the percentage difference in expression of *t*-carriers relative to that of non-carriers for genes which overlap with at least 80% 5Kb windows for which no recombination was detected (green), some/old recombination was detected (orange) and recent/extensive recombination was detected (yellow), in (F) heart, (G) kidney, (H) muscle, (I) spleen, (J) thyroid. (K-O) Boxplots showing the percentage difference in expression of *t*-carriers relative to that of non-carriers for genes overlapping or not overlapping a CNV, in (K) heart, (L) kidney, (M) muscle, (N) spleen, (O) thyroid.

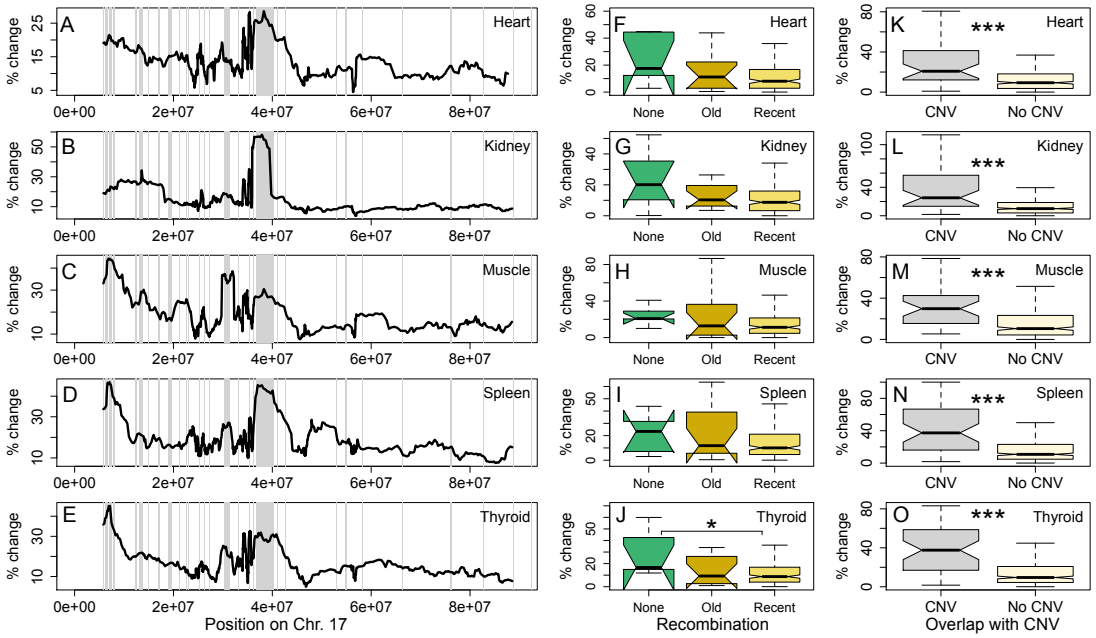


**Table S1:** Differentially expressed transcripts between *M. m. domesticus* wild type individuals and *t*-carriers in testis, heart, kidney, brain, thyroid, muscle, spleen and liver.

| **Brain** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000000579 | Dynlt1c | 17 | 36.6 | 81.7 | 0.04122 |
| ENSMUSG00000026269 | Rnpepl1 | 1 | 345.6 | 386.1 | 0.00861 |
| ENSMUSG00000036315 | Znrd1 | 17 | 94.1 | 158.9 | 0.04122 |
| ENSMUSG00000056692 | D17Wsu92e | 17 | 749.7 | 991.1 | 0.04122 |
| ENSMUSG00000071984 | Fndc1 | 17 | 240.2 | 76.3 | 0.00861 |
| ENSMUSG00000092074 | Dynlt1a | 17 | 36.6 | 81.7 | 0.04122 |
| ENSMUSG00000095677 | Dynlt1f | 17 | 36.6 | 81.7 | 0.04122 |

| **Heart** | | | | | |
|---|---|---|---|---|---|
| No significantly expressed genes found. | | | | | |

| **Kidney** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000023828 | Slc22a3 | 17 | 9.9 | 71.0 | 0.00129 |
| ENSMUSG00000024032 | Tff1 | 17 | 56.6 | 23.4 | 0.03007 |

| **Liver** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000079707 | Tcte3 | 17 | 6.7 | 40.1 | 0.04567 |

| **Spleen** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000033450 | Tagap | 17 | 305.6 | 793.1 | 0.01156 |

| **Testis** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000000579 | Dynlt1c | 17 | 1230.5 | 5557.4 | 6.76E-08 |
| ENSMUSG00000014956 | Ppp1cb | 5 | 281.3 | 2679.9 | 6.61E-08 |
| ENSMUSG00000023828 | Slc22a3 | 17 | 8.2 | 108.1 | 3.86E-06 |
| ENSMUSG00000029265 | Dr1 | 5 | 531.2 | 802.1 | 0.00115 |
| ENSMUSG00000036214 | Znrd1as | 17 | 782.0 | 3106.3 | 6.61E-08 |
| ENSMUSG00000040188 | Scamp2 | 9 | 1285.4 | 920.4 | 0.00140 |
| ENSMUSG00000046711 | Hmga1 | 17 | 384.4 | 587.7 | 0.00102 |
| ENSMUSG00000055602 | Tcp10b | 17 | 1259.7 | 2275.7 | 0.01609 |
| ENSMUSG00000059030 | Olfr128 | 17 | 51.3 | 24.4 | 0.03348 |
| ENSMUSG00000068037 | Mas1 | 17 | 344.9 | 157.9 | 0.00171 |
| ENSMUSG00000092074 | Dynlt1a | 17 | 1230.5 | 5557.4 | 6.76E-08 |
| ENSMUSG00000095677 | Dynlt1f | 17 | 1230.5 | 5557.4 | 6.76E-08 |

| **Thyroid** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000026269 | Rnpepl1 | 1 | 464.9 | 578.5 | 0.00020 |

| **Muscle** | | | | | |
|---|---|---|---|---|---|
| **Gene_ID** | **Gene Name** | **Chrom.** | **Non-carrier Average TPM** | **t-carrier Average TPM** | **q-value** |
| ENSMUSG00000073471 | Rsph3a | 17 | 310.2 | 717.2 | 0.00886 |

**Description of the Supplementary Data**

All the Supplementary Data are available at:
http://dx.doi.org/10.15479/AT:ISTA:78

**##1. Content of folders:**
"1-PASS_SNPs for the first filtering"
"2-Coverage-and_AlleleRatio-Filtered_RAW_SNPs"
"3-Coverage-Filtered_RAW_SNPs"

For each of these three SNP filtering procedures, we provide:

**1-Trees_for_all_5kb_windows** : the trees obtained for each 5Kb window using Maximum Likelyhood (ML_trees.zip), Neighbor-Joining (NJ_trees.zip), and Maximum Parsimony (MP_trees.zip).

**2-Tree_topologies_for_all_5kb_windows** : The topologies obtained for each tree and species. The topology files (e.g. Tree_results_cas_sorted) contain the following collumns:

**1.** Location on the t-complex. This corresponds to the location on Chromosome 17 minus 5000000 (for instance, the first window, 0-5000, is located on chromosome 17 from 5000000 to 5050000).
**2.** The status of t-haplotypes relative to their subspecies: 1 = nested, 0 = outgroup.
**3.** The status of t-haplotypes relative to the M. musculus species complex: 1 = nested, 0 = outgroup.
**4.** The topological color of the tree: 0 = no recombination (dark green), 1 = old recombination (light green), 2 = recent recombination (orange).

**3-TreeFile_Concatenated_Non-Recombined_Regions** : the tree obtained (with IQ-Tree, Maximum Likelihood) from the concatenated non-recombining regions.

**4-Divergence_SNPcoordinates :** The coordinates of all SNPs used to produce Figure 1 and Figure S4. The folder "Main" contains the coordinates of only heterozygous SNPs (Figure 1). The lists of heterozygous SNP locations for each *M. m. domesticus* individual on chromosome 17 are contained in a separate file named according to the respective mouse ID (see Supplementary Methods in File S1). The folder named "Controls" contains further folders corresponding to the neutral heterozygosity figure (heterozygous SNPs that are either synonymous or intergenic; used for Figure S4B), the pseudo-t vs. *M. spretus* figure (all SNPs found in pseudo-t-haplotypes and in *M. spretus* individuals; used for Figure S4C and E), and to the neutral pseudo-*t* vs *M. spretus* figure (heterozygous SNPs found in pseudo-*t*-haplotypes and *M. spretus* individuals, which are either synonymous or intergenic; used for Figure S4D and E). Each file contains a list of SNP coordinates for chromosome 17 for the individual indicated in the name of the file.

**5-Deterioration_SNPcoordinates :** The coordinates of all homozygous and heterozygous SNPs that are either missense or synonymous (used to plot Figure

4). The folder DOM_NonT contains SNP locations for non-carrier individuals from *M. m. domesticus*, the folder DOM_T contains SNP locations for pseudo-*t*-haplotypes *M. m. domesticus* individuals, while the folder SPRET contains SNP locations for *M. spretus* individuals. Each file is named in the following way: individual_ID.heterozygous/homozygous.missense/synonymous and contains the list of coordinates in the region chr17:5-40 Mb.

**##2. Content of folder "4-Expression"**
All the final gene expression values, as well as the corresponding q-values, are provided in the folder 4-Expression (one file per tissue).

**##3. Content of folder "5-CNV-regions"**
A list of all the identified CNVs is provided in the folder 5-CNV-regions.

CHAPTER 3

# Novel patterns of expression and recruitment of new genes on the *t*-haplotype, a mouse selfish chromosome

Réka K. Kelemen[a]*, Marwan Elkrewi[a], Anna K. Lindholm[b] and Beatriz Vicoso[a]*

[a] *Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

[b] *Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

***Corresponding authors**: rkelemen@ist.ac.at, beatriz.vicoso@ist.ac.at

# Abstract

The *t*-haplotype of mice is a classical model for autosomal transmission distortion. A largely non-recombining variant of the proximal region of chromosome 17, it is transmitted to >90% of the progeny of heterozygous males through the disabling of sperm carrying a standard chromosome. While extensive genetic and functional work has shed light on individual genes involved in drive, much less is known about the evolution and function of the rest of its hundreds of genes. Here, we characterize the sequence and expression of dozens of *t*-specific transcripts and of their chromosome 17 homologs. Many genes showed reduced expression of the t-allele, but an equal number of genes showed increased expression of their *t*-copy, consistent with increased activity or a newly evolved function. Genes on the *t*-haplotype had a significantly higher nonsynonymous substitution rate than their homologs on the standard chromosome, with several genes harboring dN/dS ratios above 1. Finally, the *t*-haplotype has acquired at least two genes from other chromosomes, which show high and tissue-specific expression. These results provide a first overview of the gene content of this selfish element, and support a more dynamic evolutionary scenario than expected of a large genomic region with almost no recombination.


## 3.1 Introduction

Genetic variants that increase their own transmission rate during gametogenesis will spread in the population even if neutral or detrimental with respect to the fitness of the organism [47]. Such transmission distorters, or meiotic drivers, have been found in diverse taxa, including plants, animals and fungi [41, 57]. While true meiotic drivers increase their transmission rate by manipulating female meiosis, so called "sperm killers" do so by using a poison-antidote system (the "driver" and "responder" genes) to disable sperm not carrying the driver chromosome [35]. Since recombination between the driver and responder genes leads to the creation of suicide chromosomes (that disable all sperm), sperm killers are typically found in regions of no or very low recombination that can harbor large numbers of genes. There has been considerable progress in identifying specific genes underlying the driving mechanisms of different distorters [50, 62, 38, 48, 46, 23, 14, 36, 16], but much less is known about how the rest of the gene content of these selfish haplotypes differs from that of their homologous (non-driving) genomic region, and what evolutionary pressures contributed to these changes [14, 38, 10]. Positive selection will favor mutations that enhance drive, especially if drive-suppressing mutations arise elsewhere in the genome [2]. Such evolutionary arms-races can promote the evolution of increasingly complex driving mechanisms involving multiple genes that are co-opted to increase transmission rate [31]. For this reason, many genes linked to the original driving locus may become "neofunctionalized", i.e. repurposed for segregation distortion. For instance, cooption for drive has been suggested to contribute to the differential expression of large numbers of genes in the testis of stalk eyed flies carrying a driving X-chromosome [54]. On the other hand, transmission distorters often bear the negative consequences of strong linkage between the driver and responder genes [61]. Reduced recombination between the driving region and its homologous chromosome is often achieved by large inversions, which may trap hundreds of other genes on the driving haplotype [18, 24, 51, 27]. These genes are expected to be subject to less efficient purifying selection, which may be compounded if deleterious mutations hitch-hike when new driver mutations sweep to fixation. Genetic degeneration has therefore typically been thought to be the prevalent force shaping gene content on large drivers [10, 18, 61], although occasional recombination with the non-driving homolog may alleviate

this mutation load [49, 32].

One of the best studied autosomal drivers is the *t*-haplotype of house mice, which has served as a model for segregation distortion for nearly one hundred years [17, 42]. The *t*-haplotype is a sperm killer that achieves above 90 percent transmission in heterozygous ($+/t$) males, but causes embryonic lethality or adult sterility when present in two copies. A variant form of the proximal half of chromosome 17 thought to have originated more than a million years ago [44, 25], it contains four large inversions that link together a region of about 900 genes. Only a few of the genes on the *t*-haplotype have been functionally and evolutionarily characterized, most of these directly related to the driving mechanism. Four genes (*Tagap1, Fgd2, Nme3 and Tiam2*) have been found to cumulatively distort the transmission ratio [5, 4, 3, 14], by jointly dysregulating a single target (*Smok1*). The *t*-haplotype codes for an insensitive version of the target (*Tcr*), avoiding the sperm toxicity of *Smok1* overexpression [28]. The fate of the other hundreds of genes originally located on the *t*-haplotype is largely unknown. The drive pathway still has some missing links, and it is thought that the *t*-haplotype likely contains more genes involved in transmission ratio distortion [14], but how many is currently unclear. Interestingly, some of the most differentially expressed genes between carriers and non-carriers of this transmission distorter are on other chromosomes [32, 40], but the mechanism underlying this expression upregulation is unknown. Finally, homozygous $t/t$ mice typically die as embryos, as most variants of the *t*-haplotype contain recessive lethal mutations [60], but it is unclear whether these are due to widespread degeneration of the whole non-recombining region. While limited evidence of genetic degeneration was detected, this was likely an underestimate, as it was based on short read mapping to the reference, due to the absence of an assembly for the *t*-haplotype [32].

In order to adress some of these gaps, we combined published RNA and DNA sequencing data to characterize the sequence and expression evolution of dozens of genes on the *t*-haplotype, and compared their expression and patterns of divergence to those of their homologous chromosome 17 genes. We also describe two highly expressed *t*-specific genes, which were gained from other chromosomes. These results highlight the dynamic evolution of this non-recombining selfish chromosome, at odds with a simple scenario of reduced purifying selection that is expected for a large low recombination region, and potentially suggesting that significant sections of the genome may be co-opted for transmission distortion.

## 3.2 Results

### Most putative *t*-specific sequences map to chromosome 17

We used published RNA-seq reads obtained from four wild-caught *M. m. domesticus* $+/t$ mice (mice heterozygous for the *t*-haplotype; [26]) to infer the sequence of genes on the *t*-haplotype. Since these mice also carry one copy of the non-driving chromosome 17, we used three complementary approaches to filter for reads and/or for assembled transcripts that are likely to be *t*-specific (see Supplementary figure 3.13): (1) We mapped all RNA-seq reads of $+/t$ individuals to the *M. musculus* reference genome, and retained only diverged read pairs (reads with a minimum of three mismatches). We assembled these into transcripts. To detect true *t*-derived sequences, we mapped genomic reads (also from [26]) from 12 $+/t$ (*t*-carriers) and 12 $+/+$ (non-carriers) mice to the assembled transcripts (with no mismatches allowed to avoid cross-mapping with the $+$ allele, see Methods), and selected scaffolds that had a higher genomic coverage (normalized for library size) in all $+/t$ mice than in $+/+$ mice.

(2) We identified kmers of size 31 that were found in all the RNA and DNA samples of $+/t$ mice, but in none of the DNA or RNA samples from $+/+$ mice, yielding a set of putative *t*-specific kmers. We then selected RNA-seq read pairs from $+/t$ samples that contained these *t*-specific 31-mers, and assembled them directly into putative *t*-derived transcripts. (3) To complement the assemblies based on pre-filtered reads, we also created an assembly based on all the combined RNA-seq derived from all tissues of the four $+/t$ mice. The assembled sequences were again filtered based on genomic coverage in 12 $+/t$ and 12 $+/+$ control mice. Since this last assembly does not require that reads or transcripts are divergent from the reference, it may include young *t*-specific duplicates.

Transcripts were mapped to the mouse reference genome and transcriptome, and annotated based on which genes they overlapped with (see Methods). More than 90% of our annotated transcripts map to chromosome 17 genes for all three assemblies (Figure 3.1A), supporting a low false positive rate. 3% of all assembled *t*-specific sequences did not map to the mouse reference genome or transcriptome at all (Supplementary Table 1). 45 assembled genes are found by at least two assemblies, while 66 genes are detected by a single assembly, showing that the different approaches complement each other well. We find a higher proportion of the genes in the first three inversions of the *t*-haplotype than in the fourth inversion (39 to 65% versus 5%, $p<0.001$ with a Fisher's exact test, Figure 3.1B). The fourth inversion is a large paracentric inversion thought to be younger than the second inversion [25], and where *t*-haplotypes are a mosaic of the $+$ and *t*-specific sequences, indicative of recombination events [20, 32]. The greater level of divergence between the *t* and the standard chromosome in the proximal half of the *t* complex likely gave us more power to assemble *t*-specific transcripts from this region.

## Decreased and increased expression of *t*-specific alleles are equally common

We investigated patterns of expression of *t*-derived transcripts in eight tissues obtained from four $+/t$ mice and four $+/+$ mice of the subspecies *M. m. domesticus* [26] (see Supplementary figure 3.14A). We used Kallisto [9], a software suitable for inferring allele-specific expression, to estimate transcript abundance of both putative *t*-transcripts and of their chromosome 17 homologs. We tested our power to infer *t*-specific expression by simulating reads from the sequence of both the *t* and the $+$ alleles, and re-estimating expression levels with the simulated reads. The simulated ratio of expression between the two homologs was recovered by Kallisto for all but one gene (*Mup9*), which we excluded from further analysis (see Supplementary Figure 3.5). Only transcripts that produced an alignment longer than 300 base pairs with a $+$ transcript in the *t* complex, and for which average expression was $>1$ Transcripts Per Million (TPM) for at least one tissue, were kept for further analysis (58 out of 111 putative *t*-specific genes).

In order to understand changes in gene expression that have arisen specifically on the *t*-haplotype, we compared the expression of *t* transcripts to the expression of the $+$ allele in $+/+$ mice. As a control, we also compared the expression of the $+$ allele between $+/t$ and $+/+$ mice. The (misassigned) expression level of the *t* allele in $+/+$ mice was used to correct the TPM of the $+$ and *t* alleles in $+/t$ mice (see Methods).

Overall, the *t* allele deviated significantly in expression for 51% of the tissue comparisons, while the $+$ allele deviated only for 14% of such comparisons ($p<0.0001$, Fisher's exact test). We classified each *t* allele into one of three categories based on its expression: (1)

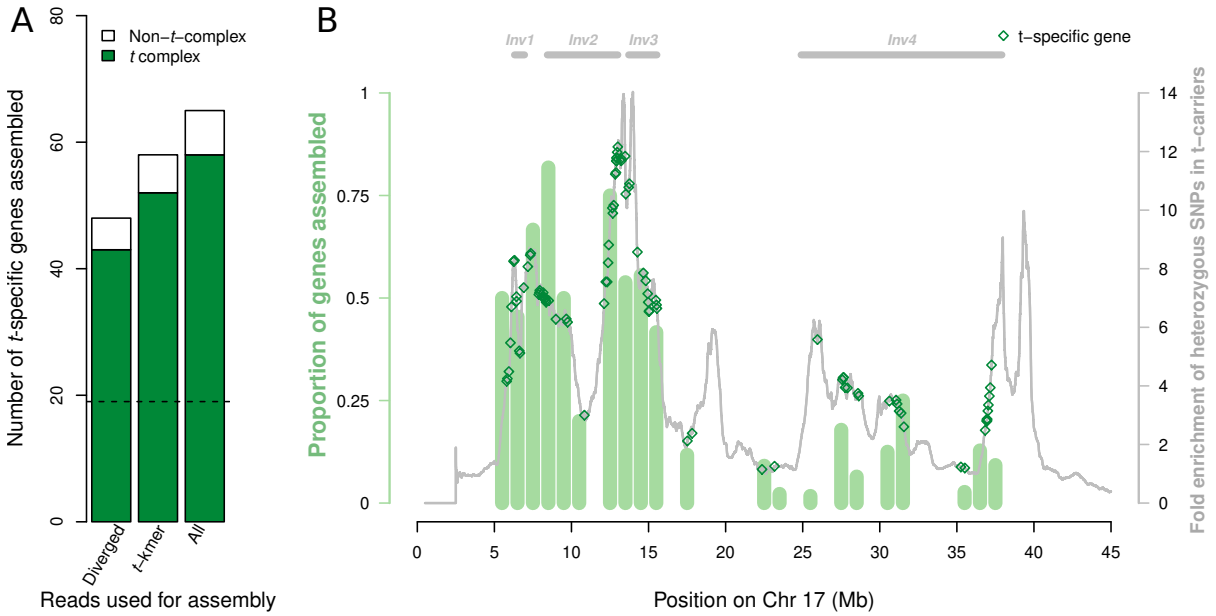



Figure 3.1: Chromosomal locations of assembled *t*-specific transcripts. (A) Numbers of genes for which *t*-specific transcripts were assembled using the three assembly strategies. The proportion of genes mapping to the *t* complex (3-42 Mb on chromosome 17) is shown in green, while those mapping elsewhere in the genome are in white. The dashed line indicates the number of genes present in all assemblies. (B) Proportion and location of genes assembled along the *t* complex. Light green bars indicate the proportions of genes in 1 Mb windows, for which a *t*-haplotype-specific sequence was assembled. The grey line shows the average excess heterozygosity of *M. m. domesticus* $+/t$ mice compared to $+/+$ mice, adapted from [32]. The locations of *t*-specific genes are shown as green empty diamonds, so mapping genes can be better visualized. The locations of the four inversions along the *t* complex, based on the coordinates of genes confirmed to be in each, are shown on top of the figure.

conserved expression, if there was no significant difference (with a Wilcoxon test) between the expression of the *t* allele and the + allele in any tissue. (2) decreased expression, if the *t* allele had a significantly lower expression compared to the + allele in at least one tissue, and was conserved otherwise. (3) increased expression, if the *t* allele had a significantly higher expression compared to the + allele in at least one tissue, which might be a sign of increased activity or a newly acquired function in the tissue(s). While 25 genes were underexpressed on the *t*-haplotype (left side of Figure 3.2), another 25 genes were overexpressed in at least one tissue on the *t*-haplotype (right side of Figure 3.2). 8 genes, shown in the middle of Figure 3.2, have conserved expression of the *t* allele in all tissues where the gene is expressed. Applying no correction for the fraction of TPM misassigned between alleles changed the classification of only one gene in the degeneration group and one gene in the conservation group (Supplementary Figure 3.6). Comparing the *t* allele's expression against the + allele's expression within $+/t$ mice changed the classification of 14 individual genes, but led to similar patterns of over- versus underexpression (Supplementary Figure 3.7).

We detected no dependence between the overexpression of + alleles and the underexpression of *t* alleles (p=0.08, binomial test, Supplementary Figure 3.8), indicating that our allele-specific expression estimation is not systematically biased towards one allele. Genes overexpressed on the *t*-haplotype are enriched for copy gain events (taken from [32]) that are either unshared or shared among the four $+/t$ mice when compared to genes in the decreased expression (Fisher's

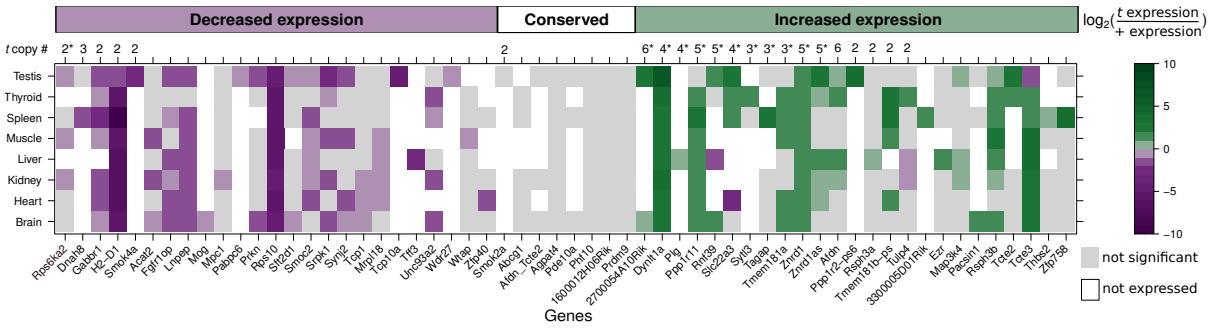



Figure 3.2: Relative expression of *t*-specific alleles in the *t* complex. Color coding shows the log2 ratio of the average expression of the *t* allele in $+/t$ mice to that of the $+$ allele in $+/+$ mice. Non-significant differences in expression are colored grey, while tissues with no expression (on average $< 1$ TPM for both alleles) are white. Genes whose *t* allele is underexpressed (purple) in at least one tissue and do not show overexpression (green) in any tissue are shown on the left. Genes with conserved *t* allele expression in all tissues with expression are shown in the middle. Genes whose *t* allele is overexpressed in at least one tissue are shown on the right. The numbers of gene copy gain detected in 4 $+/t$ vs. 4 $+/+$ *M .m. domesticus* DNA samples are indicated on top of the figure, with asterisks denoting fixed copy gain among the four $+/t$ mice.

exact test, p=0.004 and p=0.002, respectively) or conservation (Fisher's exact test, p=0.02 and p=0.03, respectively) groups, suggesting that gene amplification may be an important mechanism through which expression is modulated [32]. Varying polymorphism levels among the gained copies suggest different ages of amplification (Supplementary Figure 3.9).

## Faster accumulation of nonsynonymous mutations on the *t*-haplotype

Both reduced purifying selection and increased positive selection can lead to increases in the rate of protein coding divergence. We therefore tested if *t*-haplotype genes show an elevated ratio of nonsynonymous to synonymous substitution rates (dN/dS) when compared with their freely recombining homologs. We selected *t*-specific genes that had a coding sequence alignment of at least 100 base pairs with the $+$ allele, and that had annotated homologs in the sister species, the Algerian mouse (*M. spretus*) and the distant relative, the rat (*R. norvegicus*) (35 genes, see Supplementary figure 3.14B). We concatenated alignments of these genes, excluding four genes that had premature STOP codons (underlined genes names in Figure 3.3), and estimated dN/dS for all branches using the software PAML [67]. The branch leading to the *t*-haplotype had a dN/dS two-fold higher than the rest of the branches ($p < 0.0001$, Likelihood-ratio-test of models with branch-specific or shared dN/dS values), generally consistent with what is observed in regions of low recombination due to the accumulation of deleterious mutations [61, 15, 59].

We also plotted dN/dS values for the *t*-haplotype and *M. musculus* ($+$) branches for all individual genes that had a rate of synonymous substitution (dS) greater than 0 in both lineages (Figure 3.3). Six genes show a significantly higher dN/dS on the *t*-haplotype than on the $+$ allele. Two genes, *Ppp1r11* and *Tcte3* have dN/dS values significantly higher than one, suggesting that these genes may have undergone positive selection after becoming part of the *t*-haplotype.

