Cell Genomics, Volume *3*

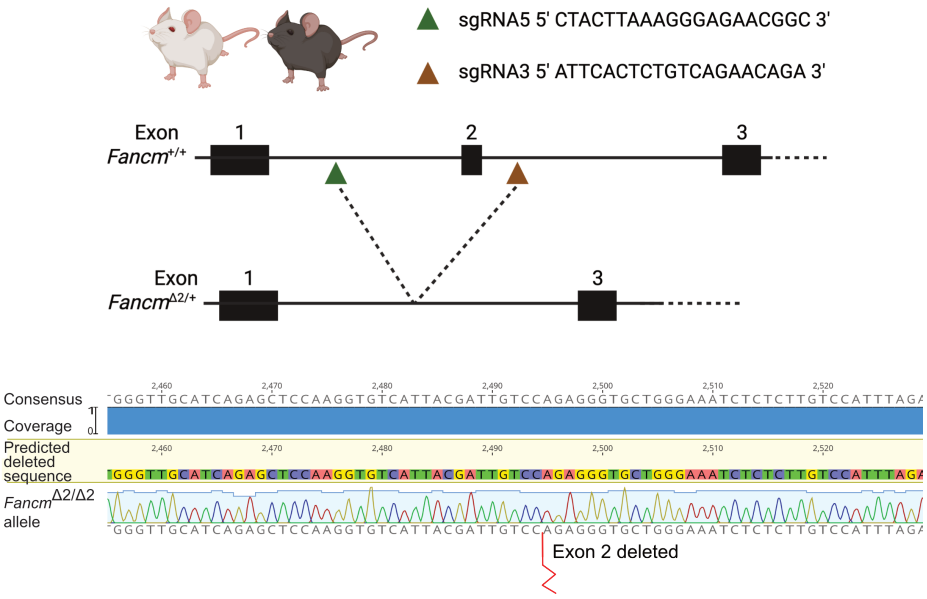
Supplemental information

Fancm has dual roles in the limiting of meiotic crossovers and germ cell maintenance in mammals

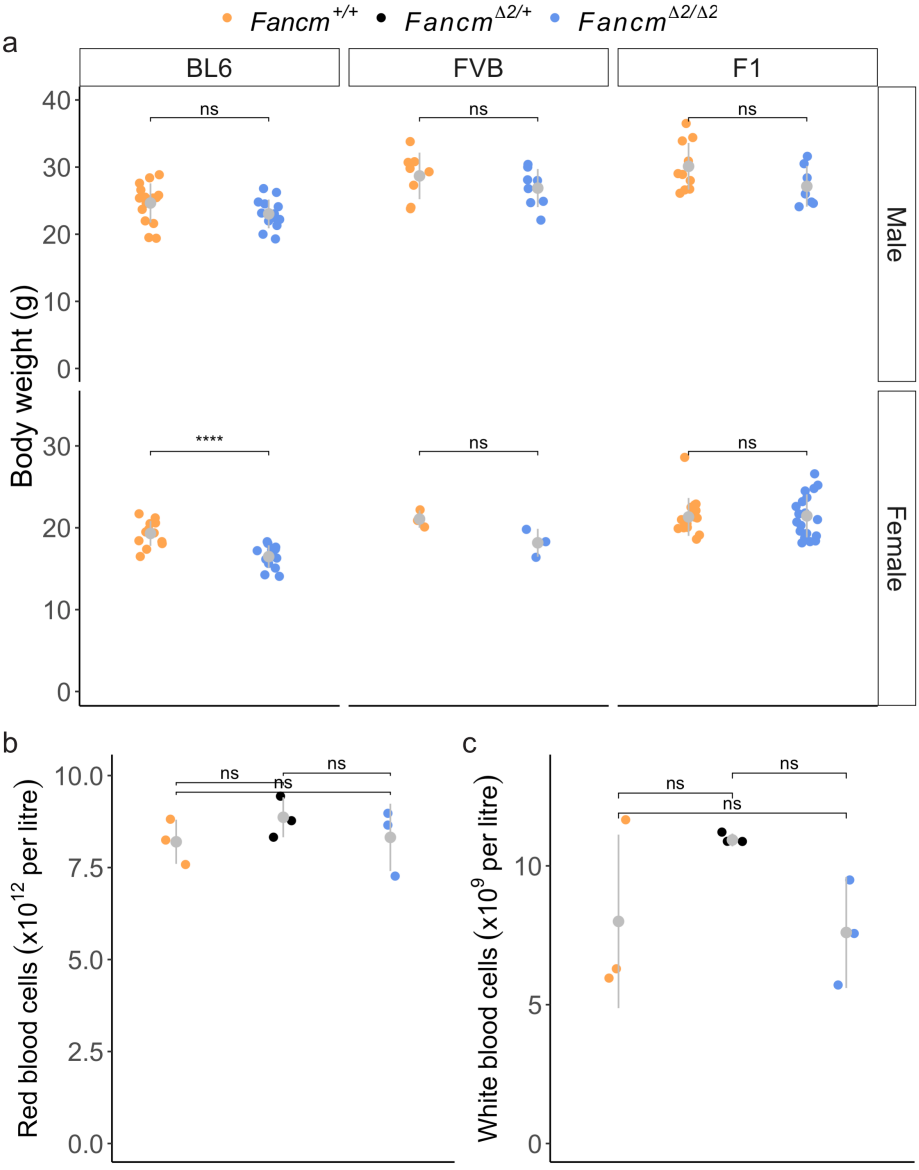
Vanessa Tsui, Ruqian Lyu, Stevan Novakovic, Jessica M. Stringer, Jessica E.M. Dunleavy, Elissah Granger, Tim Semple, Anna Leichter, Luciano G. Martelotto, D. Jo Merriner, Ruijie Liu, Lucy McNeill, Nadeen Zerafa, Eva R. Hoffmann, Moira K. O'Bryan, Karla Hutt, Andrew J. Deans, Jörg Heierhorst, Davis J. McCarthy, and Wayne Crismani