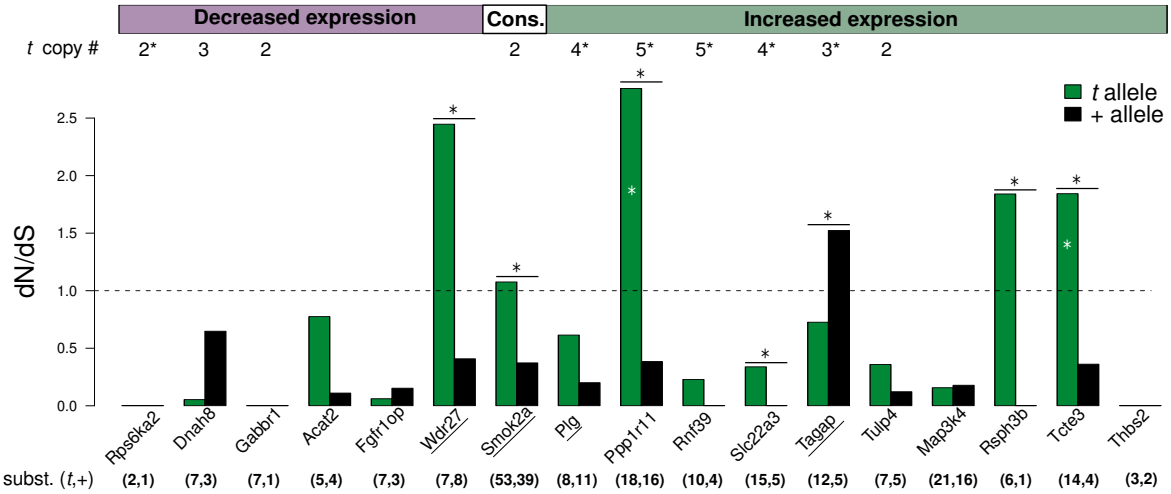



Figure 3.3: Ratios of nonsynonymous to synonymous substitution rates (dN/dS) of *t* (green) and + (black) alleles of 16 *t* complex genes. Only genes with a coding sequence alignment of at least 100 base pairs and dS>0 on both the + and *t* lineages were included. Black stars on top of the bars mean that dN/dS is significantly different between the *t* and + alleles, and white stars indicate dN/dS values significantly higher than 1 (using likelihood ratio tests, see Methods). The estimated mean copy number gained by 4 +/*t* *M .m. domesticus* mice is indicated on top of the figure, with asterisks denoting fixed copy gain among the four +/*t* mice. Underlined genes have a premature STOP codon in their *t* alleles. The numbers of substitutions in the *t* and + alleles are shown in parentheses on the bottom. The boxes on top of the figure indicate the *t* allele's expression pattern (Figure 3.2).

## The *t*-haplotype expresses modified copies of genes gained from other chromosomes

Our set of candidate *t*-specific sequences included copies of eight genes, which are located outside of chromosome 17 (one gene each from chromosomes 1, 2, 4, 5, 6, 15 and three genes from chromosome 16). The majority has very low absolute expression, or low expression relative to the parental copy (Supplementary Figure 3.11). However, two genes, *Rnpepl1* and *Ppp1cb* showed high expression, and had previously been found to be strongly overexpressed in *t*-carrier mice [32, 40]. It had been suggested that functional elements on the *t*-haplotype might be regulating these genes in *trans* [32, 40]. However, patterns of genomic read coverage of +/*t* and +/+ samples (Supplementary Figure 3.10) strongly support the presence of a copy of these genes on the *t*-haplotype itself. PCR amplification of these sequences yielded strong bands in all 10 +/*t* mice tested, and no or very faint bands in +/+ mice (Figure 3.4A, Supplementary Figure 3.12 and Supplementary Table 3.2), confirming the presence of a *t* copy of these genes.

*Rnpepl1* is overexpressed in the brains and thyroid glands, while *Ppp1cb* is overexpressed in the testes of *t*-carrier mice ([32, 40], Figure 3.4B). The current analysis shows that, for both genes, overexpression comes from the *t*-specific paralog, with the parental copy being expressed at similar levels in +/*t* and +/+ mice (Figure 3.4B). In the case of *Rnpepl1*, the *t*-haplotype expresses a nonsense-mediated decay copy of the gene, which contains only an 80-amino-acid-long truncated version of the protein. The *t*-specific paralog of *Ppp1cb* expresses a putative protein-coding transcript at a 10-fold higher level in the testis than the chromosome 5 paralog. Contrary to the original *Ppp1cb*, the *t*-specific paralog is not expressed in any other tissue. We aligned the *t*-specific *Ppp1cb* sequence to that of the paralog in *M. m.*

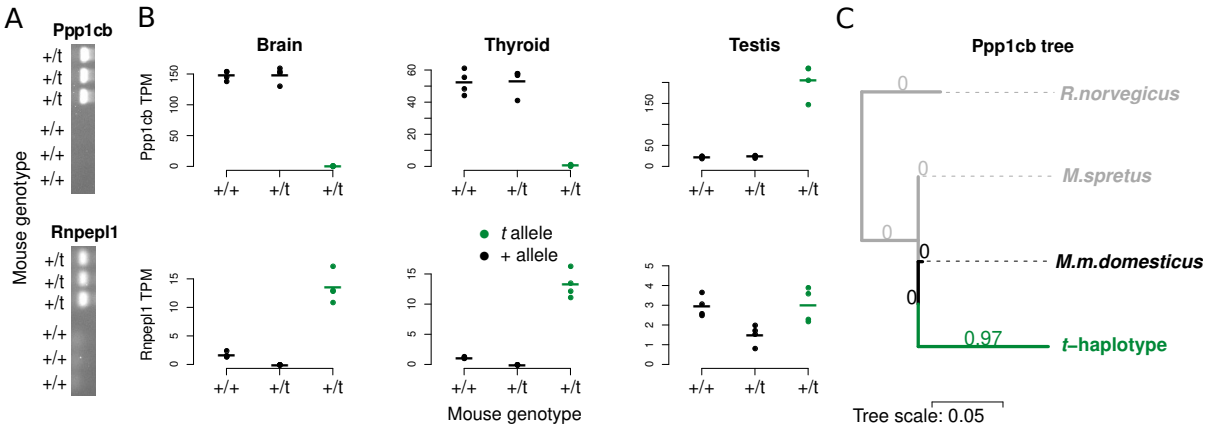



Figure 3.4: Presence, expression and sequence evolution of gained genes on the *t*-haplotype. (A) PCR bands showing the presence of the *t*-specific copies of *Ppp1cb* and *Rnpepl1* in 3 $+/t$ mice and their absence in 3 $+/+$ mice (for all 20 mice tested, see Supplementary Table 2). (B) Expression in the three tissues, where the gained genes are differentially expressed [32]. Dots show Transcripts Per Million (TPM) measured in individual mice, while the horizontal bars show the average of the four mice. Expression is shown in green for the *t*-specific copy and in black for the paralogs on the other chromosomes. (C) Phylogenetic tree estimated by PAML based on the sequence alignment of *Ppp1cb*. The ratio of nonsynonymous and synonymous substitution rates, dN/dS, was estimated for each branch separately, as this model was superior to one with shared dN/dS ($p<0.0001$, likelihood ratio test). dN/dS values are shown above each branch.

*domesticus* and the orthologs in *M. spretus* and *R. norvegicus*, and estimated nonsynonymous and synonymous substitution rates using PAML. While *Ppp1cb* is generally highly conserved, without a single nonsynonymous mutation detected on any of the non-*t* lineages, the paralog on the *t*-haplotype differs by 20 nonsynonymous substitutions, resulting in a dN/dS of 0.97 (Figure 3.4C).

## 3.3 Discussion

The *t*-haplotype has been a model for meiotic drive for nearly a century. While a lot is known about the molecular mechanism and the key genes used for achieving drive, studying the entire sequence of the *t*-haplotype has not yet been possible. Here we performed a partial characterization of the gene content of the *t*-haplotype by assembling *t*-specific transcripts from RNA-seq reads, and assessing their expression and sequence evolution. Of the 878 genes of the *t* complex, we assembled 111 genes. Since only data from $+/t$ mice was available, we were limited to regions of the *t*-haplotype that were differentiated in sequence from the homologous chromosome 17 regions and/or duplicated on the *t*-haplotype, thus yielding increased genomic read coverage in $+/t$ mice compared to $+/+$ mice. The average divergence of assembled *t*-specific sequences was 0.022. Due to our genomic-coverage-based selection of *t*-specific sequences involving samples from three subspecies we miss genes specific to certain *t* variants. Although we likely underestimate degeneration by missing unexpressed or deleted genes, copy number estimation [32] showed that there are only four genes in the *t* complex that overlap a deletion fixed among *M. m. domesticus* $+/t$ mice. Furthermore, significant underexpression in $+/t$ mice compared to $+/+$ individuals affects only a minority of *t* complex genes (4 and 77 genes, in [32] and [40], respectively). Although gene expression buffering

/ dosage compensation from the standard chromosome could mask the underexpression of degenerated genes on the *t*-haplotype, we find no correlation between the underexpression of *t* alleles and the overexpression of + alleles in this study. This suggests that degeneration due to deletions or lack of expression from *t* alleles is fairly limited. On the other hand, since only a minority of chromosome 17 genes are differentially expressed between $+/t$ and $+/+$ individuals (21 and 195 in [32] and [40], respectively), and there are signs of widespread recombination between the *t*-haplotype and the standard chromosome 17 [32], a large proportion of the genes on the *t*-haplotype are likely undifferentiated in both sequence and expression (and therefore missed here).

While our study is restricted to diverged genes, it provides a first overview of the dynamic evolution of the gene content and expression of this large transmission distorter. 43% of the *t*-specific genes in our expression analysis are overexpressed in at least one tissue when compared to the + allele. While the accumulation of neutral or deleterious mutations in regulatory regions could also lead to increased expression [21], this raises the interesting possibility that some may have acquired new functions since becoming part of the *t*-haplotype. Although no functional enrichment can be found after correcting for multiple comparisons, several of these genes with upregulated expression have a functional annotation related to plasma membrane bounded cell projection, such as sperm flagellum, cilium, microvillus and microspike (7 out of 25 genes, p=0.0018 without correcting for multiple testing, see Supplementary Table 3), making it plausible that they are involved in drive. However, differential expression of *t*-specific genes is not limited to the testis, and it is possible that some of these differentially expressed genes may give the *t*-haplotype a selective advantage without direct involvement in sperm function. For example, $+/t$ mice show behavioral differences compared to $+/+$ mice, such as increased aggression in males [39] or higher likelihood to disperse from their populations [55, 56], both of which have been hypothesized to facilitate the spread of this transmission distorter.

Our results also show that, contrary to a model of simple degeneration, selfish elements can gain genes from other chromosomes, similar to the gain of genes by nonrecombining Y chromosomes [12, 13]. While functional studies are needed to infer the role of the new copies of *Rnpepl1* and *Ppp1cb*, their high and tissue-specific expression suggests a possible contribution to the biology of the *t*-haplotype. The overexpression of the *t*-specific paralog of *Rnpepl1*, an aminopeptidase, in the brain makes it an interesting candidate for the behavioral differences associated with $+/t$ mice, but the lack of a substantial open reading frame supports at most a regulatory function. On the other hand, the *t*-specific paralog of the protein phosphatase *Ppp1cb* shows signs of very fast protein evolution and is highly and exclusively expressed in the testes of *t*-carriers. Protein phosphatase 1 complexes are important for spermatogenesis, with one of the active forms suppressing sperm motility in the epididymis [65, 66]. It is therefore possible that the new copy of *Ppp1cb* is involved in the drive exhibited by the *t*-haplotype. The fact that two other *t*-complex PPP1-related genes (*Ppp1r11* and *Ppp1r2ps6*) show highly increased expression of their *t*-derived transcripts in the testis, and that *Ppp1r11*'s rate of nonsynonymous substitution is suggestive of positive selection, provides further support for the role of these proteins in the biology of the *t*-haplotype.

Genome and transcriptome assemblies of large transmission distorters coupled with allele-specific expression and sequence evolution analysis have the prospect of showing how degenerate selfish haplotypes are and of uncovering driver-specific functionality [61]. Future genomic assemblies that include the entire *t*-haplotype will reveal the full extent of conservation and divergence in sequence and expression on this classic model for transmission distortion.

## 3.4 Methods

For a detailed description of the methods and scripts see the Supplementary methods. Pipelines are shown in Supplementary figures 3.13 and 3.14.

### Assembling diverged reads

We pooled RNA-seq reads from ten tissues sampled from four *M. m. domesticus* mice heterozygous for the *t*-haplotype ([26], https://www.ebi.ac.uk/ena/browser/view/PRJEB9450). We trimmed the first and last five base pairs off of every read using a custom perl script. Trimmomatic [8] (version 0.38, with parameters LEADING:20 TRAILING:20 SLIDINGWINDOW:4:25 MINLEN:36) was used to remove bases with quality below 20 at the beginning and end of reads, windows of 4 base pairs with an average base quality below 25, and Illumina adapters. Reads shorter than 36 base pairs after trimming were removed. To select diverged reads, we mapped trimmed RNA-seq reads to the GRCm38.p6 genome using Tophat [63] (version 2.1.1 with default settings). Reads with more than two mismatches were unmapped, and all paired unmapped reads were assembled into scaffolds using Trinity [22] (version 2.12.0 with default parameters).

### Assembling reads with *t*-specific kmers

We used genomic libraries of four $+/t$ and four $+/+$ *M. m. domesticus* mice as well as transcriptomic libraries from these mice with up to ten tissues pooled per mouse ([26], https://www.ebi.ac.uk/ena/browser/view/PRJEB9450). Following [19], we used the script kcompress.sh in the software BBMap [11] to output the unique 31 base pair k-mers in each of the four $+/t$ genomic libraries and each of the four $+/t$ RNA-seq libraries. We found 31-mers shared between all $+/t$ 31-mer sets, by setting the mincount parameter to 8 in the script kmercountexact.sh. We then removed any 31-mer present in any of the four genomic or RNA-seq libraries of the $+/+$ control mice using bbduk.sh. We recovered RNA-seq reads from *t*-carrier libraries that overlapped in at least 30% of their lengths a *t*-carrier specific k-mer, by setting the "minkmerfraction" parameter to 0.3 in bbduk.sh. The recovered reads from the four *t*-carrier mice were pooled and assembled using Trinity, as before.

### Assembling unfiltered reads

Pooled, untrimmed and unfiltered RNA-seq reads from up to ten tissues of four *M. m. domesticus* $+/t$ mice, were assembled into scaffolds with the software Trinity (default parameters).

### Filtering based on genomic reads

We masked repetitive sequences in our assembled sequences with RepeatMasker [58] (using the combined database Dfam 3.1 and rmblastn version 2.10.0+), and filtered for a minimum unmasked length of 300 base pairs. We mapped the first read in each pair of genomic reads in 12 carrier and 12 non-carrier samples to the sequences with Bowtie2 [37] (version 2.3.4.1 with default parameters). We filtered for a higher abundance of perfectly matching reads (normalized for library size) in all $+/t$samples than in $+/+$ samples.

## Annotation of assembled sequences

We mapped RepeatMasker-masked sequences against the GRCm38.p6 genome and transcriptome using BLAT [33] (version 35x1 with parameters -t=dnax -q=dnax). Sequences that overlapped multiple neighboring genes were further examined based on coding sequence overlap and assigned to a single gene whenever possible (see Supplementary methods).

## Gene-specific re-assembly and sequence selection

We grouped sequences by gene annotation from the divergence-based and kmer-based assemblies together and from the unfiltered-reads-based assembly separately, and re-assembled scaffolds into longer sequences using the software Cap3 [30] (version 02/10/15, with a maximum overhang of 80% and requiring at least 40% overlap of at least one scaffold).

## Expression estimation

We aligned *t* sequences to GRC38.p6 transcripts using BLAT (version 35x1 with parameters -t=dnax -q=dnax), and for each gene we retained the longest alignments (minimum 300 base pairs). The assembly of unfiltered reads was only used when genes were not found in the other assemblies. For all other genes, the longest transcripts were included. We used RNA-seq libraries from four $+/t$ and four $+/+$ *M. m. domesticus* mice obtained from 8 tissues [26] (Supplementary figure 3.13A). We trimmed reads using Trimmomatic, and estimated expression levels of *t* and + transcripts from each sample using the software Kallisto [9] (version 0.46.2 with default parameters). Transcript abundance estimates were normalized by library size. Genes with average expression below 1 TPM in all individuals for both the *t* and + transcripts were removed from the analysis.

## Correcting for ambiguity in read assignment

In R [53] (version 3.6.3) we calculated the proportion of ambiguity in $+/+$ samples by dividing the average TPM mis-assigned to the *t* allele by the average total TPM assigned to that gene. In each sample we subtracted this proportion from both the *t*-transcript's and the + transcript's TPM values.

## Testing for differential expression

For each gene and tissue we used a Wilcoxon signed rank test (in R) on the four corrected expression values of the *t* transcript in $+/t$ mice and the four corrected expression values of the + transcript in $+/+$ mice divided by two.

## Simulating reads for testing the expression estimation of Kallisto

Using the software ART [29] (version 2.5.8) we generated Illumina Hiseq 2000 paired-end reads (91 base pairs, standard deviation of 10, fragment size of 180 base pairs, mimicking our real reads) from all the *t* and + transcripts that were included in our expression analysis. Expression estimation was the same as on the real dataset.

## Coding sequence (CDS) alignments

We aligned each *t* transcript to the + peptide sequences of the corresponding gene using the software GeneWise [6] (version 2.4.1 with default settings), and retained the translated *t* peptide with the longest alignment, if it was longer than 100 base pairs. We used the *t* peptide sequence to align the *M. musculus* + transcript, as well as the *R. norvegicus* and *M. spretus* orthologous transcripts (obtained from the ensembl database BioMart [34] (release 104)) to it using GeneWise. For genes with orthologs in both species CDS alignments were made using TranslatorX [1] (version 1.1 with default settings).

## Estimating dN/dS

We used the *codeml* function of PAML [67] (version 4.9j) to estimate dN/dS from alignments. We used the species tree as the input tree (see Supplementary methods). To test if the total dN/dS on the *t*-haplotype is larger than that on other lineages we compared a null model of shared dN/dS among all lineages ($model = 0$) and an alternative model of only the *t*-haplotype having its own dN/dS value ($model = 2$ and a distinct branch label on the input tree). To test if a single gene has different dN/dS values on the *t*-haplotype and on the + chromosome, we compared a null model of shared dN/dS of these two lineages and an alternative model of distinct dN/dS values. To test if a gene has a dN/dS value above 1 the null model was the site-branch model with $\omega_2$ fixed at 1 ($model = 2, NSites = 2, fixomega = 1, omega = 1$), and the alternative model was the full site-branch model ($model = 2, NSites = 2, fixomega = 0, omega = 2$).

## Statistical comparison of different PAML models

We extracted log likelihood (lnL) estimates of PAML, and calculated the Akaike Information Criteria (AIC) score for each model using the formula 2k-2lnL, where k is the number of parameters (dN/dS values estimated) in a model. AIC score differences above 2 units were considered to be significant.

## Finding genes overlapping copy number variant (CNV) regions in *t*-carrier mice

We used CNVs called by the software Control-FREEC [7] (version 10.5 with parameter window=1000 or 5000) for the four *M. m. domesticus* $+/t$ mice and a pool of four $+/+$ mice as controls (same as in [32]). With BEDTools' *intersect* function [52] we found genes overlapping CNVs, and we averaged the estimated copy number inferred in the 1- and 5-Kb windows for each gene in R.

## Primer design and PCR

We designed two primer pairs each for *Ppp1cb* and *Rnpepl1* (see Supplementary methods) that contained *t*-specific mutations at their 3' ends, using the software Primer3 and its default settings [64] (version 0.4.0). The primers were tested on an independent set of *M. m. domesticus* (see [43] for population details; study design and sampling procedures were approved by the Veterinary Office, Zurich Switzerland (permit 215/2006)). All mice were genotyped using the *Hba4-ps4* and *Vil2* primers, which produce bands of different sizes in $+/+$ and $+/t$ mice. We first ran the PCR with the first set of primers per gene, on only three

$+/+$ and three $+/t$ mice (shown in Figure 3.4 and Supplementary figure 3.12). To confirm these results, we conducted PCR on another 10 $+/+$ and another 10 $+/t$ mice (summarized in Supplementary Table 2). We isolated DNA using saltchloroform extraction [45]. We used PCR conditions of 94 degrees for 7 minutes, and 32 cycles of 94 degrees for 30s, 60 degrees for 60s, and 72 degrees for 120 s and then a 20 min extension at 72 degrees. We ran the samples on a 1% agarose gel. We analysed PCR products using a 3730xl DNA Analyzer (Applied Biosystems) and Genemapper software (Applied Biosystems).

### Gene ontology (GO) enrichment analysis

We used the MouseMine website with default settings and no test correction to find enrichment in the "cellular component" ontology.

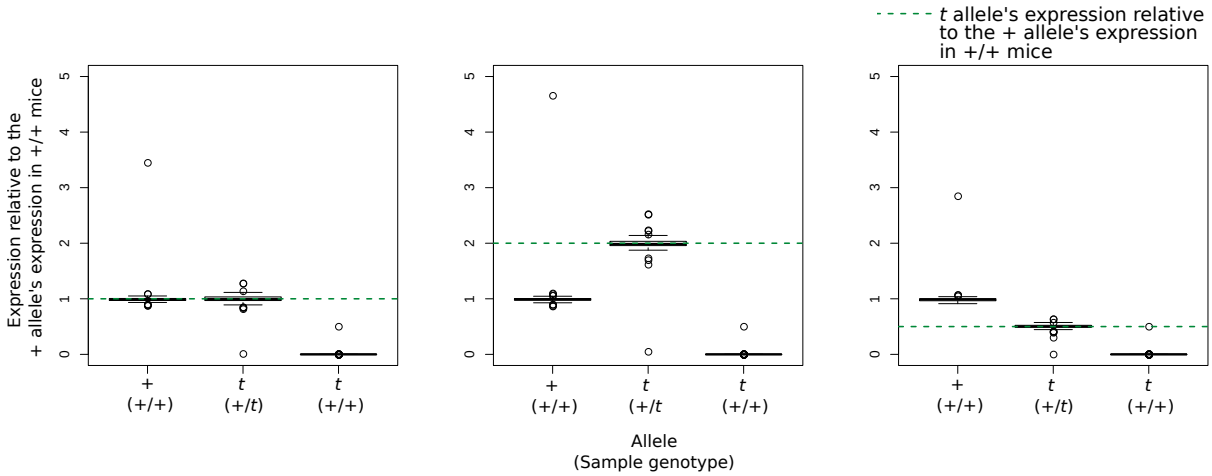
## Acknowledgments

This project has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number 715257) and from the Swiss National Science Foundation (grant number 310030_189145).

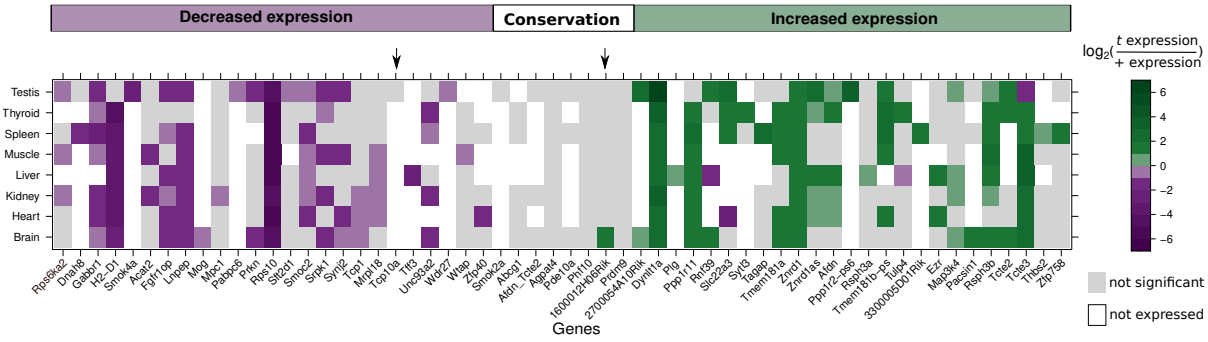
We thank Jari Garbely of the Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland, for conducting the PCR verification.

Barbara Konig, Gabi Stichel and AKL collected mouse tissue samples, from the field study led by BK.

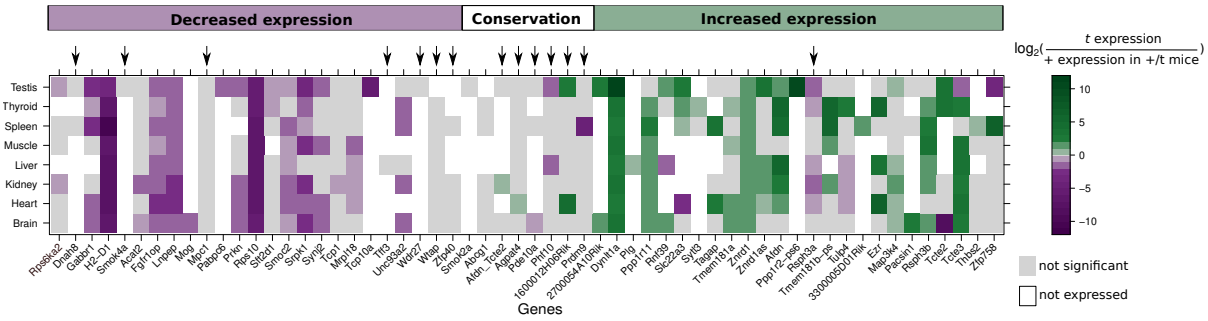
## 3.5 Supplementary figures



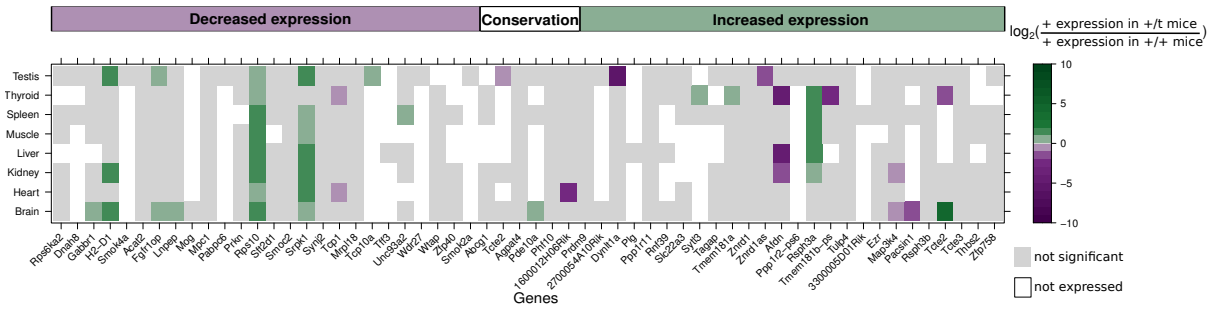

Supplementary Figure 3.5: Expression estimation of simulated reads. Allelic expression relative to the $+$ allele in $+/+$ samples was estimated using Kallisto. Read files were generated in three simulation experiments. (A) We generated a coverage of 20 for $t$-specific transcripts and a coverage of 20 for the $+$ transcripts. (B) We simulated $t$-underexpression by decreasing the read coverage of the $t$ transcript to 10, while keeping the $+$ alleles coverage at 20. (C) we simulated overexpression of the $t$ alleles by generating a coverage of 40 for them and an unchanged coverage of 20 for the $+$ alleles.



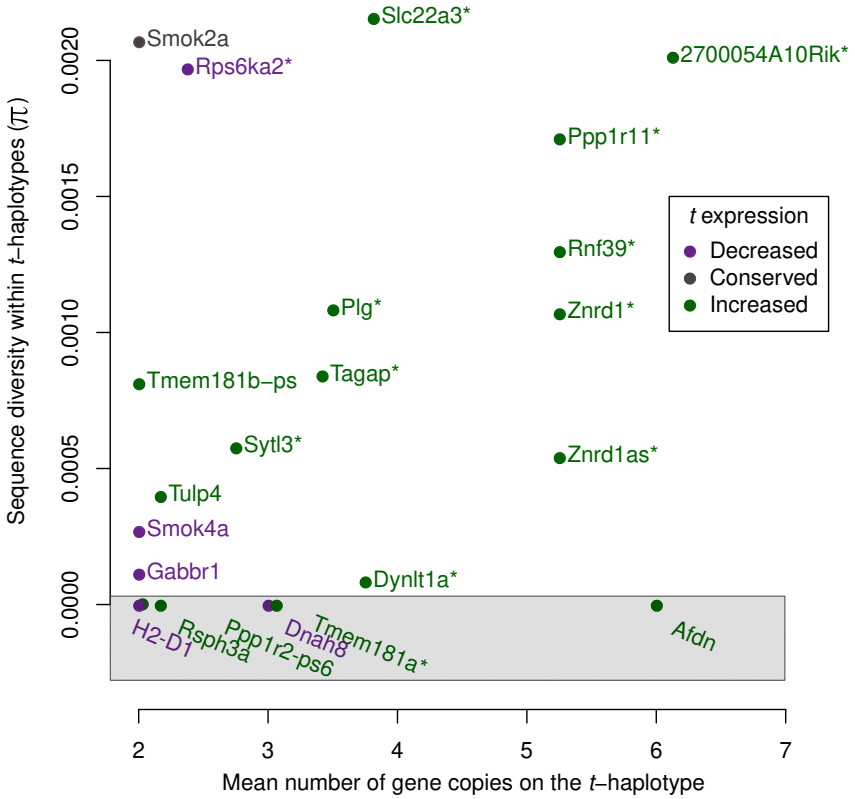

Supplementary Figure 3.6: Relative expression of $t$-specific alleles in the $t$ complex without correcting for the fraction of TPM misassigned between alleles. Expression was estimated in units of transcripts per million (TPM) with the software Kallisto. Color coding shows the log2 ratio of the average expression of the $t$ allele in $+/t$ mice to that of the $+$ allele in $+/+$ mice. Non-significant differences in expression are colored grey, while tissues with no expression (on average $< 1$ TPM for both alleles) are white. Genes with no expression in any tissue are not shown. Gene expression categories, indicated as decreased expression, conservation and increased expression, are based on the TPM values corrected for misassignment between alleles, shown in Figure 2. Arrows on the top show genes that change category when misassignment correction is not performed.



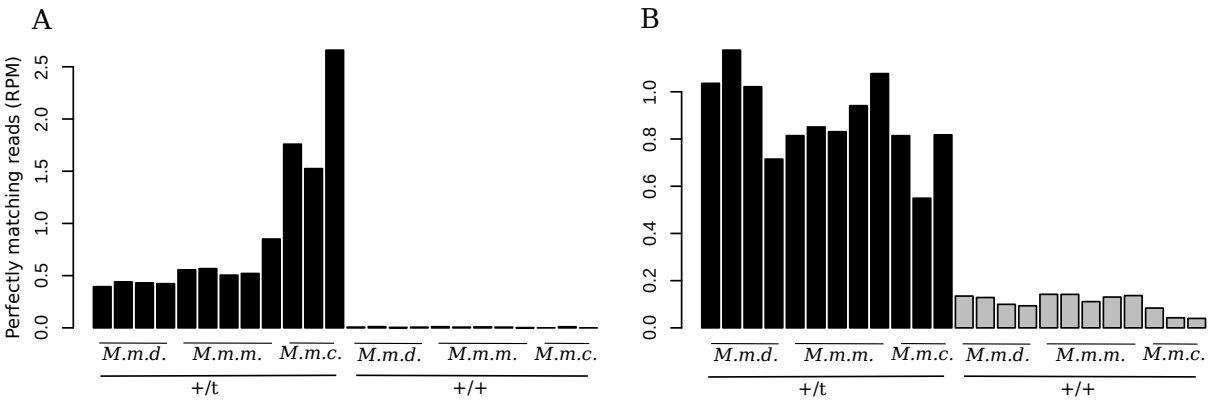

Supplementary Figure 3.7: Expression of $t$-specific alleles in the $t$ complex normalized by the expression of the $+$ alleles in $+/t$ mice. Expression was estimated in units of transcripts per million (TPM) with the software Kallisto. Color coding shows the log2 ratio of the average expression of the $t$ allele in $+/t$ mice to that of the $+$ allele in $+/t$ mice. Non-significant differences in expression are colored grey, while tissues with no expression (on average $< 1$ TPM for both alleles) are white. Genes with no expression in any tissue are not shown. Gene expression categories, indicated as decreased expression, conservation and increased expression, are based on the $t$ expression relative to the $+$ allele in $+/+$ mice, shown in Figure 2. Arrows on the top show genes that change category when the $t$ alleles expression is normalized by the $+$ alleles expression in $+/t$ mice.



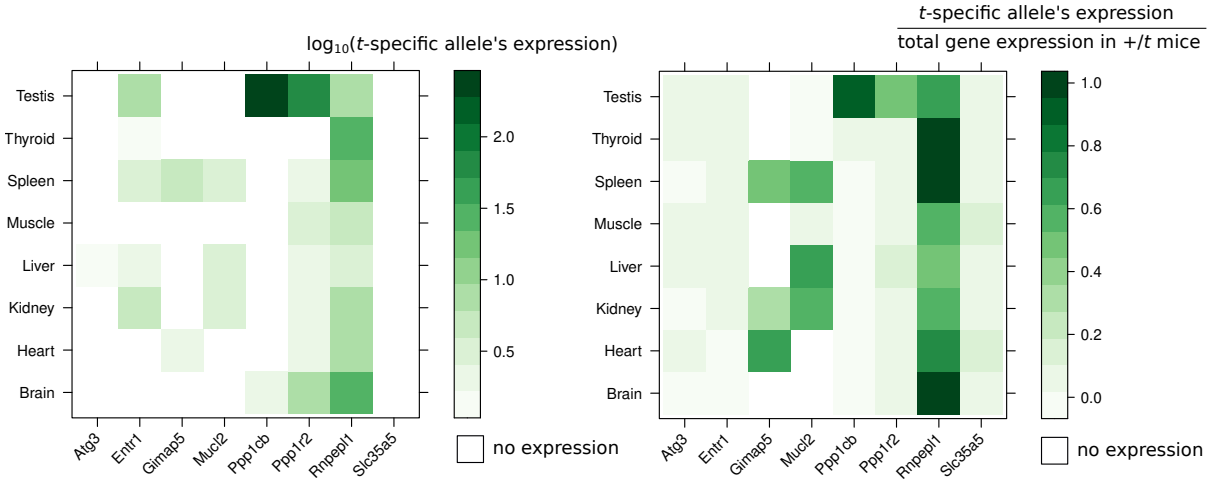

Supplementary Figure 3.8: Relative expression of + alleles on chromosome 17 measured in $+/t$ mice. Expression was estimated in units of transcripts per million (TPM) with the software Kallisto. Color coding shows the log2 ratio of the average expression of the + allele in $+/t$ mice to that of the + allele in $+/+$ mice. Non-significant differences in expression are colored grey, while tissues with no expression ($< 1$ TPM for both alleles) are white. Genes with no expression in any tissue are not included in this plot. Expression patterns observed for the $t$ allele (decreased expression, conservation and increased expression) are shown above the genes.



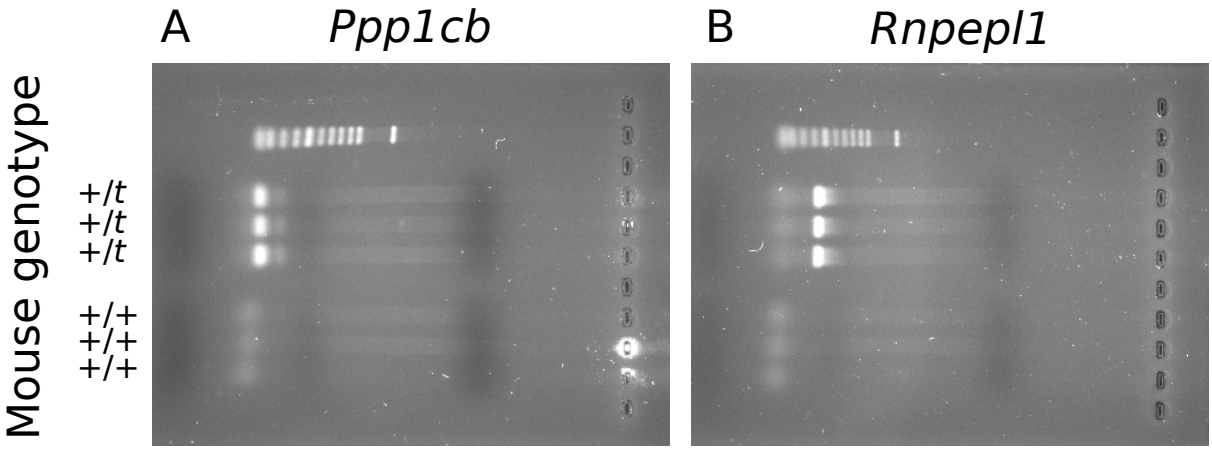

Supplementary Figure 3.9: Copy number and polymorphism among amplified genes on the $t$-haplotype. The x axis shows the estimated mean number of gene copies on the $t$-haplotype, among four *M. m. domesticus* $+/t$ mice, and the y axis indicates the nucleotide diversity ($\pi$) in these genes among $t$-haplotypes. Purple, grey and green denote genes with decreased, conserved and increased expression of the $t$ allele, respectively. Genes that showed copy number gain in all four $+/t$ samples have an asterisk after their names. Genes in the grey box on the bottom had no polymorphism among the four $+/t$ samples.



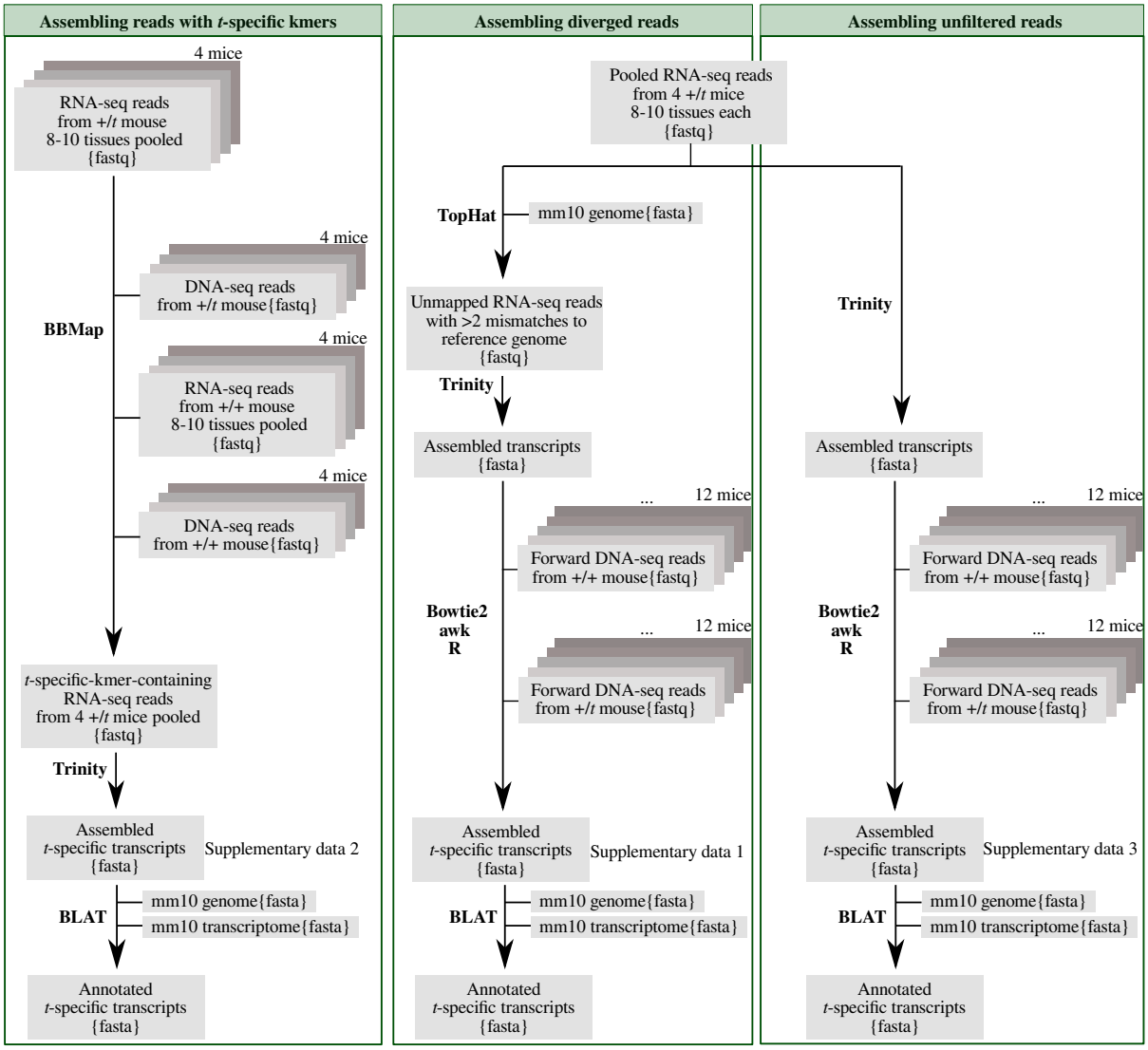

Supplementary Figure 3.10: Genomic Coverage of gained genes on the *t*-haplotype. Genomic reads from 12 $+/t$ (black bars) and 12 $+/+$ (grey bars) mouse samples from all three subspecies (indicated under the bars as *M. m. d.*, *M. m. m.* and *M. m. c.* for *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus*, respectively) of house mice were mapped against *t*-specific (A) *Ppp1cb* and (B) *Rnpepl1* sequences. Perfectly matching reads were counted and normalized by library size and shown as reads per million (RPM).



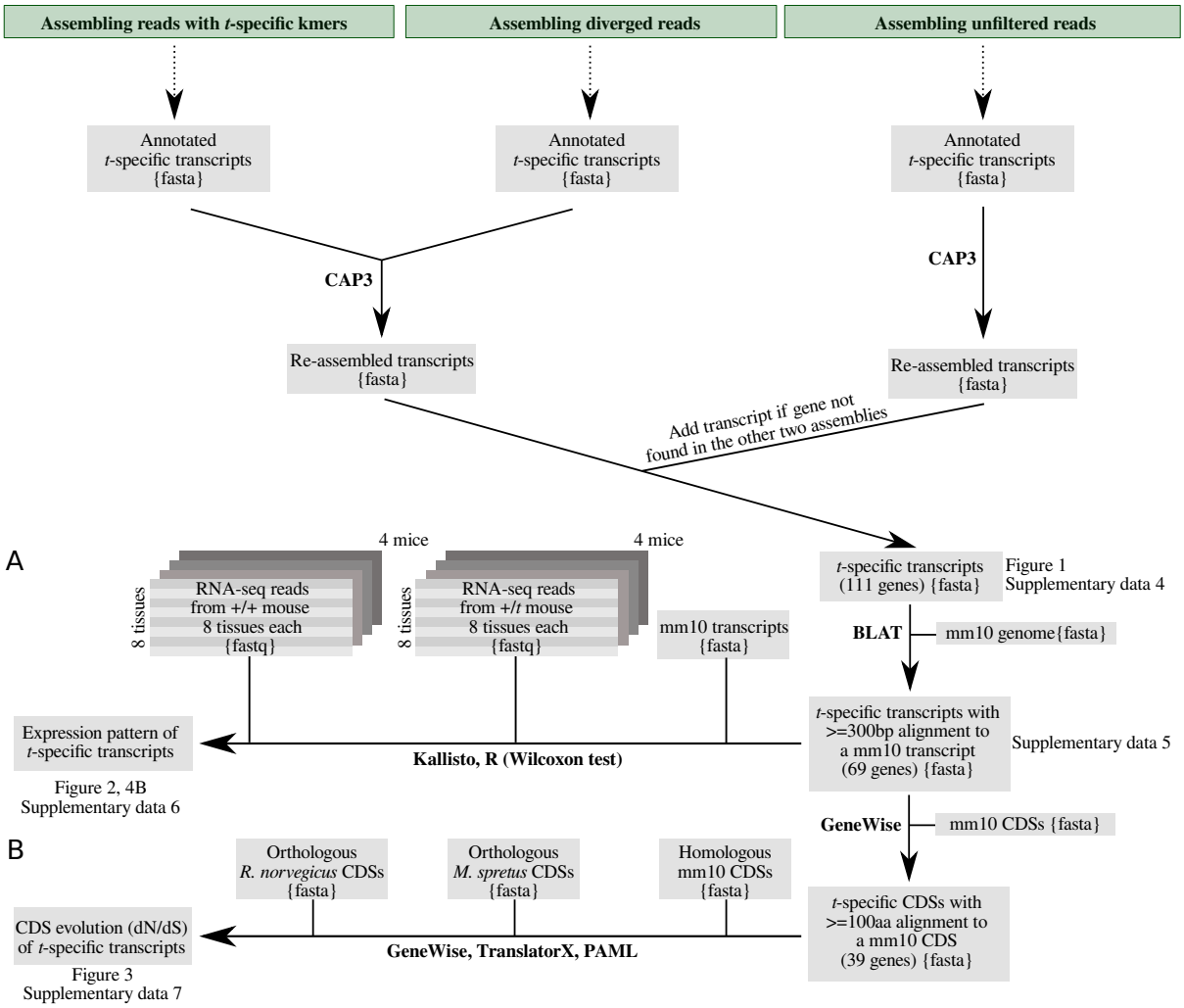

Supplementary Figure 3.11: Absolute and relative expression of the gained genes on the *t*-haplotype. (A) Expression heatmap of the *t*-specific copies of eight non-chromosome-17 genes shown on a log10 scale. (B) Expression heatmap of the *t*-specific copies of eight non-chromosome-17 genes relative to the total expression of that gene in $+/t$ mice. Allelic expression was estimated in units of transcripts per million with Kallisto for each of the four mouse samples and the average expression is shown.




Supplementary Figure 3.12: PCR gels amplifying the gained genes on the *t*-haplotype. The top three runs were based on three $+/t$ mice and the bottom three runs represent samples from three $+/+$ mice. Panel (A) shows the *t*-specific copy of *Ppp1cb* and panel (B) shows the *t*-specific copy of *Rnpepl1*. The primers were tested on an independent set of *M. m. domesticus* (see [43] for population details).




Supplementary Figure 3.13: Assembly pipelines.




Supplementary Figure 3.14: Pipelines for the expression (A) and sequence evolution (B) analyses.

## 3.6 Supplementary tables

| Assembled reads | Mapping to chr. 17 | Mapping to other chr. | Unmapped |
|---|---|---|---|
| Diverged | 135 | 15 | 7 |
| *t*-specific-kmer-containing | 187 | 15 | 16 |
| All | 390 | 12 | 2 |

Supplementary Table 3.1: Number of assembled transcripts in each assembly that map to chromosome 17, map to a different chromosome or do not map at all to the mm10 reference genome or transcriptome

| Mouse ID | Genotype | Hba-ps4 | *Hba-ps4* | *Vil2* | *Vil2* | *Ppp1cb*,E3 | *Ppp1cb*,E3 | *Ppp1cb*,E1 | *Ppp1cb*,E1 | *Rnpepl1*,3UTR | *Rnpepl1*,3UTR | *Rnpepl1*,3UTRfus | *Rnpepl1*,3UTRfus |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1466 | +/+ | 197 | 197 | 230 | 230 | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1467 | +/+ | 197 | 197 | 230 | 230 | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1468 | +/+ | 197 | 197 | 230 | 230 | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1473 | +/+ | 197 | 197 | 230 | 230 | no amp. | no amp. | 112,small | 138,small | 432,small | 432,small | 233,small | 233,small |
| 1474 | +/+ | 197 | 197 | not done | not done | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1475 | +/+ | 197 | 197 | not done | not done | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1476 | +/+ | 197 | 197 | not done | not done | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1478 | +/+ | 197 | 197 | not done | not done | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1482 | +/+ | 197 | 197 | 230 | 230 | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. | no amp. |
| 1496 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1465 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1471 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1479 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1480 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1481 | +/t | 197 | 213 | 198 | 230 | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1485 | +/t | 197 | 213 | not done | not done | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1487 | +/t | 197 | 213 | not done | not done | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1488 | +/t | 197 | 213 | not done | not done | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |
| 1490 | +/t | 197 | 213 | not done | not done | 180 | 180 | 112 | 138 | 432 | 432 | 233 | 233 |

Supplementary Table 3.2: PCR verification of *Ppp1cb* and *Rnpepl1* using 10 $+/+$ and 10 $+/t$ mice. Numbers under gene names represent PCR product sizes in basepairs. Primers designed for the genes *Hba-ps4* and *Vil2* produce PCR products with different lengths from the $+$ and $t$ chromosomes, and were used for genotyping mice. We designed two pairs of primers each for the $t$-specific copies of the gained genes, *Ppp1cb* and *Rnpepl1*, which show no amplification in 9 out of 10 $+/+$ mice. The $+/+$ mouse with a small amplification of *Ppp1cb, E1* and the *Rnpepl1* primers shows 5-10% of peak hight seen in $+t$ mice. Note: some $++$ samples had much weaker amplification that it could not clearly be distinguished from background (here considered no amplification = no amp.), nonetheless there was some signal. For other $++$ there was no signal at all.

| GO category | p-value | Gene IDs | GO ID |
|---|---|---|---|
| extrinsic component of membrane | 0.000117226368082 | MGI:1933367,MGI:97620,MGI:98642,MGI:98931 | GO:0019898 |
| cytoplasmic side of apical plasma membrane | 0.00163879828158 | MGI:98931 | GO:0098592 |
| cilium | 0.002081472182049 | MGI:1914082,MGI:3630308,MGI:98642,MGI:98931 | GO:0005929 |
| Schwann cell microvillus | 0.002457252922534 | MGI:98931 | GO:009745 |
| microspike | 0.00327507849915 | MGI:98931 | GO:0044393 |
| intraciliary transport particle | 0.005120324387421 | MGI:1914082,MGI:3630308,MGI:98642 | GO:0030990 |
| ruffle | 0.00606085910658 | MGI:1345181,MGI:98931 | GO:0001726 |
| extrinsic component of external side of plasma membrane | 0.007354786201754 | MGI:97620 | GO:0031232 |
| cell tip | 0.007354786201754 | MGI:98931 | GO:0051286 |
| uropod | 0.008982281604408 | MGI:98931 | GO:0001931 |
| cell trailing edge | 0.008982281604408 | MGI:98931 | GO:0031254 |
| plasma membrane bounded cell projection | 0.009560005006826 | MGI:1314653,MGI:1345181,MGI:1914082,MGI:3630308,MGI:98642,MGI:98931 | GO:0120025 |
| cell pole | 0.009795090887488 | MGI:98931 | GO:0060187 |
| photoreceptor ribbon synapse | 0.009795090887488 | MGI:1345181 | GO:0098684 |
| extrinsic component of plasma membrane | 0.010444684261381 | MGI:1933367,MGI:97620 | GO:0019897 |
| RNA polymerase I complex | 0.010607275157022 | MGI:1913386 | GO:0005736 |
| presynaptic endocytic zone | 0.010607275157022 | MGI:1345181 | GO:0098833 |
| myelin sheath | 0.01367661611357 | MGI:1345181,MGI:98931 | GO:0043209 |
| cell projection | 0.014900258007201 | MGI:1314653,MGI:1345181,MGI:1914082,MGI:3630308,MGI:98642,MGI:98931 | GO:0042995 |
| plasma membrane region | 0.01511353907613 | MGI:1314653,MGI:1345181,MGI:1914082,MGI:3630308,MGI:98642,MGI:98931 | GO:0098590 |
| platelet alpha granule | 0.015467280589857 | MGI:98738 | GO:0031091 |
| microvillus membrane | 0.017082302593648 | MGI:98931 | GO:0031528 |
| astrocyte projection | 0.01788888189196 | MGI:98931 | GO:0097449 |
| ribbon synapse | 0.01788888189196 | MGI:1345181 | GO:0097470 |
| synaptic ribbon | 0.01788888189196 | MGI:1345181 | GO:0098681 |
| COPI-coated vesicle | 0.020304898322542 | MGI:1345181 | GO:0030137 |
| cytoplasmic dynein complex | 0.021108998128987 | MGI:98642 | GO:0005868 |
| pore complex | 0.021108998128987 | MGI:1314653 | GO:0046930 |
| cell projection membrane | 0.024360482176328 | MGI:1345181,MGI:98931 | GO:0031253 |
| glial cell projection | 0.029913355437332 | MGI:98931 | GO:0097386 |
| nucleolus | 0.032297611163055 | MGI:1913386,MGI:2385044,MGI:98931 | GO:0005730 |
| immunological synapse | 0.036269835175556 | MGI:98931 | GO:0001772 |
| plasma lipoprotein particle | 0.037061643876243 | MGI:97620 | GO:0034358 |
| lipoprotein particle | 0.037061643876243 | MGI:97620 | GO:1990777 |
| protein-lipid complex | 0.03943341216333 | MGI:97620 | GO:0032994 |
| collagen-containing extracellular matrix | 0.039539867370968 | MGI:97620,MGI:98738 | GO:0062023 |
| cell leading edge | 0.041781978144078 | MGI:1345181,MGI:98931 | GO:0031252 |

Supplementary Table 3.3: Gene Ontology enrichment in the cellular component category of genes with increased expression of the *t*-specific allele.

## 3.7 Bibliography


[1] F. Abascal, R. Zardoya, and M. J. Telford. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic acids research*, 38(suppl_2):W7–W13, 2010.

[2] H. Bastide, P. R. G?rard, D. Ogereau, M. Cazemajor, and C. Montchamp-Moreau. Local dynamics of a fast-evolving sex-ratio system in Drosophila simulans. *Mol Ecol*, 22(21):5352–5367, Nov 2013.

[3] H. Bauer, S. Schindler, Y. Charron, J. Willert, B. Kusecek, and B. G. Herrmann. The nucleoside diphosphate kinase gene Nme3 acts as quantitative trait locus promoting non-Mendelian inheritance. *PLoS Genet.*, 8(3):e1002567, 2012.

[4] H. Bauer, N. Véron, J. Willert, and B. G. Herrmann. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes Dev*, 21(2):143–147, Jan 2007.

[5] H. Bauer, J. Willert, B. Koschorz, and B. G. Herrmann. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nat Genet*, 37(9):969–973, Sep 2005.

[6] E. Birney, M. Clamp, and R. Durbin. GeneWise and genomewise. *Genome research*, 14(5):988–995, 2004.

[7] V. Boeva, T. Popova, K. Bleakley, P. Chiche, J. Cappo, G. Schleiermacher, I. Janoueix-Lerosey, O. Delattre, and E. Barillot. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics*, 28(3):423–425, 2012.

[8] A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014.

[9] N. L. Bray, H. Pimentel, P. Melsted, and L. Pachter. Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5):525–527, 2016.

[10] A. Burt and R. Trivers. *Genes in Conflict*. Harvard University Press, Cambridge, MA and London, England, 2009.

[11] B. Bushnell. BBMap: a fast, accurate, splice-aware aligner. Technical report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2014.

[12] A. B. Carvalho, B. A. Dobo, M. D. Vibranovski, and A. G. Clark. Identification of five new genes on the Y chromosome of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 98(23):13225–13230, 2001.

[13] A. B. Carvalho, B. Vicoso, C. A. Russo, B. Swenor, and A. G. Clark. Birth of a new gene on the Y chromosome of Drosophila melanogaster. *Proceedings of the National Academy of Sciences*, 112(40):12450–12455, 2015.

[14] Y. Charron, J. Willert, B. Lipkowitz, B. Kusecek, B. G. Herrmann, and H. Bauer. Two isoforms of the RAC-specific guanine nucleotide exchange factor TIAM2 act oppositely on transmission ratio distortion by the mouse t-haplotype. *PLoS Genet*, 15(2):e1007964, 02 2019.