Supporting information

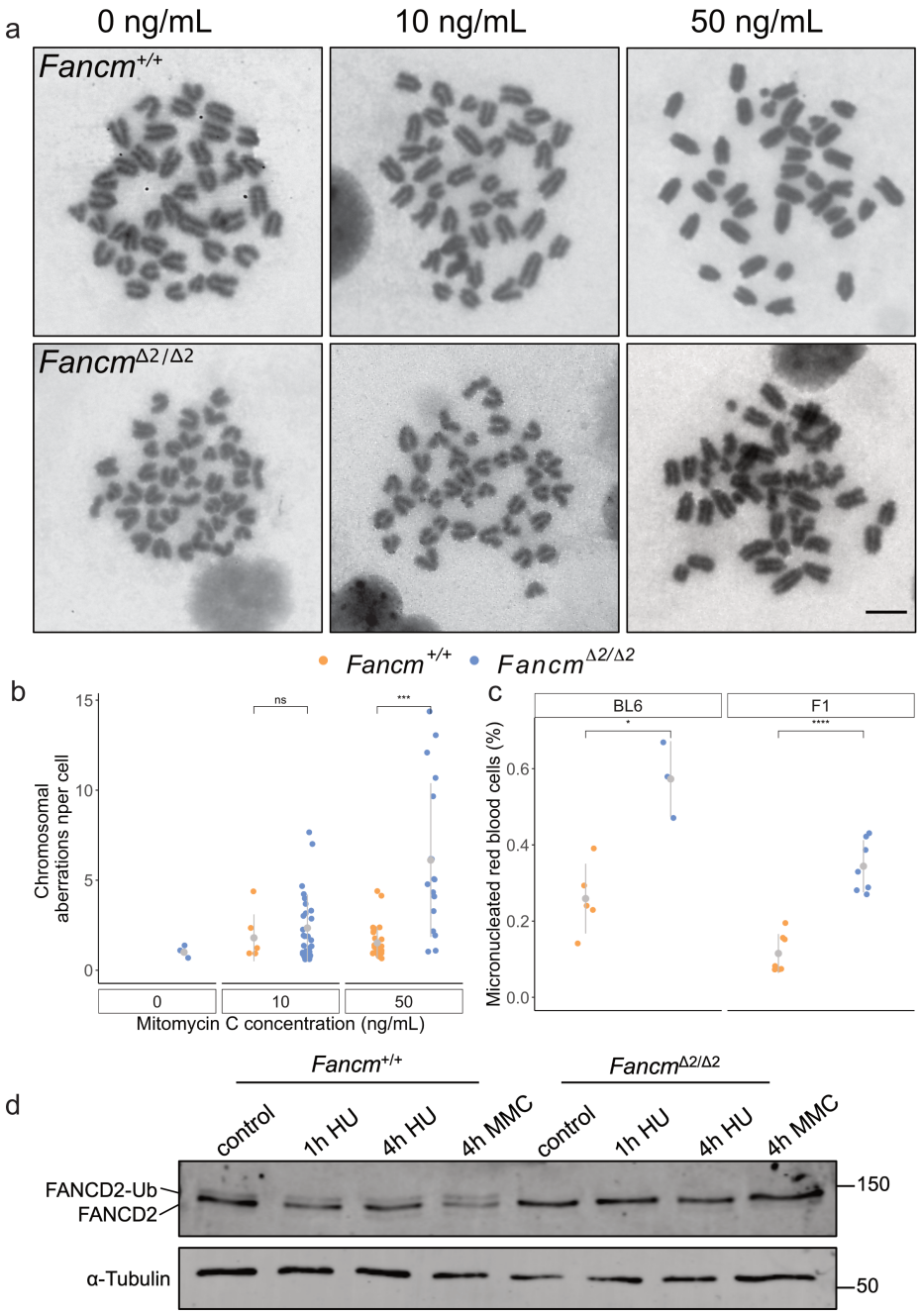
S1 Fig. **Transgenic mouse generation. Related to STAR methods.** *Fancm* mouse knockouts were made in two strains, C57BL/6J and FVB/N, with the same two guide RNAs, which resulted in the targeted deletion of exon 2.



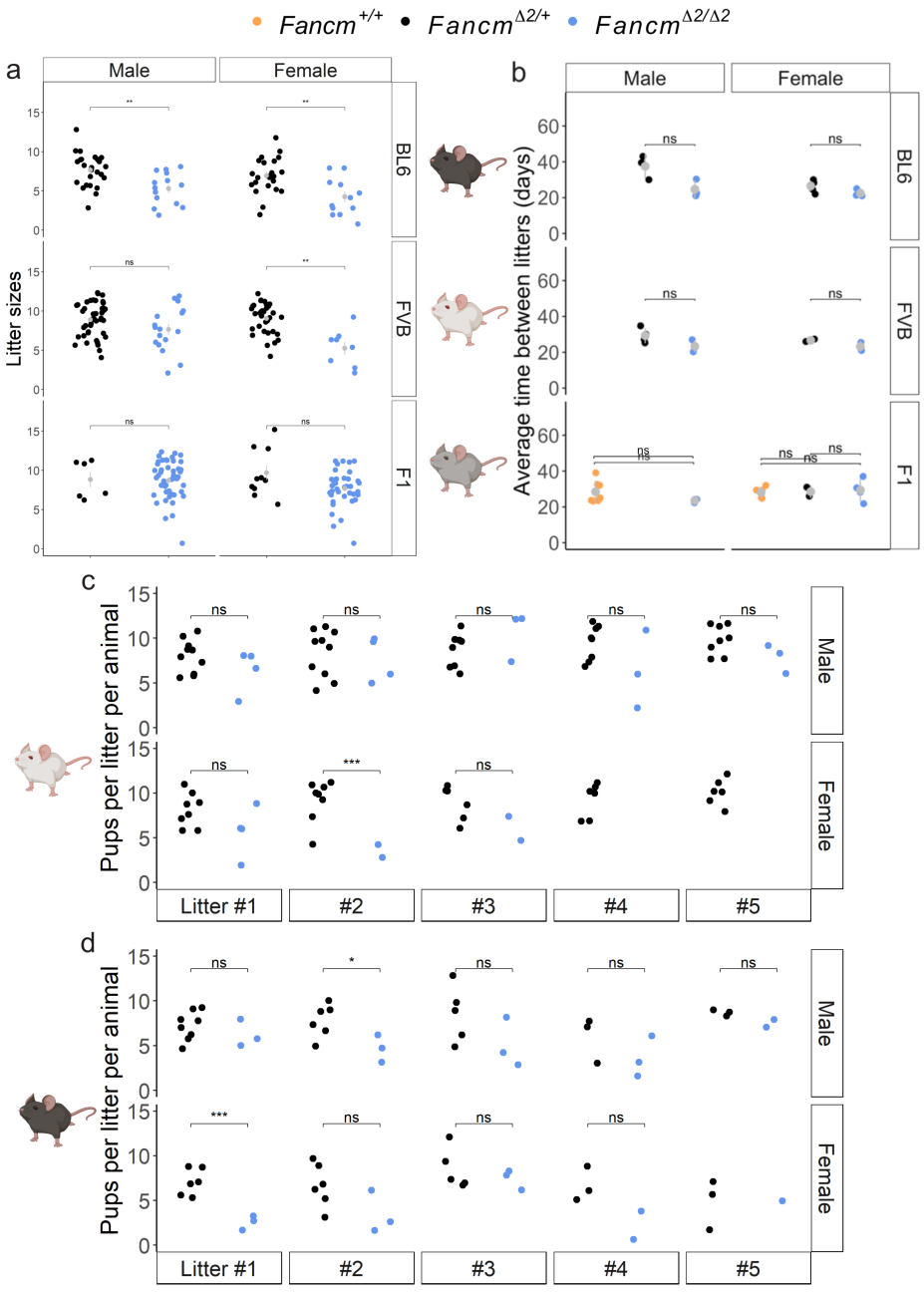
S2 Fig. **Body weight and blood parameters are normal in *Fancm*$^{\Delta 2/\Delta 2}$ mice. Related to STAR methods.** a) Body weights of 6-14 week old *Fancm*-deficient mice and control litter mates. (**** indicates p-value $\leq$ 0.0001 (unpaired t-test). b) Red blood counts from B6.*Fancm* mice and their control siblings. c) White counts from B6.*Fancm* mice and their control siblings.