[15] S. J. Christianson, C. L. Brand, and G. S. Wilkinson. Reduced polymorphism associated with X chromosome meiotic drive in the stalk-eyed fly Teleopsis dalmanni. *PLoS One*, 6(11):e27254, 2011.

[16] C. Courret, C.-H. Chang, K. H.-C. Wei, C. Montchamp-Moreau, and A. M. Larracuente. Meiotic drive mechanisms: lessons from Drosophila. *Proceedings of the Royal Society B*, 286(1913):20191430, 2019.

[17] N. Dobrovolskaia-Zavadskaia and N. Kobozieff. Sur la reproduction des souris anoures. *Comptes Rendus Séances Société de Biologie et de ses Filiales*, 97:116–119, 1927.

[18] K. A. Dyer, B. Charlesworth, and J. Jaenike. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl. Acad. Sci. U.S.A.*, 104(5):1587–1592, Jan 2007.

[19] M. Elkrewi, M. A. Moldovan, M. A. Picard, and B. Vicoso. Schistosome W-linked genes inform temporal dynamics of sex chromosome evolution and suggest candidate for sex determination. *Molecular Biology and Evolution*, 2021.

[20] M. Erhart, S. Phillips, and J. Nadeau. Contrasting patterns of evolution in the proximal and distal regions of the mouse t complex. *Genetics of Immunological Diseases*, pages 70–76, 1988.

[21] E. C. Glassberg, Z. Gao, A. Harpak, X. Lant, and J. K. Pritchard. Measurement of selective constraint on human gene expression. *BioRxiv*, page 345801, 2018.

[22] M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature biotechnology*, 29(7):644, 2011.

[23] P. Grognet, H. Lalucque, F. Malagnac, and P. Silar. Genes that bias Mendelian segregation. *PLoS Genetics*, 10(5):e1004387, 2014.

[24] M. F. Hammer, J. Schimenti, and L. M. Silver. Evolution of mouse chromosome 17 and the origin of inversions associated with t haplotypes. *Proceedings of the National Academy of Sciences*, 86(9):3261–3265, 1989.

[25] M. F. Hammer and L. Silver. Phylogenetic analysis of the alpha-globin pseudogene-4 (Hba-ps4) locus in the house mouse species complex reveals a stepwise evolution of t haplotypes. *Molecular biology and evolution*, 10(5):971–1001, 1993.

[26] B. Harr, E. Karakoc, R. Neme, M. Teschke, C. Pfeifle, . Pezer, H. Babiker, M. Linnenbrink, I. Montero, R. Scavetta, M. R. Abai, M. P. Molins, M. Schlegel, R. G. Ulrich, J. Altmuller, M. Franitza, A. Buntge, S. Kunzel, and D. Tautz. Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. *Sci Data*, 3:160075, Sep 2016.

[27] E. Hauschteck-Jungen and B. Maurer. Sperm dysfunction in sex ratio males of Drosophila subobscura. *Genetica*, 46(4):459–477, 1976.

[28] B. G. Herrmann, B. Koschorz, K. Wertz, K. J. McLaughlin, and A. Kispert. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature*, 402(6758):141–146, Nov 1999.

[29] W. Huang, L. Li, J. R. Myers, and G. T. Marth. ART: a next-generation sequencing read simulator. *Bioinformatics*, 28(4):593–594, 2012.

[30] X. Huang and A. Madan. CAP3: A DNA sequence assembly program. *Genome research*, 9(9):868–877, 1999.

[31] J. Jaenike. Sex chromosome meiotic drive. *Annual Review of Ecology and Systematics*, 32(1):25–49, 2001.

[32] R. K. Kelemen and B. Vicoso. Complex History and Differentiation Patterns of the t-Haplotype, a Mouse Meiotic Driver. *Genetics*, Nov 2017.

[33] W. J. Kent. BLATthe BLAST-like alignment tool. *Genome research*, 12(4):656–664, 2002.

[34] R. J. Kinsella, A. Kähäri, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, et al. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, 2011, 2011.

[35] A. N. Kruger and J. L. Mueller. Mechanisms of meiotic drive in symmetric and asymmetric meiosis. *Cell Mol Life Sci*, Jan 2021.

[36] T. Kubo, A. Yoshimura, and N. Kurata. Pollen killer gene S35 function requires interaction with an activator that maps close to S24, another pollen killer gene in rice. *G3: Genes, Genomes, Genetics*, 6(5):1459–1468, 2016.

[37] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357, 2012.

[38] A. M. Larracuente and D. C. Presgraves. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics*, 192(1):33–53, Sep 2012.

[39] S. Lenington, L. C. Drickamer, A. S. Robinson, and M. Erhart. Genetic basis for male aggression and survivorship in wild house mice (Mus domesticus). *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, 22(2):135–145, 1996.

[40] A. Lindholm, A. Sutter, S. Künzel, D. Tautz, and H. Rehrauer. Effects of a male meiotic driver on male and female transcriptomes in the house mouse. *Proc Biol Sci*, 286(1915):20191927, 11 2019.

[41] A. K. Lindholm, K. A. Dyer, R. C. Firman, L. Fishman, W. Forstmeier, L. Holman, H. Johannesson, U. Knief, H. Kokko, A. M. Larracuente, A. Manser, C. Montchamp-Moreau, V. G. Petrosyan, A. Pomiankowski, D. C. Presgraves, L. D. Safronova, A. Sutter, R. L. Unckless, R. L. Verspoor, N. Wedell, G. S. Wilkinson, and T. A. Price. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol. (Amst.)*, 31(4):315–326, Apr 2016.

[42] M. F. Lyon. Transmission ratio distortion in mice. *Annu. Rev. Genet.*, 37:393–408, 2003.

[43] A. Manser, B. König, and A. K. Lindholm. Polyandry blocks gene drive in a wild house mouse population. *Nature communications*, 11(1):1–8, 2020.

[44] T. Morita, H. Kubota, K. Murata, M. Nozaki, C. Delarbre, K. Willison, Y. Satta, M. Sakaizumi, N. Takahata, and G. Gachelin. Evolution of the mouse t haplotype: recent and worldwide introgression to Mus musculus. *Proc. Natl. Acad. Sci. U.S.A.*, 89(15):6851–6855, Aug 1992.

[45] R. Mullenbach. An efficient salt-chloroform extraction of dna from blood and tissues. *Trends Genet.*, 5:391, 1989.

[46] N. L. Nuckolls, M. A. B. Núñez, M. T. Eickbush, J. M. Young, J. J. Lange, S. Y. Jonathan, G. R. Smith, S. L. Jaspersen, H. S. Malik, and S. E. Zanders. Wtf genes are prolific dual poison-antidote meiotic drivers. *Elife*, 6:e26033, 2017.

[47] G. Oestergren. Parasitic nature of extra fragment chromosomes. *Bot Not*, 2:157–163, 1945.

[48] N. Phadnis and H. A. Orr. A single gene causes both male sterility and segregation distortion in Drosophila hybrids. *science*, 323(5912):376–379, 2009.

[49] K. E. Pieper and K. A. Dyer. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *J. Evol. Biol.*, 29(11):2229–2241, Nov 2016.

[50] K. E. Pieper, R. L. Unckless, and K. A. Dyer. A fast-evolving X-linked duplicate of importin-$\alpha 2$ is overexpressed in sex-ratio drive in Drosophila neotestacea. *Molecular ecology*, 27(24):5165–5179, 2018.

[51] S. Prakash. Gene differences between the sex ratio and standard gene arrangements of the X chromosome and linkage disequilibrium between loci in the standard gene arrangement of the X chromosome in Drosophila pseudoobscura. *Genetics*, 77(4):795–804, 1974.

[52] A. R. Quinlan and I. M. Hall. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.

[53] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

[54] J. A. Reinhardt, C. L. Brand, K. A. Paczolt, P. M. Johns, R. H. Baker, and G. S. Wilkinson. Meiotic drive impacts expression and evolution of X-linked genes in stalk-eyed flies. *PLoS Genetics*, 10(5):e1004362, 2014.

[55] J.-N. Runge and A. K. Lindholm. Carrying a selfish genetic element predicts increased migration propensity in free-living wild house mice. *Proceedings of the Royal Society B*, 285(1888):20181333, 2018.

[56] J.-N. Runge and A. K. Lindholm. Experiments confirm a dispersive phenotype associated with a natural gene drive system. *Royal Society open science*, 8(5):202050, 2021.

[57] L. Sandler and E. Novitski. Meiotic drive as an evolutionary force. *The American Naturalist*, 91(857):105–110, 1957.

[58] A. Smit, R. Hubley, and P. Green. RepeatModeler Open-1.0. 2008–2015. *Seattle, USA: Institute for Systems Biology. Available from: httpwww. repeatmasker. org, Last Accessed May*, 1:2018, 2015.

[59] E. Stolle, R. Pracana, P. Howard, C. I. Paris, S. J. Brown, C. Castillo-Carrillo, S. J. Rossiter, and Y. Wurm. Degenerative expansion of a young supergene. *Molecular biology and evolution*, 36(3):553–561, 2019.

[60] M. Sugimoto. Developmental genetics of the mouse t-complex. *Genes Genet Syst*, 89(3):109–120, 2014.

[61] J. Svedberg, S. Hosseini, J. Chen, A. A. Vogan, I. Mozgova, L. Hennig, P. Manitchotpisit, A. Abusharekh, T. M. Hammond, M. Lascoux, et al. Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nature communications*, 9(1):1–13, 2018.

[62] Y. Tao, L. Araripe, S. B. Kingan, Y. Ke, H. Xiao, and D. L. Hartl. A sex-ratio meiotic drive system in Drosophila simulans. ii: an X-linked distorter. *PLoS biology*, 5(11):e293, 2007.

[63] C. Trapnell, L. Pachter, and S. L. Salzberg. TopHat: discovering splice junctions with RNA-seq. *Bioinformatics*, 25(9):1105–1111, 2009.

[64] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen. Primer3 – new capabilities and interfaces. *Nucleic acids research*, 40(15):e115–e115, 2012.

[65] J. V Silva, M. J Freitas, and M. Fardilha. Phosphoprotein phosphatase 1 complexes in spermatogenesis. *Current molecular pharmacology*, 7(2):136–146, 2014.

[66] S. Vijayaraghavan, D. T. Stephens, K. Trautman, G. D. Smith, B. Khatra, E. F. da Cruz e Silva, and P. Greengard. Sperm motility development in the epididymis is associated with decreased glycogen synthase kinase-3 and protein phosphatase 1 activity. *Biology of reproduction*, 54(3):709–718, 1996.

[67] Z. Yang. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, 13(5):555–556, 1997.

CHAPTER 4

# Single-nucleus RNA sequencing uncovers candidate poisons and antidotes in testes carrying the *t*-haplotype, a model meiotic driver

Réka K. Kelemen[a], Anna K. Lindholm[b] and Beatriz Viçoso[a]

[a] *Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

[b] *Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland*

# Abstract

Meiotic drivers selfishly promote their own transmission at the expense of homologous alleles in heterozygotes, and are expected to sweep to fixation. However, most of the detected meiotic drivers are balanced by fitness costs, and are large non-recombining haplotypes with many genes involved in drive. The *t*-haplotype in house mice is a poison-antidote system where the expression of several poison genes gives strong transmission advantage to sperm that carries the antidote. Although sperm motility, morphology and fertilization ability are affected in $+/t$ mice, the molecular bases of these phenotypes remain elusive, partly because of their complex, polygenic nature. Here we conducted single nucleus RNA sequencing of $+/t$ and $+/+$ testes and found differentially expressed genes at both diploid and haploid stages of spermatogenesis, with significant enrichment in motility-, mitochondrion- and fertilization-related processes. Sequencing nuclei allowed us for the first time to compare gene expression in $+$ and *t* spermatids, and find candidate poison and antidote genes based on their timing of expression. In light of our results and recent findings about sperm motility regulation, we revisit and discuss hypotheses about the molecular basis of the *t*-haplotype's drive.


## 4.1 Introduction

Natural selection favors alleles that are on average passed on to more offspring than other alleles of the same locus. Alleles that contribute to organismal fitness increase the number or survival rate of offspring they are transmitted to, but an allele can also gain transmission advantage by overrepresenting itself in the gametes of heterozygous carriers at the expense of the homologous allele [63, 53]. Alleles that alter the processes of meiosis and gametogenesis for their own benefit, so-called meiotic drivers, are expected to quickly replace homologous alternative alleles in the population, and may explain the fast evolution of genes and sequences involved in these processes [25]. On the other hand, the meiotic drivers that are discovered are often large and ancient haplotypes found at intermediate frequencies in the population, whose spread is counteracted by fitness costs imposed on their hosts [11, 41, 58]. These selfish sequences have become specialized in biasing transmission ratios in heterozygotes, which maintains them in the population despite their negative effects on host fitness [53, 7, 29, 59, 63]. Meiotic drivers often encompass large genomic regions with many genes, which are kept as one non-recombining haplotype due to their location in regions of suppressed recombination, such as in chromosomal inversions [41, 61, 46, 14]. The mechanisms of drive are often complex, as they affect various biological processes, and involve multiple genes and sequences, distorters and suppressors, which may reflect a history of arms race between the driver and the rest of the genome. [12, 10, 46].

The *t*-haplotype is a classic and well-studied example of a meiotic driver. A selfish variant of the proximal half of the autosomal chromosome 17 (the genomic region known as the *t* complex) in house mice, it is transmitted to 90-100% of the offspring in males heterozygous for it (i.e. males carrying one *t*-haplotype and one standard chromosome 17) [46]. Most *t*-haplotypes contain recessive embryonic lethals, which eliminate homozygotes, while *t*-variants without lethals produce sterile male homozygotes and fertile female homozygotes. Despite a century of phenotypic and molecular studies, how the *t*-haplotype reaches such high transmission ratios it is not fully understood. $+/t$ males produce normal amounts of sperm, but the number of sperm reaching the site of fertilization is significantly reduced [52]. This is due to the fact that all sperm from $+/t$ mice have impaired motility – they show lower straight-line-velocity and less linear movement than sperm from $+/+$ mice, as well as morphological changes, such as

shorter tails and larger head-to-flagellum ratios [67]. Furthermore, sperm from *t*-carriers was reported to show fertilization defects, such as delayed penetration of the zona pellucida, the surrounding layer of the egg [35], and premature acrosome reaction, a process that releases enzymes necessary for egg penetration [6]. Because it is technically challenging to distinguish sperm cells carrying the *t*-haplotype and those carrying the standard chromosome 17 while assessing their performance, it is unknown how + and *t* sperm differ from each other in the affected traits. Bimodality has been reported for sperm tail beat frequency and path linearity [36, 2], but genotyping the top 25% most progressively moving sperm showed only a 60:40 ratio of *t* versus + sperm [2]. Another recent study found unimodal distributions for all tested sperm motility and morphology traits in $+/t$ mice, with the top 25% most progressive sperm in $+/t$ mice having significantly worse motility and morphology parameters than the top 25% in $+/+$ mice [67]. In line with this, the *t*-haplotype is significantly underrepresentated in litters from polyandrous matings [48]. Therefore, the *t*-haplotype seems to only partially compensate for its self-induced harm, but the biased, haploid nature of its compensation is enough to overrepresent it in the dozen or so fertilizing sperm in monogamous matings. Further, although the motility of $+/t$ sperm has been more heavily researched, the respective roles of poor sperm motility and fertilization-related defects in transmission ratio distortion are unknown.

Male meiotic drivers need to create functional differences between sperm carrying them and sperm not carrying them in order to gain their transmission advantage [39]. In poison-antidote systems, such as the *t*-haplotype, this is achieved by at least one trans-acting factor, the so-called "poison", which is shared among sperm, and at least one unshared cis-acting factor, the "antidote", which selectively saves sperm carrying the meiotic driver [39]. The sharing of the poisons may be achieved either by pre-meiotic expression (when cells are diploid), or through the diffusion across cytoplasmic bridges that connect developing haploid spermatids. On the other hand, keeping the antidote private requires haploid expression and limiting mRNA and/or protein diffusion by late expression, late translation and/or RNA tethering to the nucleus.

Six regions that contribute to transmission ratio distortion were identified in the *t* complex, while one region is responsible for the antidote function [46]. Unraveling the molecular pathways underlying drive begun by the identification of two candidate genes in two different distorter regions, which code for dyneins in the mouse sperm flagella [40, 32] – motor proteins that coordinate flagellar beating in response to signals transmitted from the central microtubules through radial spoke proteins [43]. *Dynlt1* is present in multiple copies on both the + and *t* chromosomes, with overexpression from the diverged *t*-allele [40, 37], while *Dynlt2* is present in three copies on the + chromosome, and the *t* copy is strongly underexpressed [32]. Although the involvement of dyneins in drive has never been tested [56], a homozygous deletion of the proximal *Dynlt2* copy in $+/+$ mice leads to poor sperm motility [60], which may suggest that the underexpression of *Dynlt2*$^t$ may affect sperm motility.

In 1999, Herrmann and colleagues identified one gene in the responder region whose *t*-specific copy proved to act as an antidote, and promote the transmission of the chromosome it was on in the presence of the *t*-haplotype poisons [27]. This gene shows homology to serine/threonine protein kinases, has haploid-specific expression, and localizes to sperm nuclei and the principal piece of flagella - hence the name sperm motility kinase, *Smok* [66]. The *t*-haplotype's copy, *Smok*$^{Tcr}$ (standing for *t* complex responder), codes for many amino acid changes and shows a tenfold reduction in phosphorylation activity compared to the homolog on the + chromosome [27]. Herrmann and colleagues hypothesized that an altered regulatory pathway upstream of *Smok* might hyperactivate the + allele to abnormal levels, while the hypomorphic *Smok*$^{Tcr}$'s

activity is raised to normal levels, resulting in better function of *t*-bearing sperm [5]. Their search for regulatory proteins in the distorter regions yielded four genes, whose *t*-alleles are associated with drive. All four of them (TAGAP1, FGD2, NME3 and TIAM2) are regulators of Rho GTPases, and their *t*-alleles show altered testis expression [8, 4, 5, 3]. However, it is unkown if or how these four driver proteins influence SMOK function, and whether SMOK is hyperactivated in +-sperm. Further, the function and flagellar localization of $Smok^{Tcr}$ call for the investigation of further driver candidates, especially in the light of recent advancements in understanding sperm motility.

The past decades have uncovered the central role of protein phosphorylation in controlling the function of transcriptionally and translationally silent sperm cells [62, 1, 16]. In particular, dephosphorylation must be turned off and phosphorylation has to be turned on for motility initiation and hyperactivation [16]. Since SMOK is a flagellar kinase whose hypomorphic *t*-allele results in better sperm function in combination with the *t*-specific distorter products, it is possible that proteins of the antagonistic dephosphorylation pathways or downstream phosphorylation targets of SMOK are involved in drive. Notably, an isoform of the key protein phosphatase responsible for the inhibition of sperm motility through dephosphorylation [18, 16], as well as two of its three inhibitors, are encoded on the *t*-haplotype, and their sequence and expression evolution suggest a role in drive. The *t*-specific gene encoding the phosphatase, PPP1, was acquired as a duplicate from the original chromosome 5 gene, and is absent from the standard *t*-complex. This gene, $Ppp1cb^{t}$, acquired many nonsynonymous changes on the *t*-haplotype, and shows high testis-specific expression, unlike $Ppp1cb^{chr5}$, which is expressed in many tissues [37, 18]. The PPP1 inhibitor, PPP1R2 is also encoded in the *t* complex (although in the form of several intronless genes with intact open reading frames annotated as pseudogenes), with overexpression from the *t*-allele [56, 37]. Finally, the other key PPP1 inhibitor, PPP1R11, is an ancestrally *t*-complex gene, that shows signs of positive selection on the *t*-haplotype [37]. A study found that $+/t$ mice that carried a $Ppp1r11^{MusSpretus}$ allele from the sister species *Mus spretus* instead of the $Ppp1r11^{+}$ allele showed severe flagellar abnormalities usually associated with $t/t$ mice, while sperm from $Ppp1r11^{MusSpretus}/Ppp1r11^{+}$ mice are normal [55], which implicates $Ppp1r11^{t}$ in motility defects. While the location of $Ppp1cb^{t}$ on the *t*-haplotype is unknown, *Ppp1r2* and *Ppp1r11* reside in the distorter regions of the *t*-haplotype [56].

SMOK, PPP1 and its inhibitors all localize to the principal piece of the flagellum [66, 18], the part responsible for creating the wave-like motion that drives sperm forward. SMOK was specifically shown to be near the outer dynein arms, encoded by the *t*-complex-gene *Dnah8*, whose *t*-allele has non-conservative mutations and similarly to $Ppp1r11^{t}$, $Dnah8^{t}/Dnah8^{MusSpretus}$ sperm show the flagellar abnormalities seen in $t/t$ mice [55, 56]. Other dyneins are likely targets of phosphorylation as well, as *Dynlt2*'s phosphorylation is required for sperm motility initiation in sea urchins [33]. Further, the *t*-complex-encoded radial spoke protein, RSPH1 colocalizes with SMOK in the fibrous sheath and outer dense fibers of the principal piece [31], and was reported to co-immunoprecipitate with PPP1 inhibitors [56]. Overall, the *t*-haplotype codes for important motor and structural proteins of the sperm flagellum, as well as for the catalytic and inhibitory subunits of a key negative regulator of sperm motility, PPP1. While these candidate genes show differential expression in the testis, for many of them it is not known when during spermatogenesis they are expressed. Expression timing is an important factor in achieving drive, as early expression creates higher potential in distributing mRNA or protein products among spermatids, while late expression limits the time for product distribution and may lead to functional differences between spermatids [39]. Although only $Smok^{Tcr}$ has been found to act as an antidote, there has not been a systematic screen for late-acting

differentially expressed genes in $+/t$ testes. Additionally, genes outside of the *t*-complex may also cause functional differences between $+$ and *t*-spermatids if they show differential expression late during spermatogenesis. Indeed, previous whole-tissue RNA-sequencing studies found widespread differential expression between $+/+$ and $+/t$ testes in the rest of the genome [44, 38].

Here we conduct an in-depth expression analysis of testis cell types in $+/t$ mice using single nucleus RNA-sequencing. Single nucleus RNA-sequencing allows the detection of expression changes within certain cell types, which may be undetected in whole-tissue RNA-sequencing, and further allows us to distinguish *t*-carrying and $+$-carrying spermatids within an individual. Through a genome-wide screen for differentially expressed genes in $+/t$ versus $+/+$ mice in early or late stages of spermatogenesis, we identify genes and processes that may act as poison or antidote. We further directly compare the transcriptomes of $+$- and *t*-spermatids, and investigate the molecular underpinnings of their shared impairments, and their unequal transmission rates.

# 4.2 Results

## Nuclei cluster according to cell type and allow haploid genotype inference

We isolated nuclei from testis samples of two *t*-carrier ($+/t$) mice and one non-carrier ($+/+$) mouse, which were full siblings, and conducted single nucleus 3' mRNA-sequencing on the three samples (Figure 4.1A). The *t*-haplotype is highly diverged from the homologous region of the mouse chromosome 17 [38], which may hinder the mapping of *t*-haplotype-derived reads to the mouse reference genome. Therefore, we created a pseudo-*t*-haplotype sequence by substituting previously identified *t*-specific SNPs [38] into the *mm10* reference genome. We appended the pseudo-*t*-haplotype sequence, as well as the sequences of the previously reported gained genes of the *t*-haplotype, *Ppp1cb*$^t$, *Rnpepl1*$^t$ [37], and of the diverged responder, $Smok^{Tcr}$ [27], to the reference genome. After mapping, quantifying expression and quality filtering we obtained 5886 high-quality nuclei from the three mouse samples (see Methods, Supplementary figure 4.6), which were clustered based on their gene expression patterns, and annotated using marker genes for the different mouse spermatogenesis stages [45, 19] (Supplementary figure 4.7). We recovered five somatic cell type clusters, and twelve germ cell clusters representing the various stages of spermatogenesis (Figure 4.1B), with similar proportions of nuclei in all three samples (p=0.224, Chi-square test, Supplementary figure 4.8). To infer the genotypes of the 1666 haploid spermatid nuclei in the two $+/t$ samples, we calculated a *t*-score for each nucleus by taking the difference in mean expression of pseudo-*t* genes and the reference *t* complex genes (3-42 Mb on chromosome 17, following [44]) (Figure 4.1C). Haploid spermatids showed bimodal *t*-score distributions, of which the top 90th percentiles we called *t*-spermatids, and the bottom 90th percentiles $+$-spermatids (for the more unimodal early round spermatids 1 cluster, the top and bottom 75th percentiles were taken, see Methods and Supplementary figures 4.9 and 4.10). While somatic and pre-meiotic diploid cells have unimodal distributions centered around 0, consistent with equal expression from both alleles, interestingly, meiotic cells have an expression bias towards the $+$-allele (Figure 4.1C). This is due to the increased expression of $+$ alleles in the *t* complex in meiotic cells, while the average pseudo-*t*-allele expression is more balanced across cell types (Supplementary figure 4.11). A similar increase in the expression of *t*-complex-genes in meiotic cells is seen in $+/+$ mice (Supplementary figure 4.12).

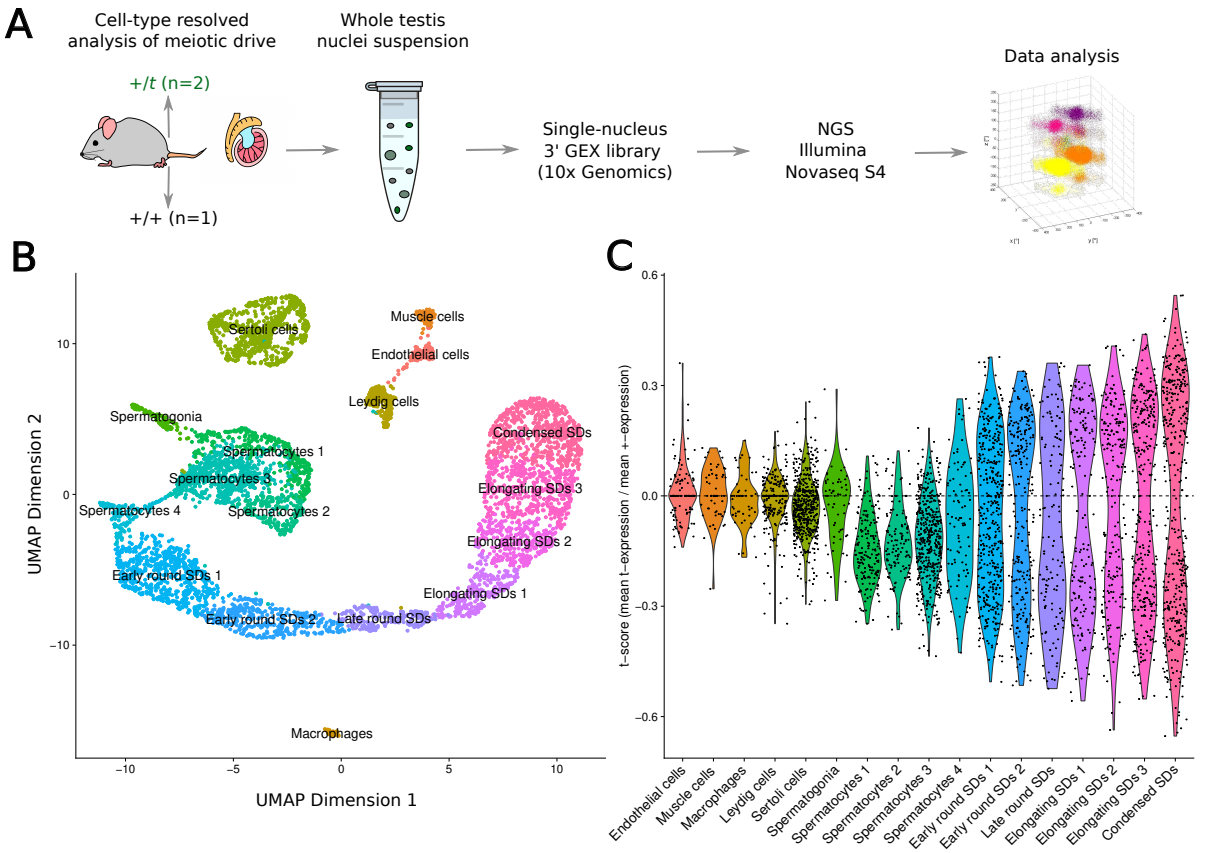



Figure 4.1: Single nucleus sequencing clusters nuclei based on cell type and genotype. (A) Experimental pipeline (B) UMAP (Uniform Manifold Approximation and Projection) plot positioning high-quality nuclei from all three samples according to their gene expression patterns. Dots (nuclei) with the same color belong to the same cluster, as inferred by a shared nearest neighbor (SNN) clustering algorithm in R (see Methods). Clusters were annotated as cell types based on the expression of testis cell type marker genes [45, 19] (Supplementary figure 4.7). (C) *t*-scores (mean *t*-expression - mean + expression) for nuclei in $+/t$ samples are shown. Nuclei are binned into cell types. SD - spermatids

## Widespread differential expression at early and late stages of spermatogenesis

In poison-antidote systems the poison is spread across cells, while the antidote acts only intracellularly to selectively save its carrier [39]. Transcriptional regulation may contribute to this difference through early expression of the poison genes, and late expression of the antidotes.