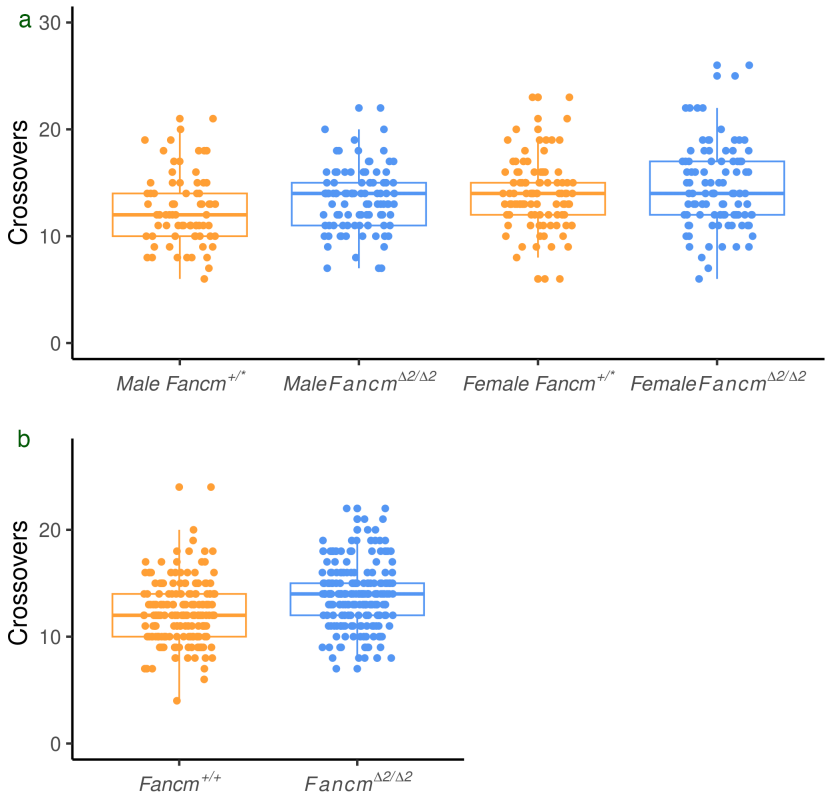
S3 Fig. **Genomic instability in *Fancm*-deficient mice. Related to STAR methods.** a) Chromosome breakage tests shows a dose-dependent response to mitomycin C (0, 10 and 50 ng/mL) in the absence of *Fancm*. Scale bar is 5 μm. b) Quantification of chromosomal aberrations per in $Fancm^{+/+}$ and $Fancm^{\Delta 2/\Delta 2}$ cells. c) Increased micronuclei in mature red blood cells reveals an increase in genomic instability in *Fancm* mutants. d) FANCD2 mono-ubiquitination in splenic B cells measured by western blot. An 8 kDa shift occurs when FANCD2 is mono-ubiquitinated. This can occur under basal conditions and treatment with hydroxyurea (10 mM) which causes replication fork stalling, or the DNA crosslinker mitomycin C (50 μg/μL). FANCD2 mono-ubiquitination is not detected in the absence of *Fancm* under any of these conditions. Error bars indicate mean $\pm$ s.d., *** indicates $p \leq 0.001$, unpaired t-test.