Genes on the *t*-haplotype can be diverged from those on the reference genome, which could bias their expression estimates. To obtain reliable estimates of gene expression, we assembled a transcriptome based on the reads in this dataset. We made use of the identified *t*-spermatids to assemble *t*-haplotype-specific transcripts, as well as the control library to assemble a transcriptome specific to our mouse strain; the two transcriptomes were then merged, and identical transcripts were removed. Our resulting transcriptome contained transcripts of three of the four known poison genes, *Nme3, Fgd2* and *Tiam2*, and of the antidote, *Smok*$^{Tcr}$. Interestingly, our assembled *Smok*$^{Tcr}$ allele only contained the *Smok* sequence, and no reads in our dataset supported its fusion to a truncated *Rps6ka2* gene sequence, as reported at its

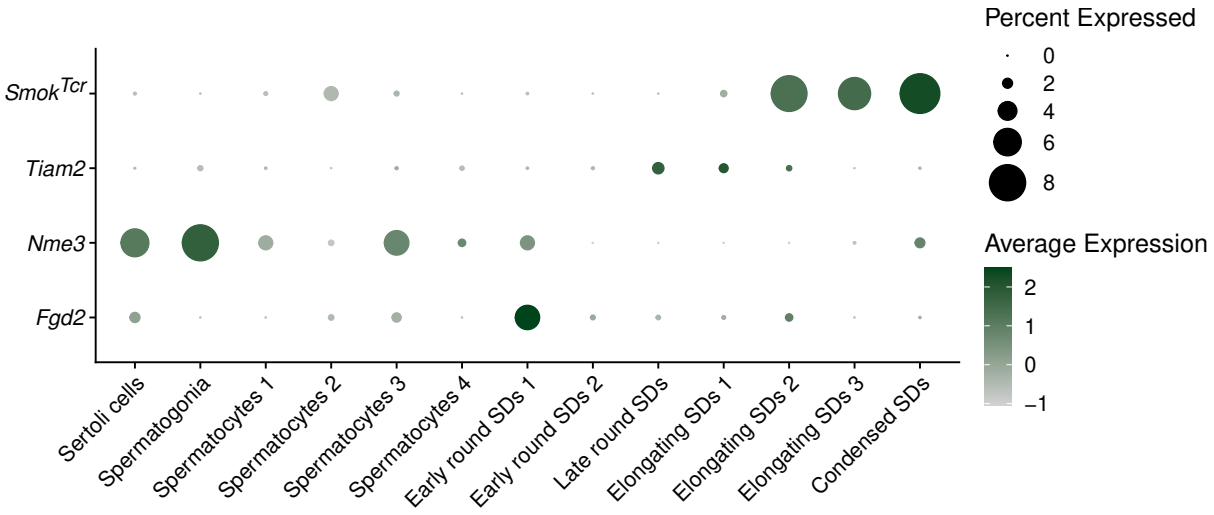



Figure 4.2: Late-stage expression of the responder ($Smok^{Tcr}$) and earlier expression of the known distorter genes (*Tiam2, Nme3* and *Fgd2*). The percent of nuclei expressing each gene is proportional to the size of the circles, and the average nucleus-normalized expression is shown by the shade of green. Only $+/t$ diploid nuclei were included in the diploid clusters, and only *t*-spermatids were included in the spermatid clusters. The distorter *Tagap1* was not expressed in our dataset. SD - spermatid

initial discovery (Supplementary figure 4.13) [27]. Although identified in a low percentage of nuclei in our dataset, $Smok^{Tcr}$ shows increasingly strong expression in elongating and condensed spermatids, while remaining mostly absent in previous stages (Figure 4.2), in line with previous studies [27, 66]. Poison genes, on the other hand, are expressed at earlier stages, with peak expression in spermatogonia, early round spermatids and late round spermatids for *Nme3, Fgd2* and *Tiam2*, respectively (Figure 4.2), consistent with other studies [8, 4, 5].

Differential expression between $+/t$ and $+/+$ mice may point to candidate genes and pathways involved in transmission ratio distortion, while the timing of differential expression could distinguish poisons from antidotes. Therefore we tested the 16,743 genes in our assembly for differential expression between $+/t$ and $+/+$ samples within each of the 17 clusters. We found 383 differentially expressed genes, with the somatic Sertoli cells having the most differentially expressed genes (165), followed by the diploid meiotic germ cells, spermatocytes (143) (Figure 4.3A). Sertoli cells are the only somatic cells present in the location of germ cell production, in the seminiferous tubules, and they control the extracellular microenvironment and instruct germ cell progression through cell-to-cell contacts [54]. Therefore, the widespread expression changes in Sertoli cells may have an effect on spermatid development and the function of the resulting mature sperm cells. Spermatocytes are the cells that undergo meiosis, and therefore represent the last diploid stage of spermatogenesis, when poison products can be easily distributed, as differentially expressed genes at this stage will likely affect both daughter cells, $+$ and *t* spermatids.

When we compared gene expression in $+$ and *t* spermatids to each other and to the $+$ spermatids in $+/+$ mice, we found 19, 62 and 76 differentially expressed genes, respectively. Interestingly, only 34 of the 383 differentially expressed genes identified between *t*-carriers and non-carriers mapped to the *t* complex, in line with the widespread genome-wide expression changes found previously [44]. In contrast, 12 out of 19 differentially expressed genes between $+$ and *t* spermatids in $+/t$ mice are *t*-complex-genes. The distorter genes, *Nme3, Fgd2* and *Tiam2*, were not differentially expressed, while $Smok^{Tcr}$ was differentially expressed between $+$

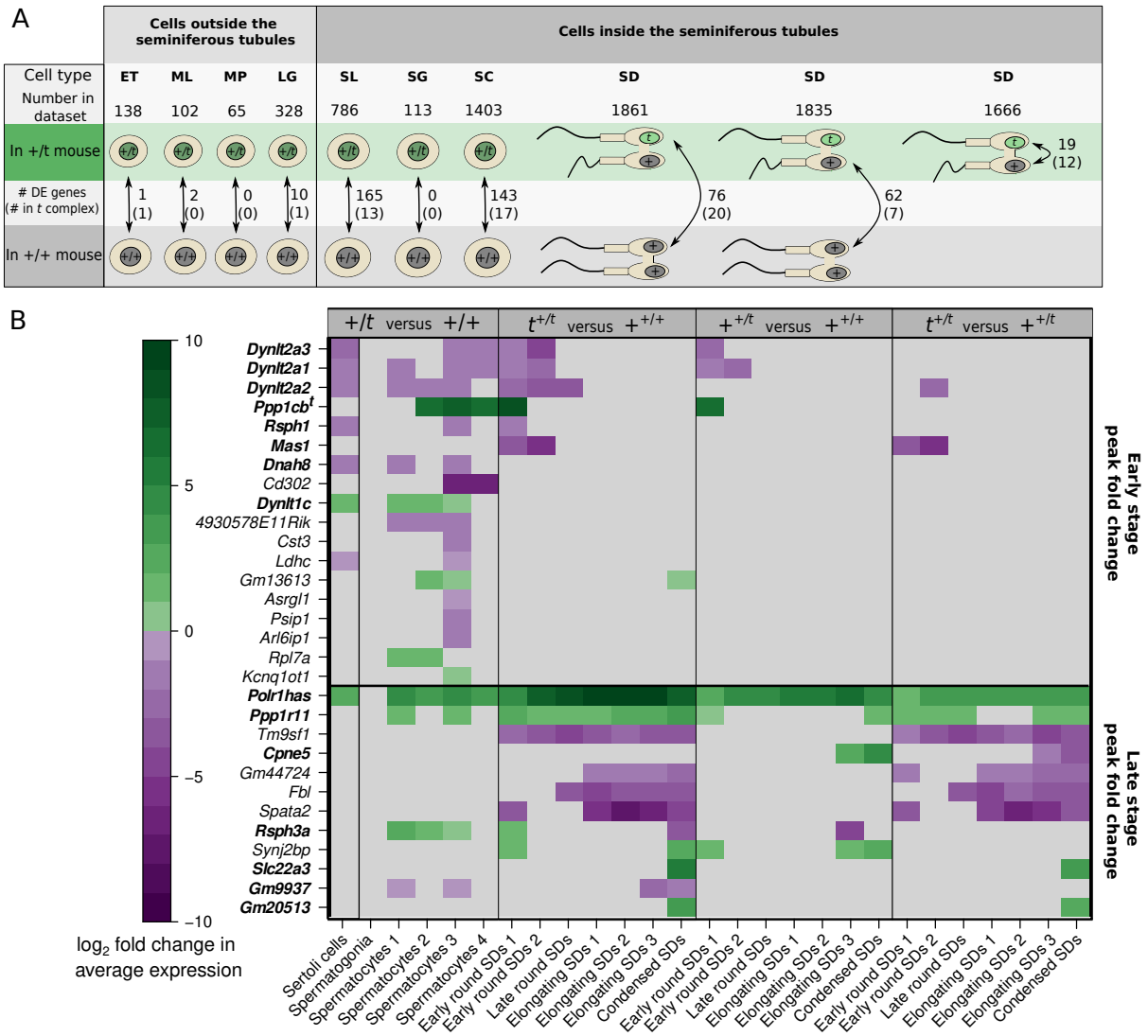



Figure 4.3: Widespread expression changes early and late during spermatogenesis affect genes in and outside of the *t* complex. (A) The number of differentially expressed genes detected in each cell type are shown, with the number of differentially expressed *t* complex genes in parentheses. The total number of nuclei in each comparison and cell type are shown on top. Cartoons show the genotype of the nuclei compared. ET - endothelial cells, ML - muscle cells, MP - macrophases, LG - Leydig cells, SL - Sertoli cells, SG - spermatogonia, SC - spermatocytes, SD - spermatids. (B) Differential expression of the 30 most significantly differentially expressed genes, classified based on having their largest absolute log fold change before the elongating spermatid stages (early) or not (late), and ordered by adjusted p-value increasing from top to bottom on each category. Shades of green and purple indicate the magnitude of log fold change of overexpressed and underexpressed genes, respectively. Log fold change is calculated based on average expression in the compared groups. Grey indicates no significant expression change. The genotypes of the compared groups are shown on top of the boxes. Genes printed with bold font are in the *t* complex.

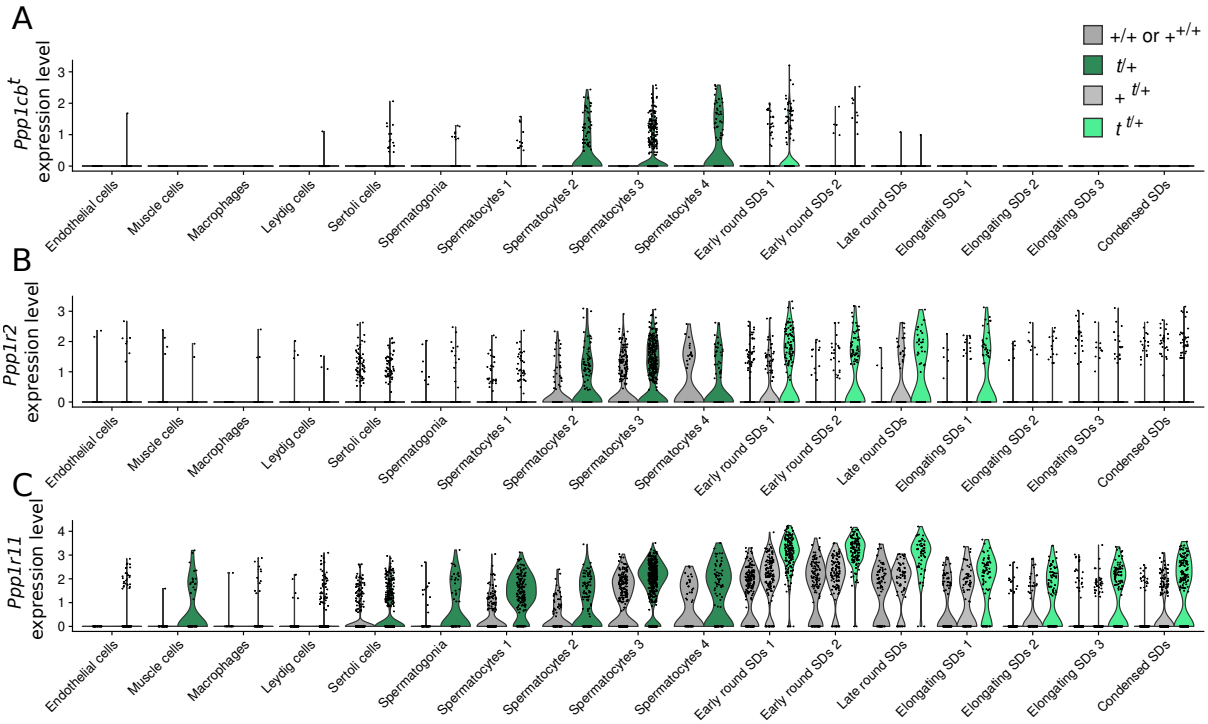



Figure 4.4: Expression of *Ppp1cb*$^t$, *Ppp1r11* and *Ppp1r2*. Dark grey violin plots denote expression in diploid and haploid stages of the $+/+$ sample. Dark green violin plots show expression in the diploid $+/t$ cells in the two $+/t$ samples. Light grey and light green violin plots represent expression in $+$-spermatids and in *t*-spermatids of the $+/t$ sample, respectively.

and *t* spermatids in the last stage, condensed spermatids.

To understand how the timing of expression of germline differentially expressed genes may relate to their potential function as poisons or antidotes, we used the expression of *Smok*$^{Tcr}$ as a guideline. We classified genes that had their highest absolute log fold change in elongating or condensed spermatid clusters as late differentially expressed genes, and those with earlier peaks in absolute log fold change as early differentially expressed genes. Of the 242 differentially expressed genes in the germline, 172 were early, and 70 were late. Figure 4.3B shows the 30 most significantly differentially expressed genes in the germline, with 18 early genes and 12 late genes. These 30 genes are significantly enriched for the biological process of axonemal dynein complex assembly (FDR=0.0354) and for the mammalian phenotype ontology of abnormal sperm physiology (FDR=0.0233), with a significant enrichment of protein-protein interactions among them (p=0.0186). Eight of the top 10 early genes are *t* complex genes: five dynein genes, *Dynlt2a1, Dynlt2a2, Dynlt2a3, Dynlt1c* and *Dnah8*, the radial spoke protein-encoding gene, *Rsph1*, and the *t*-specific allele of motility-inhibiting protein phosphatase (PPP1), *Ppp1cb*$^t$ (Figure 4.4A). Additionally, the PPP1 inhibitor, *Ppp1r2*, is overexpressed in *t* spermatids compared to $+^{+/t}$ spermatids and $+^{+/+}$ spermatids in the early haploid stages (Figure 4.4B), likely due to the overexpression of the *t*-complex-encoded *Ppp1r2* on the *t*-haplotype, whose reads may map to the chromosome-16-encoded *Ppp1r2*.

The late genes include the *t* complex gene, *Ppp1r11*, an inhibitor of PPP1, which is overexpressed in *t*-carrying cells all throughout spermatogenesis, and shows its largest log fold change between $+^{+/t}$ and *t*-spermatids in the late haploid stages (Figure 4.4C). A radial spoke protein-encoding gene, *Rsph3b*, switches from overexpression in spermatocytes to stronger underexpression in haploid spermatids of $+/t$ mice, which classifies it as a late-acting

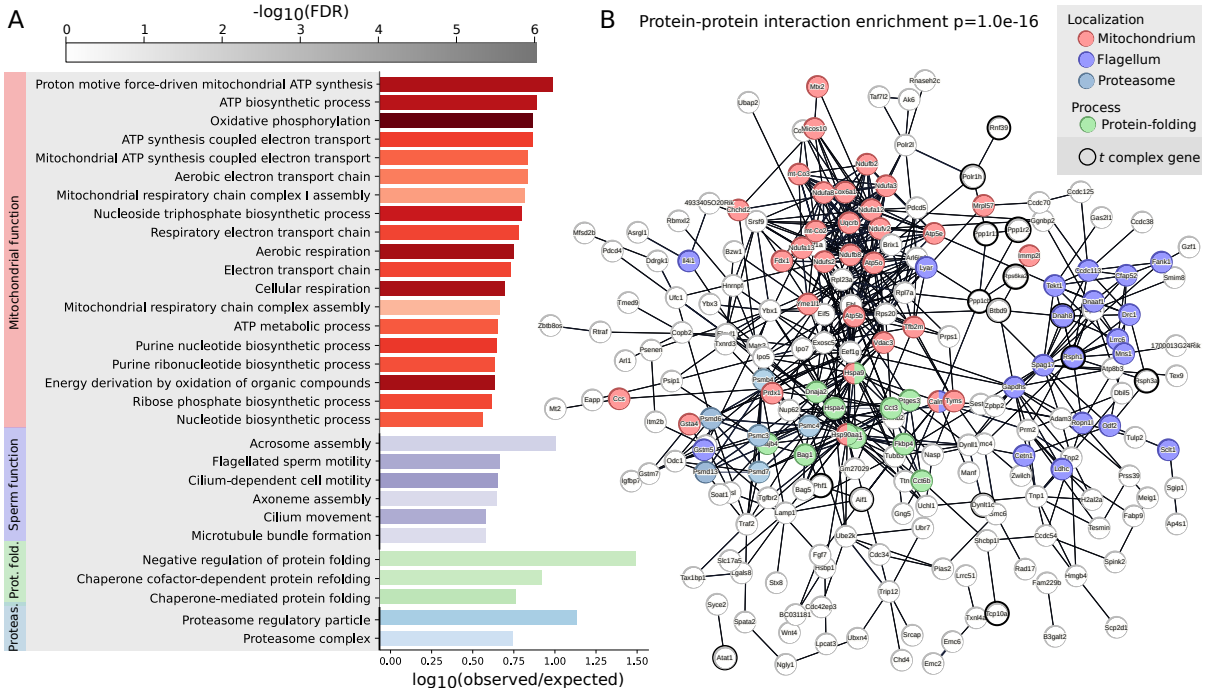



Figure 4.5: Differentially expressed genes (in any of the comparisons shown in Figure 4.3A) in the cells of the seminiferous tubules are enriched for mitochondrial, sperm-related, chaperonin and protease functions. (A) 30 strongest gene ontology enrichments of differentially expressed genes in cells of the seminiferous tubules. Red stands for mitochondrial functions, purple for sperm-related functions, green for chaperonin (protein folding) functions and blue for protease (protein degradation) functions. The x axis shows the strength of enrichment as the log (observed number of genes / expected number of genes), and the shading of the colors is proportional to the -log10(False Discovery Rate). (B) Protein-protein interaction network of the 302 differentially expressed genes in the seminiferous tubules, including the physical interactions, as well as co-expression, and co-appearance in publications. Only connected genes are shown. Red, purple, blue and green circles stand for genes with cellular localization in the mitochondria, sperm flagella, proteasome and with function in protein-folding, respectively. *t* complex genes are surrounded by thick black circles.

differentially expressed gene.

Two genes from an imprinted gene cluster of the *t* complex, related to embryonic growth are among the top 30 differentially expressed genes, consistently with our previous findings based on bulk RNA-seq [38]. Our data give insights into the timing of their expression: *Mas1* is underexpressed only in *t*-carrying round spermatids, while *Slc22a3* shows overexpression at the very last stage of spermatogenesis, in *t*-carrying condensed spermatids (Figure 4.3B). Mice homozygous for the sequenced *t*-haplotype strain die as embryos, and the underlying recessive embryonic lethal is unknown. The misexpression of *Slc22a3* and *Mas1* in *t* spermatids raises the interesting possibility that these genes alter the embryonic development of offspring inheriting the *t*-haplotype.

### Differential expression affects mitochondrial, flagellar, chaperonin and proteasomal genes

The previous section suggests that there is widespread expression misregulation in the cells of the seminiferous tubules: the germ cells, as well as the essential Sertoli cells whose expression differences may affect germ cell function [54, 50]. To obtain a better understanding of what pathways and cellular mechanisms may be affected, we looked for enrichment in biological processes and cellular components (Figure 4.5A), as well as enrichment in protein-protein interactions within this set of 377 differentially expressed genes (Figure 4.5B). Our gene ontology enrichment analysis yielded 60 significantly enriched biological processes (Supplementary table 4.1), and 32 significantly enriched cellular components (Supplementary table 4.2). The majority of resulting biological processes could be grouped into three major functional categories that were in agreement with the cellular component in which differentially expressed genes were enriched. The largest functional category consisted of mitochondrium-related processes (Figure 4.5A), which also showed a significant excess of interaction among the proteins they code for (p=1.0e-16). Mitochondria localize to the mid piece of sperm flagella, and their respiratory efficiency correlates with sperm motility [15], while expression aberrations in both mitochondrial and nuclear-encoded mitochondrial proteins is associated with less progressive sperm movement in humans [34].

We also find an enrichment of differentially expressed genes involved in sperm-related processes, such as the assembly of the acrosome, the structure of the sperm head that contains the enzymes necessary for sperm-egg fusion (Figure 4.5A). Further sperm function related processes that were enriched included the assembly of the axoneme, the cytoskeletal structure of the sperm tail, and of its component, the microtubule bundle, as well as processes of flagellar motility. Interestingly, we find an enrichment in chaperonin-mediated protein folding, including elements and interactors of the CCT protein complex on the sperm head surface that may mediate binding to the zona pellucida, the extracellular glycoprotein layer of the egg [13]. Since sperm from $+/t$ mice show delayed penetration of the zona pellucida, expression changes of CCT-complex-associated genes are candidates for underlying this phenotypic aberration.

## 4.3 Discussion

Since its discovery about a hundred years ago, the *t*-haplotype has been the object of extensive research. These many studies have uncovered the complex phenotypic effects of this selfish element on sperm function, as well as the polygenic underpinnings of its drive. Our single nucleus sequencing of $+/t$ and $+/+$ testes uncovered the extent of differential expression between the cells of these two genotypes, and it also allowed comparing gene expression in *t*- and +-spermatids within $+/t$ testes. However, as only one $+/+$ sample was used, caution has to be taken while interpreting the results of the former comparison. Further, when comparing gene expression of +-spermatids in $+/+$ vs. $+/t$ testes, one has to keep in mind that the sequenced mRNA in +-spermatids in $+/t$ testes may be contaminated with ambient RNA of the $+/t$ testis, or with mRNA that was shared through cytoplasmic bridges from *t*-spermatids, and may not solely reflect expression from +-spermatids. On the other hand, the 19 differentially expressed genes between *t*- and +-spermatids within $+/t$ samples likely stem from true expression differences between these cells.

The high proportion of differentially expressed genes outside of the *t* complex (Figure 4.3A) suggests a regulatory effect of the *t*-haplotype on the expression of certain genes. A high

impact on the genes throughout the genome was found in earlier whole-tissue RNA-sequencing studies [44, 38], which is in line with the strong influence of the genetic background on the strength of transmission ratio distortion [21]. What causes such widespread expression changes is unknown. One possibility is that the *t*-haplotype encodes for variants of transcription factors that can modulate the expression of many genes at once. These may also provide an explanation for why sets of genes involved in specific biological processes may change their expression concurrently. For example, some transcription factors coordinate the expression of nuclear-encoded mitochondrial genes in a cell-type-specific manner and respond to physiological stimuli, such as hormones [17]. We detected a significant underexpression of *Taf7l2*, a germline-specific transcription factor component, in *t*-carrier mice. *Taf7l*-null males are known to show reduced sperm motility and abnormal, folded tails [9]. *TAF7l2* is connected in the top part of our protein-protein interaction network (Figure 4.5) to the RNA polymerase II *POLR2L*, and indirectly to *POLR1H*, a *t*-complex-encoded RNA polymerase I subunit. Interestingly our most significantly differentially expressed gene with constitutive overexpression in all cell types is *Polr1has*, which encodes an antisense lncRNA of *Polr1h* (Figure 4.3B). A significantly underexpressed *t* complex gene in spermatocytes and Sertoli cells is *Tcp10a*, a negative regulator of transcription by RNA polymerase II. Therefore, our dataset may give insights into the primary, secondary and downstream effects of the *t*-haplotype on gene expression, which may result in the "poisonous" functional aberrations that lead to transmission ratio distortion.

We observed widespread expression changes at the diploid spermatocyte stages despite the fact that cytological studies of $+/t$ testes did not detect any aberrations in meiosis or sperm development [28], and most effects seem to be confined to the motility and fertilization ability of mature sperm. Indeed, distorter genes in other meiotic drive systems are often expressed at the diploid stages, while the cytological effects are observed at later stages. For example, the *Paris sex-ratio* distorters are expressed in the pre-meiotic germ cells, spermatogonia, but the cytological effects, namely the aberrant segregation of the Y chromosome, are observed only at meiosis II in spermatocytes [24]. Similarly, the *SD* system of *Drosophila melanogaster* expresses its distorter early in meiosis, but the chromatin compaction problems arise post-meiotically [47]. The fact that we found the majority of differentially expressed genes in meiotic cells may reflect that proteins needed for mature sperm function are expressed already in spermatocytes, or may point to the advantage of unleashing distorters at diploid stages.

Our single nucleus sequencing analysis uncovered that the somatic Sertoli cells are highly affected by the *t*-haplotype. While Sertoli cells have 165 differentially expressed genes when comparing 786 nuclei, Leydig cells have only 10, despite comparing 328 nuclei. Some of this difference may result from the smaller number of nuclei obtained from Leydig cells compared with Sertoli cells, which reduced our power to detect differential expression. However, Sertoli cells are very dynamic transcriptionally, as they simultaneously support germ cells at different developmental stages and are involved in the complex endocrine and paracrine regulation of spermatogenesis [69, 20]. A recent single cell sequencing study comparing young and aged primate testes also found the transcriptomes of Sertoli cells to be most affected [30]. Because of their central roles in spermatogenesis [68], differentially expressed genes in Sertoli cells could have an effect on developing germ cells. Notably, most of the differentially expressed genes of Sertoli cells (132 out of 165) are not differentially expressed in other cell types, but are enriched for processes shown in Figure 4.5: oxidative phosphorylation (FDR=0.015) and flagellar sperm motility (FDR=0.015) . It is possible that the *t*-haplotype employs Sertoli cells to impact sperm development, which may act as distributors of the "poisonous" effect.

The widespread transcriptomic changes induced by the *t*-haplotype were significantly enriched

for several biological processes. This may suggest that the *t*-haplotype distorts transmission ratios by disrupting multiple processes at once. The misregulation of genes involved in mitochondrial processes is associated with less progressive sperm motility in humans [34]. Further, the respiratory function of mitochondria is known to influence sperm motility [15], and is required for capacitation, a crucial sperm maturation step, during which many mitochondrial proteins undergo tyrosine phosphorylation [64]. Although the underlying gene has not been identified, one *t* complex locus is involved in mitochondrial organization in the midpiece of sperm flagella [57]. When the *t*-allele at this locus is paired with the allele from the sister species, *Mus spretus*, the mitochondria do not go through condensation during sperm maturation and show poorly organized mitochondrial sheaths in the midpiece, while this is not seen for the *Mus spretus*/+ allele combination. It is plausible that a *t*-specific variant of such a locus could similarly lead to mitochondrial defects.

Sperm from $+/t$ mice was shown to be delayed in penetrating the zona pellucida of the egg [35] and to have different sperm-zona pellucida receptor activity [65], phenotypes that affect fertilization ability and therefore could be related to drive. The *t* complex encodes a mutated chaperonin, *Tcp1*, which is highly expressed in the testis and together with seven other non-*t*-complex-encoded chaperonins constitute the testis-biased chaperonin-containing-Tcp1-complex (CCT). CCT was shown to aid in the nuclear compaction and cellular matrix reorganization of spermatids, and, importantly, relocates to the sperm head surface after capacitation, the final maturation step of sperm in the female genital tract [13]. CCT shows high affinity for the zona pellucida and binds the zona-pellucida-binding protein 2, *Zpbp2*, whose knock-out causes impaired acrosome compaction and zona-pellucida-binding [42]. Our differentially expressed genes were functionally enriched for chaperonin-function (Figure 4.5), and showed the underexpression of two CCT components, *Cct3* and *Cct6b*, in spermatocytes, as well as the underexpression of three other chaperonins, *Hspa4*, *Hspa9* and *Hsp90aa1*, which might interact with CCT [13]. Further, the zona-binding protein, *Zpbp2*, is significantly underexpressed in Sertoli cells in $+/t$ mice, emphasizing the potential involvment of CCT and zona-binding in the mechanism of drive, and the potential role of Sertoli cells in the associated sperm maturation steps.

Sperm motility depends on three components: an intact flagellar morphology, the ability to produce energy for flagellar movement and properly functioning signal transduction pathways that transmit external signals into internal ones [16]. The *t*-haplotype seems to affect all three of these aspects of sperm motility. Lindholm *et al.* found shorter sperm tails in $+/t$ mice when compared to $+/+$ mice, and the significant underexpression of five dynein genes and a radial spoke gene in $+/t$ mice might contribute to this (Figure 4.3B). The strong enrichment of our differentially expressed genes in mitochondrial processes suggests that oxidative phosphorylation in sperm from $+/t$ mice might be affected. However, evidence is growing that an alternative process, glycolysis, carried out by enzymes along the entire flagellum is an important source of ATP for sperm motility [26]. In particular, a sperm-specific lactase dehydrogenase, LDHC, is essential for sperm ATP production and motility [51], which is one of our most significantly underexpressed genes in $+/t$ mice (Figure 4.3B).

Finally, we show the strong overexpression of PPP1 and its repressors in *t*-carrier testes, which may imply altered motility-related signaling in + and *t* sperm. *Ppp1cb*$^t$ and its repressors, *Ppp1r2* and *Ppp1r11*, are overexpressed in the early stages of spermatogenesis in $+/t$ versus $+/+$ mice, facilitating their incorporation into both + and *t* spermatids. Interestingly, *Ppp1r11* is differentially expressed between + and *t* spermatids in $+/t$ mice, with a three-fold higher expression in *t* cells than in + cells at the elongating and condensed spermatid stages (Figure

4.4C). One possible interpretation is that *Ppp1r11*$^t$ is a poison, and the *t*-haplotype shares its allele with + spermatids by expressing it at diploid stages, and by expressing it at haploid stages where it will be distributed through the cytoplasmic bridges. *Ppp1r11* was shown to be enriched in chromatoid bodies, large, spermatid-specific structures that might facilitate mRNA transport across cytoplasmic bridges [49], which would support it being a poison. Another interpretation is that *Ppp1r11*$^t$ is an antidote, possibly to the deleterious effect of *Ppp1cb*$^t$, and its strong overexpression at the late haploid stages may enrich it in *t* spermatids even in the face of equalization through the chromatoid bodies. Its coexpression with *Ppp1cb*$^t$ in the diploid stages may be necessary to regulate PPP1's enzymatic activity along spermatogenesis [18].

More work is necessary to better understand the possible role of PPP1 and its inhibitors in the *t*-haplotype's drive. Bioinformatic prediction of *Ppp1cb*$^t$'s, *Ppp1r11*$^t$'s and *Ppp1r2*$^t$'s protein structures would allow their comparison to the structures of their non-*t*-specific homologs and paralogs. This would give insights into whether *Ppp1cb*$^t$'s catalytic subunit is functional, and if it is predicted to bind its *t*-specific or non-*t*-specific repressors. In turn, the repressors themselves are regulated by phosphorylation at known residues [18], and *Ppp1r11*$^t$'s mutations were previously predicted to affect its phosphorylation and PPP1-binding [22]. The fast evolution [37] and differential expression of these three functionally related genes on the *t*-haplotype raises the possibility that their protein structures co-evolved to alter sperm motility. *In silico* predictions of protein structures, functions and interactions would ultimately provide a basis for conducting experimental validation of PPP1-involvment in drive.

Overall, our study provided new insights into the timing of gene expression aberrations in $+/t$ testes and highlighted the misexpression of many genes outside the *t* complex. The most significantly differentially expressed genes in our study supported the hypothesis that structural and motor proteins of the sperm flagella as well as the PPP1-centric dephosphorylation pathway may be involved in drive. The strong differential expression of five dynein genes and two radial-spoke-encoding genes at the early stages of spermatogenesis supports them being "poison" genes, in line with their location in the distorter regions of the *t*-haplotype. Although the location of the *t*-specific *Ppp1cb* is not known, our results uncovered that it may be an early-acting "poison", just like its inhibitor, *Ppp1r2*, which is encoded in a distorter region of the *t*-haplotype. On the other hand, *Ppp1r11*'s strong late-stage overexpression in *t* spermatids raises the question if this distorter-region-encoded protein could function as an "antidote". Our results uncovered new candidate genes that may underlie the impaired motility and fertilization phenotypes of sperm from $+/t$ mice, such as key genes related to sperm energy production (LDHC and mitochondrial proteins) and zona-pellucida-binding (CCT complex genes and ZPBP2). This was the first study of the expression changes induced by a meiotic driver at the single cell level, a resolution that provides unprecedented and much awaited insights into the complex biology of drive. We provided a single-cell-atlas of gene expression in testes carrying the *t*-haplotype, a resource for future research into its transmission ratio distortion as well as into sperm function.

## 4.4 Materials and methods

### Single nucleus sequencing

We isolated nuclei from testis samples of three 99-day-old *M. m. domesticus* mice of the ILL strain which originating from Illnau, Switzerland, and was maintained in the lab for ten

generations. Mice were full siblings raised under standardized conditions: two were of $+/t^{wILL}$ genotype (tested by PCR amplification of the *t*-specific allele of *Hba-ps4*), and one of $+/+$ genotype. Nuclei were isolated using the 10x Genomics nuclei isolation kit, and sequenced using Illumina Novaseq S4 PE150 XP.

## Finding high-quality nuclei

We created a pseudo-*t*-haplotype reference sequence by substituting *t*-specific genomic SNPs found by Kelemen *et al.* [38] of a *M. m. domesticus* mouse from France sampled by Harr *et al.* [23]. We appended this sequence to the *mm10* mouse reference genome. We annotated the pseudo-*t*-haplotype sequence by finding the coordinates of the best BLAT hits reference *t* complex exons. CellRanger (version 7.1.0) was used to map and count the UMIs (unique molecular identifiers) for each gene and barcode in a library. The fraction of intronic reads per nucleus was calculated with the nuclear_fraction_tag function of the DropletQC library in R (version 4.3.2). We used CellBender (version 0.3.0) to remove barcodes that were predicted not to represent nuclei but ambient RNA, and reads from each predicted nucleus, which were predicted to be derived from ambient RNA. Using the Seurat package (version 5.0.1) in R we filtered for nuclei with $<$5% mitochondrial UMI counts, $>$300 genes expressed, $>$300 UMI counts and an intronic read fraction $>$0.55.

## Cell type identification

We clustered nuclei within each sample based on gene expression outside of the *t* complex (chr. 17:3-42 Mb) using Seurat, and removed suspected doublets (barcodes representing two nuclei) using the DoubletFinder package (version 2.4.0) in R. Then we integrated all samples using the FindIntegrationAnchors and IntegrateData functions of Seurat by taking the 10,000 most variable genes, outside of the *t* complex for anchoring. After clustering and dimenstionality reduction using the RunUmap function we plotted the mean cluster expression of marker genes [45, 19] to annotate the clusters. We removed two clusters that showed haploid spermatid gene expression patterns, but significantly lower intronic read fractions than other haploid clusters and unimodal X chromosome expression distribution, indicative of whole cells (Supplementary figure 4.6).

## Identification of *t*- and +-spermatids

We computed a preliminary *t*-score for each nucleus by subtracting the mean + allele expression from that of the *t* allele expression of each expressed *t* complex gene. We fit two normal distributions to the resulting bimodal distributions of *t*-scores in the haploid cell clusters of the $+/t$ samples using the *mix* function of the *mixdist* package (version 0.5-5) in R with parameters mixparam(mu=c(-0.02,0.02), sigma=c(0.01,0.01)), where *mu* parameters were estimated from the data using the *multimode* package (version 1.05) in R. Nuclei with *t*-scores above the 10th percentiles of the high-*t*-scores distributions were tested against nuclei below the 90th percentiles of the low-*t*-scores distributions for differential expression of *t* complex gene alleles using the MAST test of the FindMarkers function of Seurat. We used the 25th and 75th percentiles for the early round spermatid 1 cluster, because of less bimodal *t*-scores (Supplementary figure 4.9). Genes with differentially expressed alleles and an FDR-adjusted p-value below 0.05 were used to compute a refined *t*-score using the same formula as above. We fit two normal distributions to the bimodal *t*-score distributions in each haploid cluster. Nuclei above the 10th percentile of the high-*t*-score distribution were classified as *t*-spermatids, and

those below the 90th percentile of the low-*t*-score distribution were classified as +-spermatids. We used the 25th and the 75th percentiles, respectively, for the early round spermatids 1.

## Assembling transcriptomes of *t*-spermatids and +/+ testes

We extracted reads from the raw FASTQ files of the two $+/t$ samples using umi_tools' *extract* function in Python (version 3.9.7) and the list of our *t*-spermatid barcodes. We excluded nuclei from the early round spermatid 1 cluster, because their *t*-score distributions were less bimodal than the later clusters (Supplementary figure 4.10). Reads were trimmed with *fqtrim* (version 0.9.7) with parameters -5 AAGCAGTGGTATCAACGCAGAGTACATGGG (template switch oligo sequence) -a 10 -q 20 -l 50. Files containing read 2's from both $+/$t samples were concatenated and assembled in a forward stranded mode (–SS_lib_type F) using Trinity (version 2.15.1). Reads from the $+/+$ library were trimmed with the above parameters and read 2's were assembled using the same procedure as mentioned above. We filtered for minimum 500 basepair long contigs with faFilter (version v.357), and retained contigs that showed a minimum TPM of 0.5, estimated using kallisto (version 0.50.1), in all libraries of the focal genotype. Contigs with an overlap of at least 100 basepairs and less than 0.01 sequence differences, as inferred by BLAT, were collapsed by choosing the longest of such matching contigs using a custom-made Perl script. The *t*-spermatid assembly and the $+/+$ testis assembly were concatenated at this point, and contigs with an overlap of at least 100 basepairs and less than 0.01 sequence differences, as inferred by BLAT, were collapsed by choosing the longest of such matching contigs using a custom-made Perl script, as before. We mapped each contig to the *mm10* transcriptome and to the *t*-specific genes *Ppp1cb*$^t$, *Rnpepl1*$^t$ and *Smok*$^{Tcr}$, identified in [37] and [27], and we assigned each to a gene if it had a minimum of 200 basepair alignment to one of its transcripts.

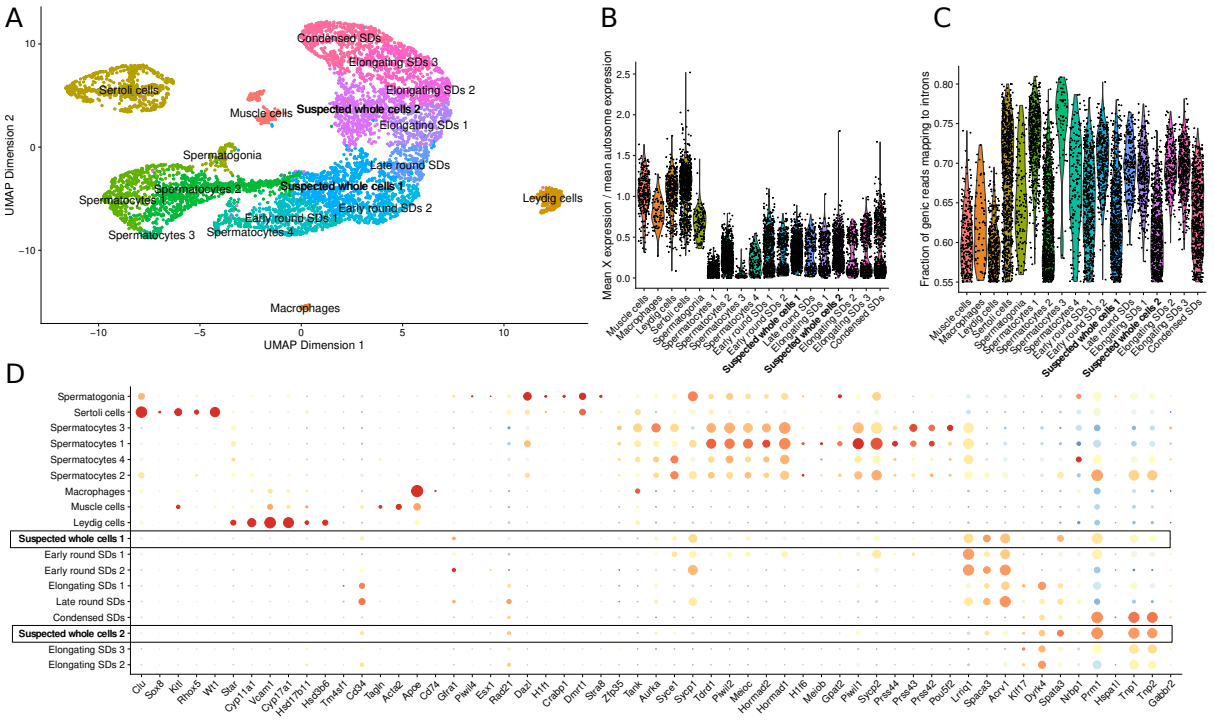
## Differential expression analysis

We quantified the expression of each contig using CellRanger. Then for each gene in our assembly and high-quality nucleus identified before we added up the UMI counts of the contigs mapping to it, and created three Seurat objects, which we integrated using the FindIntegrationAnchors and IntegrateData functions of Seurat by taking the 10,000 most variable genes outside of the *t* complex for anchoring. We then transferred the cluster information for each nucleus from the reference-mapping-based Seurat object to the assembly-mapping-based object, and also the previously inferred genotype information for the haploid spermatids. We then used the built-in MAST test of the FindMarkers function of Seurat to find differentially expressed genes between the two $+/t$ and the single $+/+$ sample within each diploid cell cluster. For haploid cell clusters we compared $+^{+/t}$ spermatids to $+^{+/+}$ spermatids, $t^{+/t}$ spermatids to $+^{+/+}$ spermatids and $t^{+/t}$ spermatids to $+^{+/t}$ spermatids. We considered only genes with an adjusted p-value below 0.05 to be differentially expressed.

## Gene ontology enrichment analysis

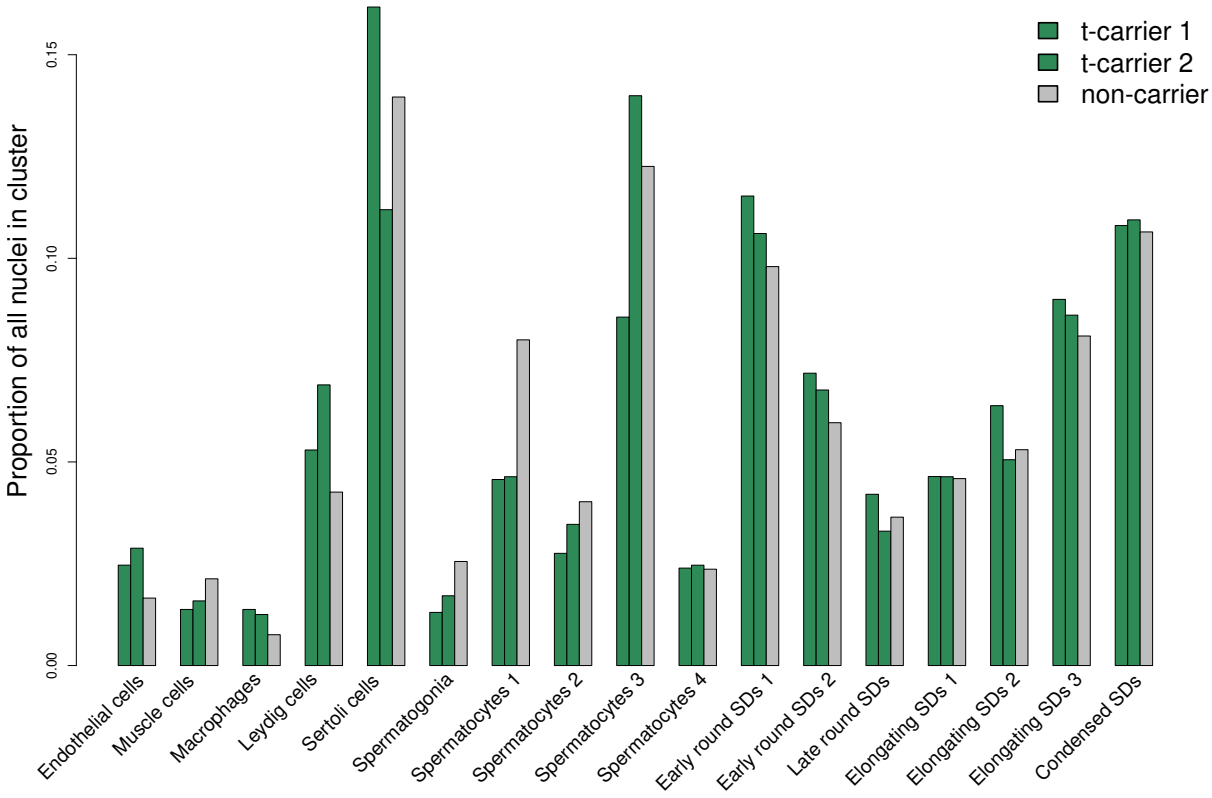
We provided our list of differentially expressed genes to the online database and gene ontology search tool, STRING (https://string-db.org/, version 12.0) to identify enriched gene ontology categories.

## 4.5 Supplementary figures

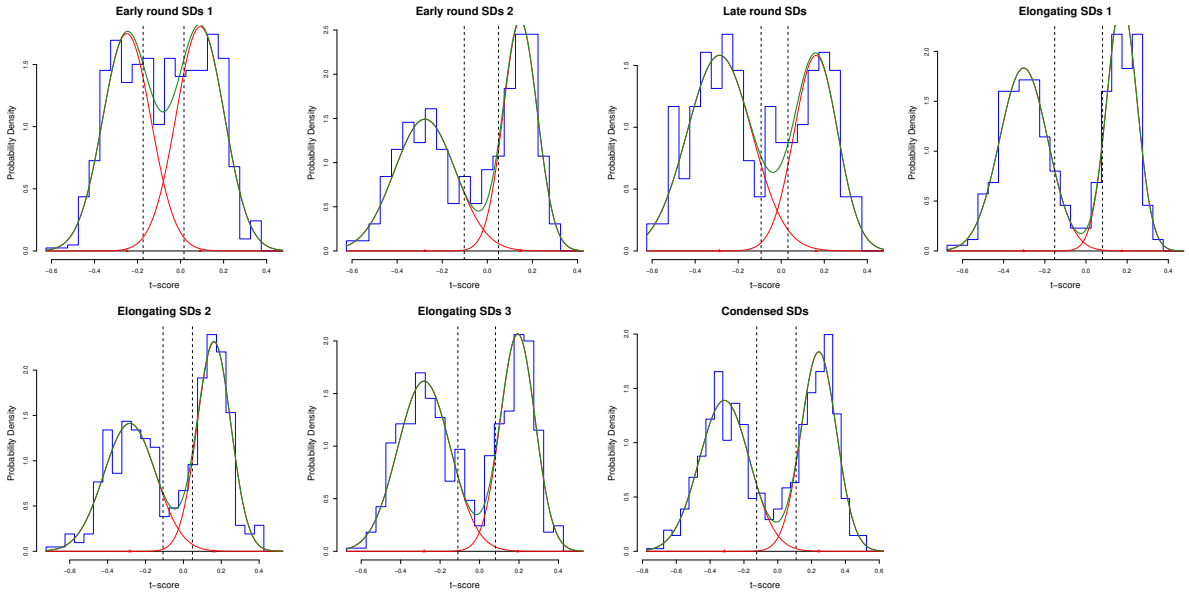



Supplementary Figure 4.6: Supporting evidence for removing two clusters that contain suspected whole cells. (A) UMAP plot showing the nuclei in a space that summarizes their high-dimensional gene expression information in two dimensions. Clusters were named based on their expression of cell-type marker genes [45, 19]. Two clusters were denoted as "suspected whole cells" because they expressed haploid spermatid marker genes (D), but they did not show clearly bimodal X chromosome to autosome expression ratios (B), and a lower fraction of their genic reads mapped to introns than seen in other haploid spermatid clusters (C).

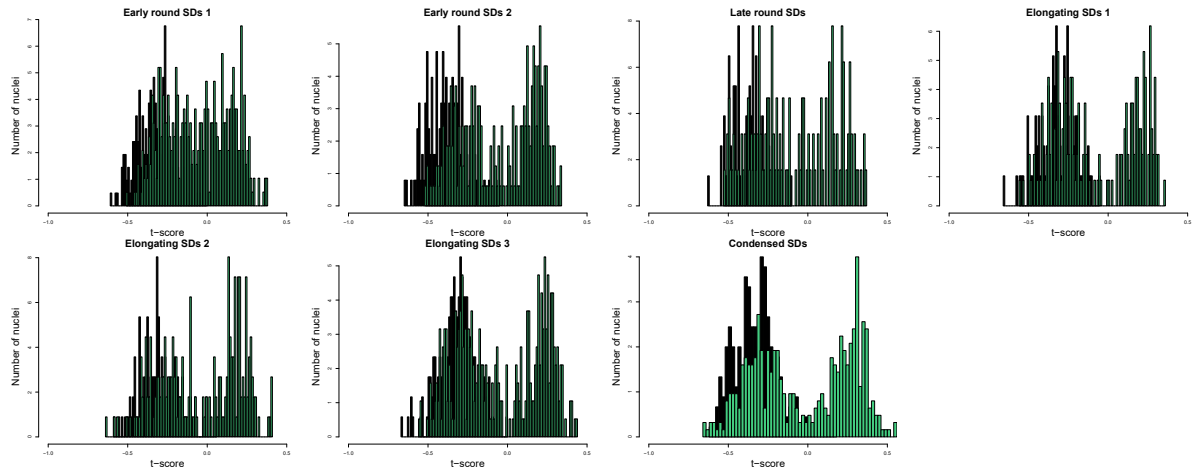

Supplementary Figure 4.7: Expression of known testis cell type marker genes [45, 19] in our clusters of nuclei.



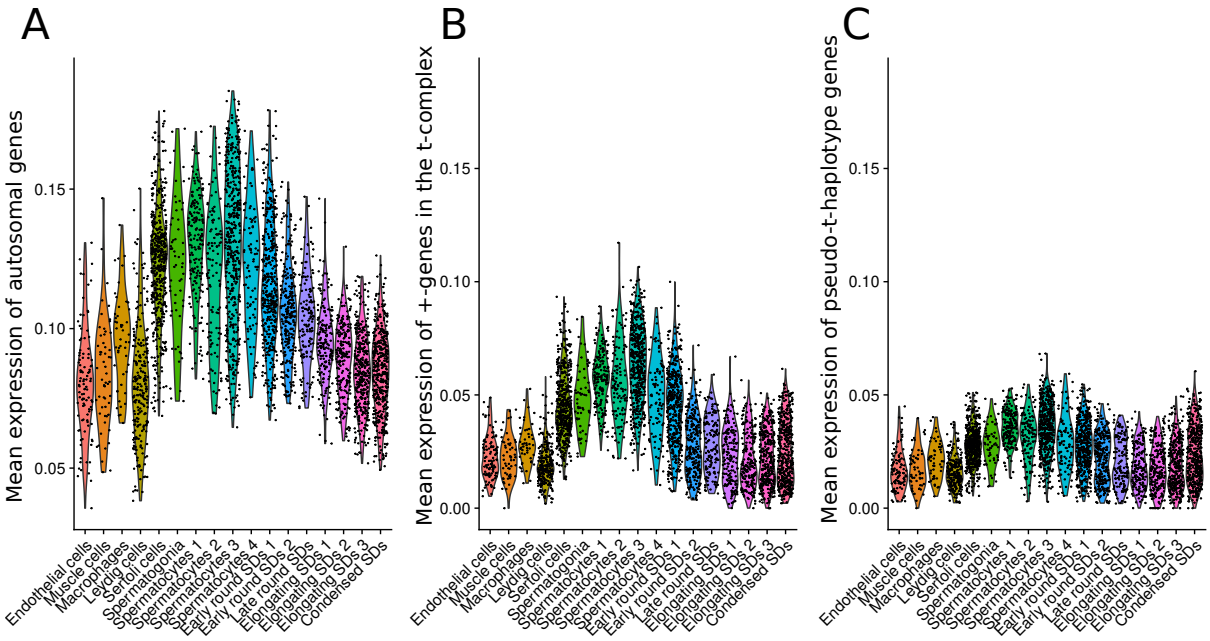

Supplementary Figure 4.8: The fraction of nuclei belonging to each cluster in each of the three samples.



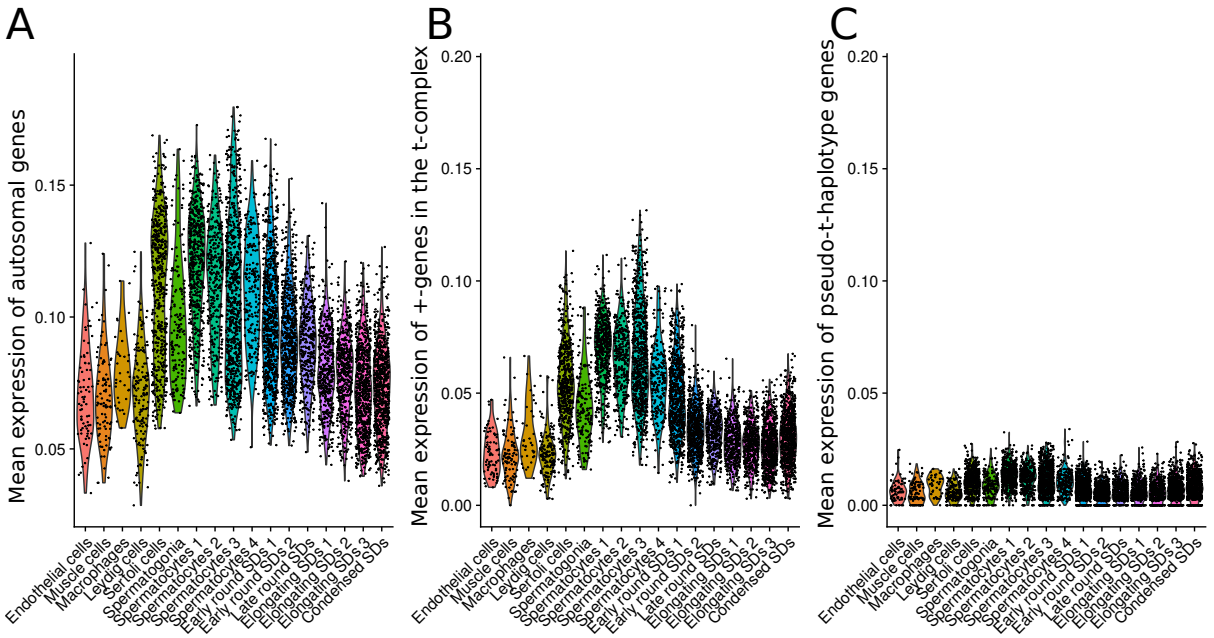

Supplementary Figure 4.9: Distributions of $t$-scores in haploid clusters of $+/t$ samples and normal distributions fitted to them. Blue histograms show the frequency of t-scores. Red smooth lines represent the two normal distributions fitted to the bimodal distributions, while the green smooth lines show the bimodal distribution fitted to the data. Dotted lines mark the 90th and 10th percentiles of the fitted low-$t$-score and high-$t$-score dsitributions, respectively, which were used to call spermatid genotypes. Nuclei in the "Early round spermatids 1" cluster showed a more unimodal distribution, therefore the 25th and 75th percentiles were used for genotype calling.



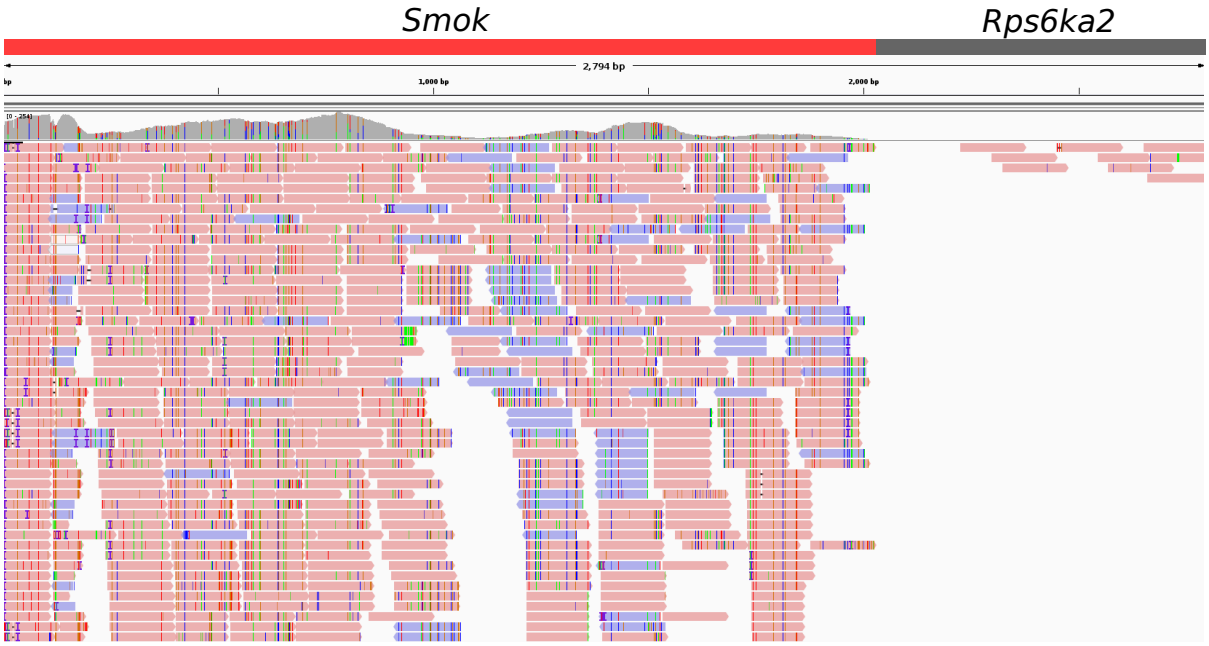

Supplementary Figure 4.10: Distributions of $t$-scores in haploid clusters of $+/t$ (green) samples and the $+/+$ (black) sample.




Supplementary Figure 4.11: Mean expression of autosomal (A), $+$-$t$-complex (B) and pseudo-$t$-haplotype genes in the different cell types of $+/t$ mice (two samples pooled).




Supplementary Figure 4.12: Mean expression of autosomal (A), +-*t*-complex (B) and pseudo-*t*-haplotype genes in the different cell types of the +/+ mouse.




Supplementary Figure 4.13: The alignment of reads from *t*-spermatids to $Smok^{Tcr}$ from Herrmann *et al.* [27], as visualized by the Integrated Genome Viewer (IGV). There are no reads supporting the *Smok-Rps6ka2* fusion at basepair number 2023.

| Gene ontology ID | Biological process | Observed gene count | Background gene count | Strength | FDR |
|---|---|---|---|---|---|
| GO:0006119 | Oxidative phosphorylation | 16 | 120 | 0.98 | 9.18e-07 |
| GO:0009060 | Aerobic respiration | 17 | 168 | 0.86 | 5.79e-06 |
| GO:0015980 | Energy derivation by oxidation of organic compounds | 21 | 275 | 0.74 | 5.79e-06 |
| GO:0006091 | Generation of precursor metabolites and energy | 24 | 378 | 0.66 | 6.83e-06 |
| GO:0007276 | Gamete generation | 40 | 982 | 0.47 | 6.83e-06 |
| GO:0019953 | Sexual reproduction | 43 | 1115 | 0.45 | 6.83e-06 |
| GO:0045333 | Cellular respiration | 18 | 204 | 0.8 | 6.83e-06 |
| GO:0042776 | Proton motive force-driven mitochondrial ATP synthesis | 11 | 62 | 1.11 | 8.43e-06 |
| GO:0007283 | Spermatogenesis | 33 | 726 | 0.52 | 8.51e-06 |
| GO:0006754 | ATP biosynthetic process | 12 | 85 | 1.01 | 1.34e-05 |
| GO:0009142 | Nucleoside triphosphate biosynthetic process | 13 | 115 | 0.91 | 2.78e-05 |
| GO:0022414 | Reproductive process | 56 | 1844 | 0.34 | 3.30e-05 |
| GO:0044085 | Cellular component biogenesis | 69 | 2631 | 0.28 | 0.00012 |
| GO:0006164 | Purine nucleotide biosynthetic process | 15 | 194 | 0.75 | 0.00014 |
| GO:0003006 | Developmental process involved in reproduction | 42 | 1271 | 0.38 | 0.00017 |
| GO:0022607 | Cellular component assembly | 64 | 2396 | 0.29 | 0.00017 |
| GO:0042773 | ATP synthesis coupled electron transport | 10 | 75 | 0.98 | 0.00017 |
| GO:0022900 | Electron transport chain | 12 | 125 | 0.84 | 0.00024 |
| GO:0022904 | Respiratory electron transport chain | 11 | 102 | 0.89 | 0.00024 |
| GO:0046390 | Ribose phosphate biosynthetic process | 15 | 209 | 0.72 | 0.00027 |
| GO:0009152 | Purine ribonucleotide biosynthetic process | 14 | 185 | 0.74 | 0.00035 |
| GO:0043933 | Protein-containing complex organization | 43 | 1388 | 0.35 | 0.00047 |
| GO:0009165 | Nucleotide biosynthetic process | 16 | 253 | 0.66 | 0.00049 |
| GO:0046034 | ATP metabolic process | 13 | 165 | 0.76 | 0.00050 |
| GO:0007281 | Germ cell development | 22 | 468 | 0.53 | 0.00053 |
| GO:0042775 | Mitochondrial ATP synthesis coupled electron transport | 9 | 73 | 0.95 | 0.00076 |
| GO:0022412 | Cellular process involved in reproduction in multicellular organism | 25 | 608 | 0.47 | 0.00093 |
| GO:0007286 | Spermatid development | 17 | 307 | 0.6 | 0.0010 |
| GO:0060285 | Cilium-dependent cell motility | 12 | 151 | 0.76 | 0.0010 |
| GO:0009141 | Nucleoside triphosphate metabolic process | 14 | 220 | 0.66 | 0.0016 |
| GO:0065003 | Protein-containing complex assembly | 38 | 1227 | 0.35 | 0.0016 |
| GO:0006996 | Organelle organization | 79 | 3482 | 0.22 | 0.0017 |
| GO:0030317 | Flagellated sperm motility | 11 | 136 | 0.77 | 0.0020 |
| GO:0019646 | Aerobic electron transport chain | 8 | 65 | 0.95 | 0.0022 |
| GO:0003341 | Cilium movement | 13 | 198 | 0.68 | 0.0023 |
| GO:0007018 | Microtubule-based movement | 18 | 412 | 0.5 | 0.0084 |
| GO:0032981 | Mitochondrial respiratory chain complex I assembly | 7 | 59 | 0.93 | 0.0092 |
| GO:0070925 | Organelle assembly | 26 | 790 | 0.38 | 0.0164 |
| GO:0006457 | Protein folding | 11 | 178 | 0.65 | 0.0165 |
| GO:0001675 | Acrosome assembly | 5 | 27 | 1.13 | 0.0183 |
| GO:0044782 | Cilium organization | 16 | 362 | 0.5 | 0.0183 |
| GO:0006163 | Purine nucleotide metabolic process | 17 | 404 | 0.48 | 0.0191 |
| GO:0060271 | Cilium assembly | 15 | 325 | 0.52 | 0.0191 |
| GO:0061077 | Chaperone-mediated protein folding | 7 | 69 | 0.87 | 0.0202 |
| GO:0033108 | Mitochondrial respiratory chain complex assembly | 8 | 98 | 0.77 | 0.0256 |
| GO:0090407 | Organophosphate biosynthetic process | 19 | 503 | 0.44 | 0.0271 |
| GO:0019693 | Ribose phosphate metabolic process | 17 | 422 | 0.46 | 0.0294 |
| GO:0071840 | Cellular component organization or biogenesis | 108 | 5681 | 0.14 | 0.0308 |
| GO:0035082 | Axoneme assembly | 8 | 102 | 0.75 | 0.0312 |
| GO:1903333 | Negative regulation of protein folding | 3 | 5 | 1.64 | 0.0312 |
| GO:1905198 | Manchette assembly | 3 | 5 | 1.64 | 0.0312 |
| GO:0120031 | Plasma membrane bounded cell projection assembly | 17 | 427 | 0.46 | 0.0313 |
| GO:0007017 | Microtubule-based process | 26 | 840 | 0.35 | 0.0332 |
| GO:0046661 | Male sex differentiation | 11 | 199 | 0.6 | 0.0332 |
| GO:0009150 | Purine ribonucleotide metabolic process | 16 | 390 | 0.47 | 0.0335 |
| GO:0043486 | Histone exchange | 5 | 33 | 1.04 | 0.0335 |
| GO:0051085 | Chaperone cofactor-dependent protein refolding | 5 | 33 | 1.04 | 0.0335 |
| GO:0001578 | Microtubule bundle formation | 9 | 137 | 0.68 | 0.0372 |
| GO:0009117 | Nucleotide metabolic process | 18 | 484 | 0.43 | 0.0400 |
| GO:0008584 | Male gonad development | 10 | 173 | 0.62 | 0.0427 |

Supplementary Table 4.1: Significantly enriched biological processes among the 377 differentially expressed genes in the seminiferous tubules. FDR - False discovery rate

## 4.6 Supplementary tables

| Gene ontology ID | Biological process | Observed gene count | Background gene count | Strength | FDR |
|---|---|---|---|---|---|
| GO:0098800 | Inner mitochondrial membrane protein complex | 18 | 149 | 0.94 | 3.27e-08 |
| GO:0005622 | Intracellular anatomical structure | 251 | 14759 | 0.09 | 1.05e-06 |
| GO:0005746 | Mitochondrial respirasome | 13 | 90 | 1.02 | 1.07e-06 |
| GO:0098803 | Respiratory chain complex | 13 | 90 | 1.02 | 1.07e-06 |
| GO:0005737 | Cytoplasm | 212 | 11788 | 0.11 | 2.59e-06 |
| GO:0098798 | Mitochondrial protein-containing complex | 20 | 297 | 0.69 | 4.76e-06 |
| GO:0043227 | Membrane-bounded organelle | 215 | 12435 | 0.1 | 6.89e-05 |
| GO:0043226 | Organelle | 227 | 13445 | 0.09 | 0.00010 |
| GO:0031967 | Organelle envelope | 42 | 1324 | 0.36 | 0.00014 |
| GO:0043229 | Intracellular organelle | 222 | 13119 | 0.09 | 0.00014 |
| GO:0005747 | Mitochondrial respiratory chain complex I | 8 | 50 | 1.06 | 0.00018 |
| GO:0005740 | Mitochondrial envelope | 31 | 854 | 0.42 | 0.00021 |
| GO:0005743 | Mitochondrial inner membrane | 24 | 556 | 0.49 | 0.00021 |
| GO:0019866 | Organelle inner membrane | 25 | 617 | 0.47 | 0.00033 |
| GO:0031514 | Motile cilium | 16 | 282 | 0.61 | 0.00040 |
| GO:0005929 | Cilium | 28 | 773 | 0.42 | 0.00058 |
| GO:0031966 | Mitochondrial membrane | 28 | 795 | 0.41 | 0.00091 |
| GO:0043232 | Intracellular non-membrane-bounded organelle | 99 | 4826 | 0.17 | 0.0013 |
| GO:0005739 | Mitochondrion | 50 | 1956 | 0.27 | 0.0019 |
| GO:0098796 | Membrane protein complex | 36 | 1230 | 0.33 | 0.0021 |
| GO:0015630 | Microtubule cytoskeleton | 39 | 1390 | 0.31 | 0.0022 |
| GO:0005838 | Proteasome regulatory particle | 5 | 23 | 1.2 | 0.0026 |
| GO:0043231 | Intracellular membrane-bounded organelle | 199 | 11954 | 0.08 | 0.0043 |
| GO:1990204 | Oxidoreductase complex | 9 | 122 | 0.73 | 0.0055 |
| GO:0045261 | Proton-transporting ATP synthase complex, catalytic core F(1) | 3 | 6 | 1.56 | 0.0131 |
| GO:0001669 | Acrosomal vesicle | 10 | 180 | 0.6 | 0.0189 |
| GO:0070069 | Cytochrome complex | 5 | 40 | 0.96 | 0.0222 |
| GO:0005829 | Cytosol | 83 | 4213 | 0.15 | 0.0230 |
| GO:1902494 | Catalytic complex | 39 | 1582 | 0.25 | 0.0230 |
| GO:0000502 | Proteasome complex | 6 | 67 | 0.81 | 0.0276 |
| GO:0005751 | Mitochondrial respiratory chain complex IV | 4 | 25 | 1.06 | 0.0341 |
| GO:0097729 | 9+2 motile cilium | 10 | 199 | 0.56 | 0.0341 |

Supplementary Table 4.2: Significantly enriched cellular components among the 377 differentially expressed genes in the seminiferous tubules. FDR - False discovery rate

## 4.7 Bibliography


[1] A. Abou-haila and D. R. Tulsiani. Signal transduction pathways that regulate sperm capacitation and the acrosome reaction. *Archives of biochemistry and biophysics*, 485(1):72–81, 2009.

[2] A. Amaral and B. G. Herrmann. RAC1 controls progressive movement and competitiveness of mammalian spermatozoa. *PLoS Genetics*, 17(2):e1009308, 2021.

[3] H. Bauer, S. Schindler, Y. Charron, J. Willert, B. Kusecek, and B. G. Herrmann. The nucleoside diphosphate kinase gene Nme3 acts as quantitative trait locus promoting non-Mendelian inheritance. *PLoS genetics*, 8(3):e1002567, 2012.

[4] H. Bauer, N. Véron, J. Willert, and B. G. Herrmann. The t-complex-encoded guanine nucleotide exchange factor Fgd2 reveals that two opposing signaling pathways promote transmission ratio distortion in the mouse. *Genes & development*, 21(2):143–147, 2007.

[5] H. Bauer, J. Willert, B. Koschorz, and B. G. Herrmann. The t complex–encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nature genetics*, 37(9):969–973, 2005.

[6] J. Brown, J. A. Cebra-Thomas, J. D. Bleil, P. M. Wassarman, and L. M. Silver. A premature acrosome reaction is programmed by mouse t haplotypes during sperm differentiation and could play a role in transmission ratio distortion. *Development*, 106(4):769–773, 1989.

[7] D. Bruck. Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Proceedings of the National Academy of Sciences*, 43(1):152–158, 1957.

[8] Y. Charron, J. Willert, B. Lipkowitz, B. Kusecek, B. G. Herrmann, and H. Bauer. Two isoforms of the RAC-specific guanine nucleotide exchange factor TIAM2 act oppositely on transmission ratio distortion by the mouse t-haplotype. *PLoS Genetics*, 15(2):e1007964, 2019.

[9] Y. Cheng, M. G. Buffone, M. Kouadio, M. Goodheart, D. C. Page, G. L. Gerton, I. Davidson, and P. J. Wang. Abnormal sperm in mice lacking the Taf7l gene. *Molecular and cellular biology*, 27(7):2582–2589, 2007.

[10] C. Courret, C.-H. Chang, K. H.-C. Wei, C. Montchamp-Moreau, and A. M. Larracuente. Meiotic drive mechanisms: lessons from Drosophila. *Proceedings of the Royal Society B*, 286(1913):20191430, 2019.

[11] R. K. Dawe. The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Research*, 30(2-3):205–216, 2022.

[12] R. K. Dawe, E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, D. M. Higgins, J. Ross-Ibarra, J. G. Wallace, L. B. Kanizay, M. Alabady, et al. A kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*, 173(4):839–850, 2018.

[13] M. D. Dun, N. D. Smith, M. A. Baker, M. Lin, R. J. Aitken, and B. Nixon. The chaperonin containing TCP1 complex (CCT/TRiC) is involved in mediating sperm-oocyte interaction. *Journal of Biological Chemistry*, 286(42):36875–36887, 2011.

[14] K. A. Dyer, B. Charlesworth, and J. Jaenike. Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proceedings of the National Academy of Sciences*, 104(5):1587–1592, 2007.

[15] A. Ferramosca, S. P. Provenzano, L. Coppola, and V. Zara. Mitochondrial respiratory efficiency is positively correlated with human sperm motility. *Urology*, 79(4):809–814, 2012.

[16] M. J. Freitas, S. Vijayaraghavan, and M. Fardilha. Signaling mechanisms in mammalian sperm motility. *Biology of Reproduction*, 96(1):2–12, 2017.

[17] S. Goffart and R. Wiesner. Regulation and co-ordination of nuclear gene expression during mitochondrial biogenesis. *Experimental physiology*, 88(1):33–40, 2003.

[18] S. Goswami, L. Korrodi-Gregório, N. Sinha, S. Bhutada, R. Bhattacharjee, D. Kline, and S. Vijayaraghavan. Regulators of the protein phosphatase PP1$\gamma$2, PPP1R2, PPP1R7, and PPP1R11 are involved in epididymal sperm maturation. *Journal of cellular physiology*, 234(3):3105–3118, 2019.

[19] C. D. Green, Q. Ma, G. L. Manske, A. N. Shami, X. Zheng, S. Marini, L. Moritz, C. Sultan, S. J. Gurczynski, B. B. Moore, et al. A comprehensive roadmap of murine spermatogenesis defined by single-cell RNA-seq. *Developmental cell*, 46(5):651–667, 2018.

[20] M. D. Griswold. 50 years of spermatogenesis: Sertoli cells and their interactions with germ cells. *Biology of Reproduction*, 99(1):87–100, 2018.

[21] G. R. Gummere, P. J. McCormick, and D. Bennett. The influence of genetic background and the homologous chromosome 17 on t-haplotype transmission ratio distortion in mice. *Genetics*, 114(1):235–245, 1986.

[22] Y. Han, X.-X. Song, H.-L. Feng, C.-K. Cheung, P.-M. Lam, C.-C. Wang, and C. J. Haines. Mutations of t-complex testis expressed gene 5 transcripts in the testis of sterile t-haplotype mutant mouse. *Asian journal of andrology*, 10(2):219–226, 2008.

[23] B. Harr, E. Karakoc, R. Neme, M. Teschke, C. Pfeifle, Ž. Pezer, H. Babiker, M. Linnenbrink, I. Montero, R. Scavetta, et al. Genomic resources for wild populations of the house mouse, Mus musculus and its close relative Mus spretus. *Scientific data*, 3(1):1–14, 2016.

[24] Q. Helleu, P. R. Gérard, R. Dubruille, D. Ogereau, B. Prudhomme, B. Loppin, and C. Montchamp-Moreau. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy of Sciences*, 113(15):4110–4115, 2016.

[25] S. Henikoff, K. Ahmad, and H. S. Malik. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, 293(5532):1098–1102, 2001.

[26] T. Hereng, K. Elgstøen, F. Cederkvist, L. Eide, T. Jahnsen, B. Skålhegg, and K. Rosendal. Exogenous pyruvate accelerates glycolysis and promotes capacitation in human spermatozoa. *Human Reproduction*, 26(12):3249–3263, 2011.

[27] B. G. Herrmann, B. Koschorz, K. Wertz, K. J. McLaughlin, and A. Kispert. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature*, 402(6758):141–146, 1999.

[28] N. Hillman and M. Nadijcka. A comparative study of spermiogenesis in wild-type and T: t-bearing mice. *Development*, 44(1):243–261, 1978.

[29] Y. Hiraizumi, L. Sandler, and J. F. Crow. Meiotic drive in natural populations of Drosophila melanogaster. iii. Populational implications of the segregation-distorter locus. *Evolution*, pages 433–444, 1960.

[30] D. Huang, Y. Zuo, C. Zhang, G. Sun, Y. Jing, J. Lei, S. Ma, S. Sun, H. Lu, Y. Cai, et al. A single-nucleus transcriptomic atlas of primate testicular aging reveals exhaustion of the spermatogonial stem cell reservoir and loss of Sertoli cell homeostasis. *Protein & Cell*, 14(12):888, 2023.

[31] L. Hui, J. Lu, Y. Han, and S. H. Pilder. The mouse T complex gene Tsga2, encoding polypeptides located in the sperm tail and anterior acrosome, maps to a locus associated with sperm motility and sperm-egg interaction abnormalities. *Biology of reproduction*, 74(4):633–643, 2006.

[32] L.-Y. Huw, A. S. Goldsborough, K. Willison, and K. Artzt. Tctex2: a sperm tail surface protein mapping to the t-complex. *Developmental Biology*, 170(1):183–194, 1995.

[33] K. Inaba, O. Kagami, and K. Ogawa. Tctex2-related outer arm dynein light chain is phosphorylated at activation of sperm motility. *Biochemical and Biophysical Research Communications*, 256(1):177–183, 1999.

[34] M. Jodar, S. Kalko, J. Castillo, J. L. Ballescà, and R. Oliva. Differential RNAs in the sperm cells of asthenozoospermic patients. *Human reproduction*, 27(5):1431–1438, 2012.

[35] L. Johnson, S. Pilder, J. Bailey, and P. Olds-Clarke. Sperm from mice carrying one or two t haplotypes are deficient in investment and oocyte penetration. *Developmental Biology*, 168(1):138–149, 1995.

[36] D. F. Katz, R. P. Erickson, and M. Nathanson. Beat frequency is bimodally distributed in spermatozoa from T/t12 mice. *Journal of Experimental Zoology*, 210(3):529–535, 1979.

[37] R. K. Kelemen, M. Elkrewi, A. K. Lindholm, and B. Vicoso. Novel patterns of expression and recruitment of new genes on the t-haplotype, a mouse selfish chromosome. *Proceedings of the Royal Society B*, 289(1968):20211985, 2022.

[38] R. K. Kelemen and B. Vicoso. Complex history and differentiation patterns of the t-haplotype, a mouse meiotic driver. *Genetics*, 208(1):365–375, 2018.

[39] A. N. Kruger and J. L. Mueller. Mechanisms of meiotic drive in symmetric and asymmetric meiosis. *Cellular and Molecular Life Sciences*, 78:3205–3218, 2021.

[40] E. Lader, H.-S. Ha, M. O'Neill, K. Artzt, and D. Bennett. tctex-1: a candidate gene family for a mouse t complex sterility locus. *Cell*, 58(5):969–979, 1989.

[41] A. M. Larracuente and D. C. Presgraves. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics*, 192(1):33–53, 2012.

[42] Y.-N. Lin, A. Roy, W. Yan, K. H. Burns, and M. M. Matzuk. Loss of zona pellucida binding proteins in the acrosomal matrix disrupts acrosome biogenesis and sperm morphogenesis. *Molecular and cellular biology*, 27(19):6794–6805, 2007.

[43] C. B. Lindemann and K. A. Lesich. Flagellar and ciliary beating: the proven and the possible. *Journal of cell science*, 123(4):519–528, 2010.

[44] A. Lindholm, A. Sutter, S. Künzel, D. Tautz, and H. Rehrauer. Effects of a male meiotic driver on male and female transcriptomes in the house mouse. *Proceedings of the Royal Society B*, 286(1915):20191927, 2019.

[45] S. Lukassen, E. Bosch, A. B. Ekici, and A. Winterpacht. Characterization of germ cell differentiation in the male mouse through single-cell RNA sequencing. *Scientific reports*, 8(1):6521, 2018.

[46] M. F. Lyon. Transmission ratio distortion in mice. *Annual review of genetics*, 37(1):393–408, 2003.

[47] E. J. Mange. Temperature sensitivity of segregation-distortion in Drosophila melanogaster. *Genetics*, 58(3):399, 1968.

[48] A. Manser, B. König, and A. K. Lindholm. Polyandry blocks gene drive in a wild house mouse population. *Nature communications*, 11(1):5590, 2020.

[49] O. Meikar, V. V. Vagin, F. Chalmel, K. Sõstar, A. Lardenois, M. Hammell, Y. Jin, M. Da Ros, K. A. Wasik, J. Toppari, et al. An atlas of chromatoid body components. *Rna*, 20(4):483–495, 2014.

[50] F.-D. Ni, S.-L. Hao, and W.-X. Yang. Multiple signaling pathways in Sertoli cells: recent findings in spermatogenesis. *Cell death & disease*, 10(8):541, 2019.

[51] F. Odet, S. Gabel, R. E. London, E. Goldberg, and E. M. Eddy. Glycolysis and mitochondrial respiration in mouse LDHC-null sperm. *Biology of reproduction*, 88(4):95–1, 2013.

[52] P. Olds. Effect of the T locus on sperm distribution in the house mouse. *Biology of Reproduction*, 2(1):91–97, 1970.

[53] G. Ostergren. Parasitic nature of extra fragment chromosomes. *Bot. Not.*, 2:157–163, 1945.

[54] L. ODonnell, L. B. Smith, and D. Rebourcet. Sertoli cells as key drivers of testis function. In *Seminars in Cell & Developmental Biology*, volume 121, pages 2–9. Elsevier, 2022.

[55] S. Pilder, J. Lu, Y. Han, L. Hui, S. Samant, O. Olugbemiga, K. Meyers, L. Cheng, and S. Vijayaraghavan. The molecular basis of" curlicue": a sperm motility abnormality linked to the sterility of t haplotype homozygous male mice. *Society of Reproduction and Fertility supplement*, 63:123–133, 2007.

[56] S. H. Pilder. Does dynein influence the non-mendelian inheritance of chromosome 17 homologs in male mice? In *Dyneins*, pages 538–559. Elsevier, 2012.

[57] S. H. PILDER, P. OLDS-CLARKE, J. M. ORTH, W. F. JESTER, and L. DUGAN. Hst7: a male sterility mutation perturbing sperm motility, flagellar assembly, and mitochondrial sheath differentiation. *Journal of andrology*, 18(6):663–671, 1997.

[58] T. A. R. Price, R. Verspoor, and N. Wedell. Ancient gene drives: an evolutionary paradox. *Proc Biol Sci*, 286(1917):20192267, Dec 2019.

[59] T. Prout. Some effects of variations in the segregation ratio and of selection on the frequency of alleles under random mating. *Acta Genetica Et Statistica Medica*, 4(2-3):148–151, 1953.

[60] S. Rashid, P. Grzmil, J.-D. Drenckhahn, A. Meinhardt, I. Adham, W. Engel, and J. Neesen. Disruption of the murine dynein light chain gene Tcte3-3 results in asthenozoospermia. *Reproduction*, 139(1):99, 2010.

[61] J. A. Reinhardt, R. H. Baker, A. V. Zimin, C. Ladias, K. A. Paczolt, J. H. Werren, C. Y. Hayashi, and G. S. Wilkinson. Impacts of sex ratio meiotic drive on genome structure and function in a stalk-eyed fly. *Genome Biology and Evolution*, 15(7):evad118, 2023.

[62] A. M. Salicioni, M. D. Platt, E. V. Wertheimer, E. Arcelay, A. Allaire, J. Sosnik, P. E. Visconti, et al. Signalling pathways involved in sperm capacitation. *Society of Reproduction and Fertility supplement*, 65:245, 2007.

[63] L. Sandler and E. Novitski. Meiotic drive as an evolutionary force. *The American Naturalist*, 91(857):105–110, 1957.

[64] S. Shivaji, V. Kota, and A. B. Siva. The role of mitochondrial proteins in sperm capacitation. *Journal of reproductive immunology*, 83(1-2):14–18, 2009.

[65] B. D. Shur. Galactosyltransferase activities on mouse sperm bearing multiple tlethal and tviable haplotypes of the T/t-complex. *Genetics Research*, 38(3):225–236, 1981.

[66] N. Véron, H. Bauer, A. Y. Weiße, G. Lüder, M. Werber, and B. G. Herrmann. Retention of gene products in syncytial spermatids promotes non-mendelian inheritance as revealed by the t complex responder. *Genes & development*, 23(23):2705–2710, 2009.

[67] L. Winkler and A. K. Lindholm. A meiotic driver alters sperm form and function in house mice: a possible example of spite. *Chromosome Research*, 30(2):151–164, 2022.

[68] S. Wu, M. Yan, R. Ge, and C. Y. Cheng. Crosstalk between Sertoli and germ cells in male fertility. *Trends in molecular medicine*, 26(2):215–231, 2020.

[69] C. Zimmermann, I. Stévant, C. Borel, B. Conne, J.-L. Pitetti, P. Calvel, H. Kaessmann, B. Jégou, F. Chalmel, and S. Nef. Research resource: the dynamic transcriptional profile of sertoli cells during the progression of spermatogenesis. *Molecular endocrinology*, 29(4):627–642, 2015.

CHAPTER 5

# Discussion

## 5.1 A century of meiotic drive research

Meiotic drivers were first observed in fruit flies and house mice about a hundred years ago [13, 17], and subsequent theoretical and empirical studies uncovered the strong impacts of allelic selection on the evolution of the genomes, organisms and populations in which they are present [36, 4, 22, 43, 45]. Some meiotic drivers, such as the *wtf* element in yeast [12] or the *Paris* driver on the *Drosophila simulans* X chromosome, are characterized by repeated sweeps to fixation and by the evolution of suppressors against them, events that are expected to be commonly associated with self-promoting genetic elements [23]. On the other hand, many meiotic drivers are ancient haplotypes that are segregating at stable and relatively low equilibrium frequencies in their populations [26, 42, 28, 10]. These large haplotypes often harbor hundreds of genes in inversions that suppress recombination with the non-driving homologs [44, 5], which are expected to accumulate deleterious mutations over their long evolutionary existence [33]. How degenerate meiotic drivers are and how they persist over hundreds of thousands of years in the absence of free recombination have remained open questions.

Early molecular methods allowed the assessment of a few genes and their expression patterns, while next generation sequencing technologies provided a more comprehensive view of the sequence and expression evolution of meiotic drivers. Studies in both technological eras uncovered that fast sequence evolution [39], repeat expansions [15, 10], novel genes [10], gene amplifications [25, 32, 44, 12], gene truncations [20, 30] and fusions [21], and highly altered expression levels [25, 44, 39, 10] are common features of many meiotic drivers, as well as of the suppressor regions, which are in an arms race with the selfish chromosomes [32, 8, 19]. However, many meiotic drivers, such as the classical case of the *t*-haplotype in house mice, remained understudied in the next generation sequencing era. This thesis presented a comprehensive characterization of the sequence and expression evolution of the *t*-haplotype, and the key results are discussed in the context of meiotic drive research.

## 5.2 Meiotic drivers accumulate genetic load

Starting in the 1940s, theoretical and empirical studies showed that the self-promoting parts of many genomes can be maintained in populations despite significant fitness costs imposed on their carriers [4, 22, 43, 45]. While many studies focused on the phenotypic costs associated with meiotic drivers, degeneration has not been assessed at the molecular level. The low recombination rates and effective population sizes of many meiotic drivers increase the effects of genetic drift [33], which is expected to lead to degeneration, for example, through higher rates of nonsynonymous amino acid substitutions or transposable element insertions.

The *t*-haplotype in house mice is a great model meiotic driver, as it shares many of the features of other meiotic drivers. It is a large [27] and ancient [31] non-recombining haplotype with several loci involved in embryonic lethality [46], male sterility [40] and drive [28]. We made use of a published dataset of genomic and transcriptomic reads from wild-caught house mice to assess the patterns of sequence variation on the *t*-haplotype and its +-homolog. We have found significantly higher nonsynonymous to synonymous SNP ratios among frequent SNPs on the *t*-haplotype, when compared to the homologous regions in house mice and the sister species *Mus spretus* (Figure 4A in Chapter 2). We have also found significantly elevated dN/dS for several *t*-specific genes (Figure 3.3 in Chapter 3), which is consistent with a decreased efficiency of purifying selection on the *t*-haplotype. Recently, the autosomal meiotic driver of

*Drosophila melanogaster*, *SD*, was also reported to harbor an excess of nonsynonymous SNPs and transposons compared to its non-driving homolog [34]. However, comprehensive studies of degeneration levels on meiotic drivers remain to be rare.

## 5.3 Occasional recombination may alleviate genetic load on meiotic drivers

We have found that the divergence between the *t*-haplotype and its non-driving homolog varied strongly across the *t* complex (Figure 1 in Chapter 2), and phylogenetic analysis suggested that the *t*-haplotype exchanged genetic material with its non-driving homolog for large parts of its sequence (Figure 2 in Chapter 2). The observation that the *t*-haplotype has a nonsynonymous to synonymous fixed SNP ratio similar to that of its non-driving homologs in regions that showed evidence of recombination (Figure 4B in Chapter 2) suggested that occasional recombination may have alleviated the *t*-haplotype's genetic load. While studies that focused on a handful of genes in the *t* complex found alleles that showed a mosaic of typically *t*-haplotype-associated and typically +-associated sequences [14], the large scale of sequence exchange that we detected along the *t*-haplotype was unexpected given the strongly suppressed recombination observed in the *t* complex. A similar result was published for the selfish X chromosome of *Drosophila neotestacea* in 2016, which showed gene flow between driving and non-driving X chromosomes for all of the 11 loci assayed, despite the presence of inversions [38]. A recent study of degeneration patterns on *SD* in *Drosophila melanogaster* also found less nonsynonymous SNPs in regions that showed genetic exchange with the non-driving homolog [34]. More population genomic studies of meiotic drivers are necessary to determine how widespread their recombination is, and whether occasional recombination is a "recovery" mechanism that is commonly associated with selfish haplotypes.

## 5.4 Revisiting the history of the *t*-haplotype

Using those regions of the *t*-haplotype that did not show evidence of recombination we could reconstruct the evolutionary history of 15 *t*-haplotypes and 40 +-*t*-complexes from geopgraphically widespread populations of three house mouse subspecies (Figure 3C in Chapter 2). This showed that the *t*-haplotype predates the split of the three house mouse subspecies, confirming previous findings based on a single *t* complex gene [31]. In contrast to this study, our data contained considerable variation between *t*-haplotypes to cluster them by subspecies (Figure 3C in Chapter 1). However, the most recent common ancestor of the 15 *t*-haplotypes seems to have lived later than that of the 40 +-*t*-complexes, suggesting that the *t*-haplotypes crossed between house mouse subspecies, which is interesting in light of the partial hybrid sterility between *M. m. musculus* and *M. m. domesticus* [16]. The *t*-haplotype contains a highly diverged allele of one of the major hybrid incompatibility loci, *Prdm9* [24]. PRDM9 determines recombination hotspots as its zinc finger array binds a specific DNA sequence, and induces a double strand break [37]. However, any mutation that escapes PRDM9-binding will be preferentially transmitted to the next generation, as it will serve as a template for repairing the asymmetric double strand break [2]. This causes the fast erosion of PRDM9 binding sites and fast evolution of PRDM9 zinc fingers [3]. Hybrid sterility arises from carrying genomes that eroded different PRDM9 binding sites, causing too many asymmetrical PRDM9-binding, and, ultimately, leading to failed meiosis [9]. *Prdm9*$^{t}$'s zinc finger is highly diverged from those of all other *Prdm9* alleles found in any of the *Mus musculus* subspecies, but identical among

*t*-haplotypes from different subspecies [24]. The co-evolution of *Prdm9*$^t$ with the genomes of all three subspecies of house mice may rescue the fertility of hybrids carrying the *t*-haplotype, and facilitate the *t*-haplotype's crossing of hybrid zones. However, this hypothesis has not been tested yet. Further, whether the diverged zinc finger of *Prdm9*$^t$ alters the recombination lanscape in $+/t$ mice will be tested using our single nucleus sequencing data of $+/t$ testes.

## 5.5 Novel expression patterns and copy number changes

We have found that 50 of our 58 assembled *t*-specific alleles are significantly differentially expressed when compared to the homologous alleles on the standard chromosome 17 (Figure 3.2 in Chapter 3). 25 of the *t*-specific alleles were consistently overexpressed, and 25 were consistently underexpressed in the tissues assayed, showing the ubiquitous effects of *t*-specific expression changes. Some of the expression changes are likely to be signs of degeneration, as expected for a large haplotype with reduced recombination. However, high expression may be a sign of functionality of a gene. Gene copy number increase seems to be associated with increased expression on the *t*-haplotype, as more than half of the overexpressed *t*-specific genes have gained copies, in contrast to only 5 of the 25 underexpressed genes (Figure 3.2 in Chapter 3). Similarly to our findings, widespread expression changes were observed also for the ancient driving X chromosome of stalk eyed flies, where of the 596 differentially expressed genes about 40% were overexpressed on the driving chromosome in the testis [44]. Further, copy number increase of genes was significantly correlated with overexpression on the driving X chromosome [44].

Copy amplification is often associated with the causative genes and suppressors of drive. The yeast driver, *wtf*, is found in many yeast species spanning 100 million years of evolution in as many as 83 copies per genome [12]. The female meiotic driver, Ab10, in maize, contains ten copies of the distorter, *Kindr*, which is highly expressed during meiosis [10]. There are amplified *Kindr*-related genes on the non-driving chromosome 10 that produce high levels of small RNAs, which are thought to suppress *Kindr* [11]. A protamine-encoding gene of the so-called *Winters* meiotic driver has amplified to a total of 22 copies in three different *Drosophila* species as a result of an arms race between the chromosomes [32]. The multicopy gene complex, *Stellate*, on the *Drosophila melanogaster* X chromosome is thought to be a cryptic driver that is suppressed by small hairpin-RNAs produced from the paralogous amplified gene complex on the Y chromosome [1].

While suppressors of the *t*-haplotype's drive have not been found yet, there are some interesting candidates on the homologous chromosome 17. The dynein, *Dynlt1* is present in multiple copies on both the *t*-haplotype and the standard chromosome 17, with very high expression during meiosis [41]. In contrast, another dynein gene *Dynlt2a* is present in three copies on the standard chromosome 17, and the *t*-allele is strongly underexpressed and contains potentially deleterious mutations [41]. Our data also shows that all three *Dynlt2a* paralogs, *Dynlt2a1*, *Dynlt2a2* and *Dynlt2a3*, are significantly underexpressed in $+/t$ mice compared to $+/+$ mice during meiosis (Figure 4.3 in Chapter 4), probably due to the missing expression from the *t* allele. This raises the interesting possibility that the $+$-chromosome's *Dynlt2a* copies act as suppressors of the *t*-haplotype's drive. Assessing if *Dynlt1* or *Dynlt2a* produce small interfering RNAs against the *t*-alleles would be a step towards investigating this hypothesis.

## 5.6 Fast protein evolution and the gain of new genes

Next to copy number increase and elevated expression, fast protein evolution is a common feature of meiotic drivers. An accelerated rate of nonsynonymous amino acid substitutions relative to the rate of synonymous substitutions (dN/dS) may indicate selection for functional innovation of a protein. Assembling *t*-specific coding sequences allowed us to detect two genes (*Ppp1r11* and *Tcte3*) with evidence of positive selection (Figure 3.3 in Chapter 3). On the driving X chromosome of stalk-eyed flies, dN/dS increase was correlated with being testis-specific and novel, and 25 genes showed dN/dS above 1, indicative of positive selection [44].

The redundancy of gene duplicates on meiotic drivers may allow their fast evolution, and proteins involved in drive often go through functional changes compared to their paralogs in the rest of the genome [7]. The *Drosophila simulans* X chromosome *Paris* drive is caused in part by a fast evolving duplicate of a Y-heterochromatin-binding gene, *HP1D2*, which lacks the protein-protein-interaction-domain , and leads to the misregulation of Y-chromatin during meiosis [20]. This bears striking resemblance to the recently identified *JASPer* duplicates on the driving X chromosome of stalk-eyed flies, which contain the heterochromatin-binding domain, but not the regulatory domain, and might contribute to the failed development of Y-bearing spermatids [44]. The driver of the autosomal *SD* system in *Drosophila melanogaster* is a truncated duplicate of *RanGAP*, missing certain domains and regulatory sites, and is mislocalized to the nucleus, leading to chromatin-compaction problems in sperm not carrying *SD* [30]. The *Ab10*-specific motor protein encoded by *Kindr* is required for the preferential segregation of the *Ab10* haplotype to the egg, and it acquired a novel cargo-binding function compared to its closest kinesin homologs in the maize genome [11]. The *t*-haplotype's responder, $Smok^{Tcr}$ is a duplicate of the sperm motility kinase *Smok*, with many amino acid changes and decreased kinase activity [21]. We have found a *t*-haplotype-specific duplicate of the highly conserved phosphatase-encoding *Ppp1cb*, which acquired dozens of substitutions and a dN/dS close to one, despite the absence of any nonsynonymous substitutions between the rat and the (non-*t*-specific) mouse alleles (Figure 3.4 in Chapter 3). $Ppp1cb^{t}$ also acquired a novel expression pattern that is exclusive to the testis (Figure 3.4 in Chapter 3), similarly to the PPP1-isoform, *Ppp1cc2*, that is relevant for sperm motility initiation [18]. The fact that $Ppp1cb^{t}$'s inhibitor, $Ppp1r11^{t}$, shows signs of positive selection opens the possibility of co-evolution of these genes on the *t*-haplotype – a hypothesis that could be assessed with *in silico* prediction of protein structure and binding, and ultimately with functional assays and transgenic mice.

## 5.7 Future directions

Recent advancements in DNA sequencing technologies have resulted in gapless genome assemblies that for the first time uncover previously "hidden" regions of genomes, such as transposon-rich regions, large duplications or repeat-arrays covering millions of basepairs – features that are commonly associated with regions of reduced recombination, such as selfish haplotypes. The handful of available assemblies of meiotic drivers uncovered the enrichment of transposable elements, duplicated genes and repeat arrays, and allowed the systematic assessment of sequence and expression evolution on these chromosomes [11, 44]. Assemblies of the *t*-haplotype have been started by some research groups, and have the potential to uncover the transposable element content, gene copy gains and possible novel gene duplicates, such as $Ppp1cb^{t}$ on this meiotic driver.

An open question about the *t*-haplotype concerns the molecular causes of homozygous male sterility [28]. Most *t*-haplotypes carry a recessive embryonic lethal [46], which eliminate homozygotes, but if a male mouse inherits two *t*-haplotypes without lethals or with complementing recessive embryonic lethals, it is always sterile. Embryonic lethality of homozygotes was hypothesized to have evolved to avoid uterine resource investment into sterile sons [6], or the extinction of the typically small demes of house mice due to male sterility [28]. However, it is unknown how homozygous male sterility evolved – whether it is due to the homozygosity of the genes involved in transmission ratio distortion, or to other genes that may have acquired deleterious mutations [28]. Sperm from $t/t$ males have flagellar and motility impairments that inhibit them from reaching the site of fertilization [35], and they are unable to fertilize eggs *in vitro* [29] – phenotypes that seem to be more severe manifestations of the motility- and fertilization-related impairments seen in sperm from $+/t$ mice. To investigate the degree of expression aberrations and biological processes associated with homozygous male sterility, and to compare them to those found in heterozygous *t*-carriers, we have conducted single nucleus RNA-sequencing of testes from mice homozygous for a *t*-haplotype.

Finally, transgenic mice have the potential to validate the role of certain candidate genes in meiotic drive. Work in this thesis presented the discovery of *Ppp1cb*$^t$, which acquired testis-specific expression and altered amino acid sequence. Single nucleus sequencing uncovered that *Ppp1cb*$^t$ likely gets incorporated into $+$ spermatids, but how it affects sperm motility and function remain to be investigated. Experiments are planned that use transgenic mice carrying $+/+$ *t* complexes and a copy of *Ppp1cb*$^t$ to assess the effect of this acquired *t*-specific gene on sperm motility.

## 5.8 Bibliography


[1] A. A. Aravin, M. S. Klenov, V. V. Vagin, F. Bantignies, G. Cavalli, and V. A. Gvozdev. Dissection of a natural RNA silencing process in the Drosophila melanogaster germ line. *Molecular and cellular biology*, 24(15):6742–6750, 2004.

[2] C. L. Baker, S. Kajita, M. Walker, R. L. Saxl, N. Raghupathy, K. Choi, P. M. Petkov, and K. Paigen. PRDM9 drives evolutionary erosion of hotspots in Mus musculus through haplotype-specific initiation of meiotic recombination. *PLoS genetics*, 11(1):e1004916, 2015.

[3] Z. Baker, M. Schumer, Y. Haba, L. Bashkirova, C. Holland, G. G. Rosenthal, and M. Przeworski. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. *Elife*, 6:e24133, 2017.

[4] D. Bruck. Male segregation ratio advantage as a factor in maintaining lethal alleles in wild populations of house mice. *Proceedings of the National Academy of Sciences*, 43(1):152–158, 1957.

[5] A. Burt and R. Trivers. Genes in conflict: the biology of selfish genetic elements. In *Genes in Conflict*. Harvard University Press, 2008.

[6] B. Charlesworth. The evolution of lethals in the t-haplotype system of the mouse. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 258(1352):101–107, 1994.

[7] C. Courret, C.-H. Chang, K. H.-C. Wei, C. Montchamp-Moreau, and A. M. Larracuente. Meiotic drive mechanisms: lessons from drosophila. *Proceedings of the Royal Society B*, 286(1913):20191430, 2019.

[8] C. Courret, P. R. Gérard, D. Ogereau, M. Falque, L. Moreau, and C. Montchamp-Moreau. X-chromosome meiotic drive in Drosophila simulans: a QTL approach reveals the complex polygenic determinism of Paris drive suppression. *Heredity*, 122(6):906–915, 2019.

[9] B. Davies, E. Hatton, N. Altemose, J. G. Hussin, F. Pratto, G. Zhang, A. G. Hinch, D. Moralli, D. Biggs, R. Diaz, et al. Re-engineering the zinc fingers of PRDM9 reverses hybrid sterility in mice. *Nature*, 530(7589):171–176, 2016.

[10] R. K. Dawe. The maize abnormal chromosome 10 meiotic drive haplotype: a review. *Chromosome Research*, 30(2-3):205–216, 2022.

[11] R. K. Dawe, E. G. Lowry, J. I. Gent, M. C. Stitzer, K. W. Swentowsky, D. M. Higgins, J. Ross-Ibarra, J. G. Wallace, L. B. Kanizay, M. Alabady, et al. A kinesin-14 motor activates neocentromeres to promote meiotic drive in maize. *Cell*, 173(4):839–850, 2018.

[12] M. De Carvalho, G.-S. Jia, A. N. Srinivasa, R. B. Billmyre, Y.-H. Xu, J. J. Lange, I. M. Sabbarini, L.-L. Du, and S. E. Zanders. The wtf meiotic driver gene family has unexpectedly persisted for over 100 million years. *Elife*, 11:e81149, 2022.

[13] N. Dobrovolskaia-Zavadskaia. Sur la mortification spontanee de la queuw chez la spuris nouveau et sur l'existence d'un caractere (facteur) hereditaire non viable. *Compr. Soc. Biol.*, 97:114–119, 1927.

[14] M. A. Erhart, S. Lekgothoane, J. Grenier, and J. H. Nadeau. Pattern of segmental recombination in the distal inversion of mouse t haplotypes. *Mammalian genome*, 13(8), 2002.

[15] L. Fishman and A. Saunders. Centromere-associated female meiotic drive entails male fitness costs in monkeyflowers. *Science*, 322(5907):1559–1562, 2008.

[16] J. Forejt, P. Jansa, and E. Parvanov. Hybrid sterility genes in mice (Mus musculus): a peculiar case of PRDM9 incompatibility. *Trends in Genetics*, 37(12):1095–1108, 2021.

[17] S. Gershenson. A new sex-ratio abnormality in Drosophila obscura. *Genetics*, 13(6):488, 1928.

[18] S. Goswami, L. Korrodi-Gregório, N. Sinha, S. Bhutada, R. Bhattacharjee, D. Kline, and S. Vijayaraghavan. Regulators of the protein phosphatase PP1$\gamma$2, PPP1R2, PPP1R7, and PPP1R11 are involved in epididymal sperm maturation. *Journal of cellular physiology*, 234(3):3105–3118, 2019.

[19] Q. Helleu, C. Courret, D. Ogereau, K. L. Burnham, N. Chaminade, M. Chakir, S. Aulard, and C. Montchamp-Moreau. Sex-Ratio meiotic drive shapes the evolution of the Y chromosome in Drosophila simulans. *Molecular Biology and Evolution*, 36(12):2668–2681, 2019.

[20] Q. Helleu, P. R. Gérard, R. Dubruille, D. Ogereau, B. Prudhomme, B. Loppin, and C. Montchamp-Moreau. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proceedings of the National Academy of Sciences*, 113(15):4110–4115, 2016.

[21] B. G. Herrmann, B. Koschorz, K. Wertz, K. J. McLaughlin, and A. Kispert. A protein kinase encoded by the t complex responder gene causes non-mendelian inheritance. *Nature*, 402(6758):141–146, 1999.

[22] Y. Hiraizumi, L. Sandler, and J. F. Crow. Meiotic drive in natural populations of Drosophila melanogaster. iii. populational implications of the segregation-distorter locus. *Evolution*, pages 433–444, 1960.

[23] L. D. Hurst. A century of bias in genetics and evolution. *Heredity*, 123(1):33–43, 2019.

[24] H. Kono, M. Tamura, N. Osada, H. Suzuki, K. Abe, K. Moriwaki, K. Ohta, and T. Shiroishi. Prdm9 polymorphism unveils mouse evolutionary tracks. *Dna Research*, 21(3):315–326, 2014.

[25] E. Lader, H.-S. Ha, M. O'Neill, K. Artzt, and D. Bennett. tctex-1: a candidate gene family for a mouse t complex sterility locus. *Cell*, 58(5):969–979, 1989.

[26] A. M. Larracuente and D. C. Presgraves. The selfish Segregation Distorter gene complex of Drosophila melanogaster. *Genetics*, 192(1):33–53, 2012.

[27] M. Lyon, J. Zenthon, E. Evans, M. Burtenshaw, and K. Willison. Extent of the mouse t complex and its inversions shown by in situ hybridization. *Immunogenetics*, 27:375–382, 1988.

[28] M. F. Lyon. Transmission ratio distortion in mice. *Annual review of genetics*, 37(1):393–408, 2003.

[29] J. McGRATH and N. Hillman. Sterility in mutant (t lx/t ly) male mice iii. in vitro fertilization. *Development*, 59(1):49–58, 1980.

[30] C. Merrill, L. Bayraktaroglu, A. Kusano, and B. Ganetzky. Truncated RanGAP encoded by the Segregation Distorter locus of Drosophila. *Science*, 283(5408):1742–1745, 1999.

[31] T. Morita, H. Kubota, K. Murata, M. Nozaki, C. Delarbre, K. Willison, Y. Satta, M. Sakaizumi, N. Takahata, and G. Gachelin. Evolution of the mouse t haplotype: recent and worldwide introgression to Mus musculus. *Proceedings of the National Academy of Sciences*, 89(15):6851–6855, 1992.

[32] C. A. Muirhead and D. C. Presgraves. Satellite DNA-mediated diversification of a sex-ratio meiotic drive gene family in Drosophila. *Nature Ecology & Evolution*, 5(12):1604–1612, 2021.

[33] H. J. Muller. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9, 1964.

[34] B. Navarro-Dominguez, C.-H. Chang, C. L. Brand, C. A. Muirhead, D. C. Presgraves, and A. M. Larracuente. Epistatic selection on a selfish Segregation Distorter supergene–drive, recombination, and genetic load. *Elife*, 11:e78981, 2022.

[35] P. Olds-Clarke and L. R. Johnson. t haplotypes in the mouse compromise sperm flagellar function. *Developmental biology*, 155(1):14–25, 1993.

[36] G. Ostergren. Parasitic nature of extra fragment chromosomes. *Bot. Not.*, 2:157–163, 1945.

[37] K. Paigen and P. M. Petkov. PRDM9 and its role in genetic recombination. *Trends in Genetics*, 34(4):291–300, 2018.

[38] K. E. Pieper and K. A. Dyer. Occasional recombination of a selfish X-chromosome may permit its persistence at high frequencies in the wild. *Journal of Evolutionary Biology*, 29(11):2229–2241, 2016.

[39] K. E. Pieper, R. L. Unckless, and K. A. Dyer. A fast-evolving X-linked duplicate of importin-$\alpha 2$ is overexpressed in sex-ratio drive in Drosophila neotestacea. *Molecular ecology*, 27(24):5165–5179, 2018.

[40] S. Pilder, J. Lu, Y. Han, L. Hui, S. Samant, O. Olugbemiga, K. Meyers, L. Cheng, and S. Vijayaraghavan. The molecular basis of "curlicue": a sperm motility abnormality linked to the sterility of t haplotype homozygous male mice. *Society of Reproduction and Fertility supplement*, 63:123–133, 2007.

[41] S. H. Pilder. Does dynein influence the non-mendelian inheritance of chromosome 17 homologs in male mice? In *Dyneins*, pages 538–559. Elsevier, 2012.

[42] C. A. Pinzone and K. A. Dyer. Association of polyandry and sex-ratio drive prevalence in natural populations of Drosophila neotestacea. *Proceedings of the Royal Society B: Biological Sciences*, 280(1769):20131397, 2013.

[43] T. Prout. Some effects of variations in the segregation ratio and of selection on the frequency of alleles under random mating. *Acta Genetica Et Statistica Medica*, 4(2-3):148–151, 1953.

[44] J. A. Reinhardt, R. H. Baker, A. V. Zimin, C. Ladias, K. A. Paczolt, J. H. Werren, C. Y. Hayashi, and G. S. Wilkinson. Impacts of sex ratio meiotic drive on genome structure and function in a stalk-eyed fly. *Genome Biology and Evolution*, 15(7):evad118, 2023.

[45] L. Sandler and E. Novitski. Meiotic drive as an evolutionary force. *The American Naturalist*, 91(857):105–110, 1957.

[46] M. Sugimoto. Developmental genetics of the mouse t-complex. *Genes & Genetic Systems*, 89(3):109–120, 2014.