a
0 ng/mL
10 ng/mL
50 ng/mL
Fancm+/+
FancmΔ2/Δ2
b
Fancm+/+
FancmΔ2/Δ2
Chromosomal aberrations nper cell
ns
***
0
10
50
Mitomycin C concentration (ng/mL)
c
BL6
F1
Micronucleated red blood cells (%)
d
Fancm+/+
FancmΔ2/Δ2
control
1h HU
4h HU
4h MMC
FANCD2-Ub
FANCD2
α-Tubulin
150
50

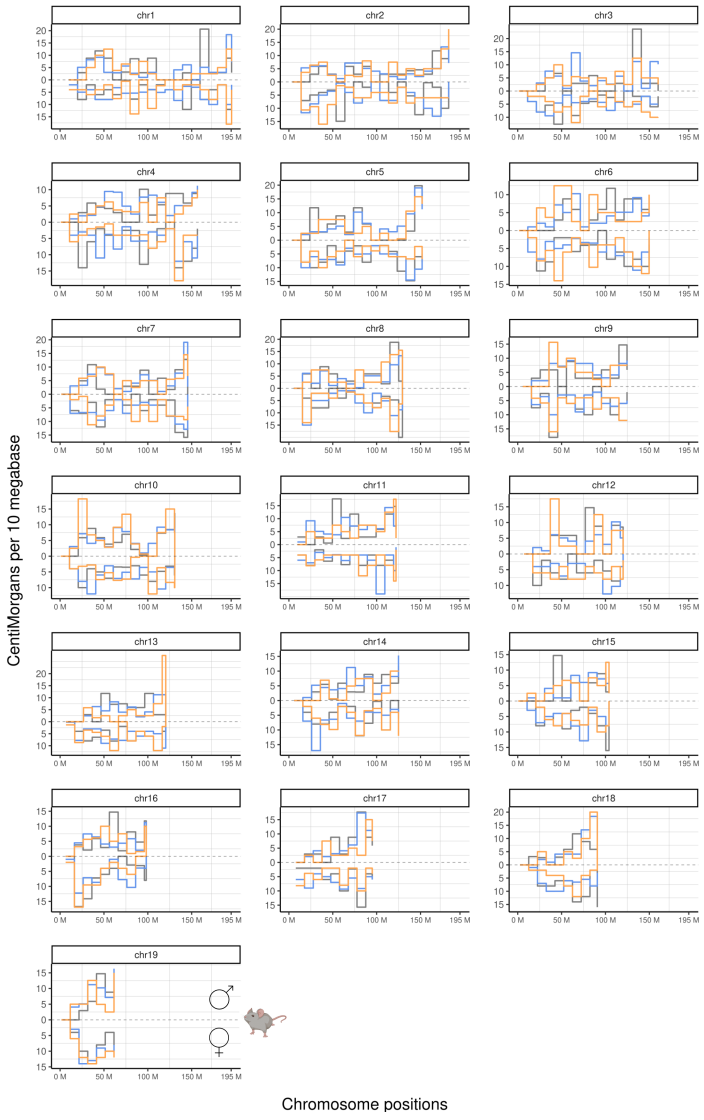
S4 Fig. **Litter sizes from *Fancm*-deficient mice and litter mate controls. Related to Figure 4.** a) Average litter sizes over five litters per mouse in C57BL/6J, FVB/N and F1 strains. b) Average time between litters in days in C57BL/6J, FVB/N and F1 mice. The average pups per litter was also considered in c) FVB/N and b) C57BL/6J mice. Mice were allowed to breed for five litters or 6 months. For the C57BL/6J and FVB/N data, their mate of the opposite sex had a wild-type *Fancm* allele. In the case of F1 mice, their mate of the opposite sex had either a wild-type or heterozygous *Fancm* allele. Data shown are mean $\pm$ s.d. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$, unpaired t-test.


Fancm+/+ Fancm∆2/+ Fancm∆2/∆2
a
Male
Female
BL6
FVB
F1
Litter sizes
b
Average time between litters (days)
c
Pups per litter per animal
Litter #1
#2
#3
#4
#5
Male
Female
d
Pups per litter per animal
Litter #1
#2
#3
#4
#5
Male
Female

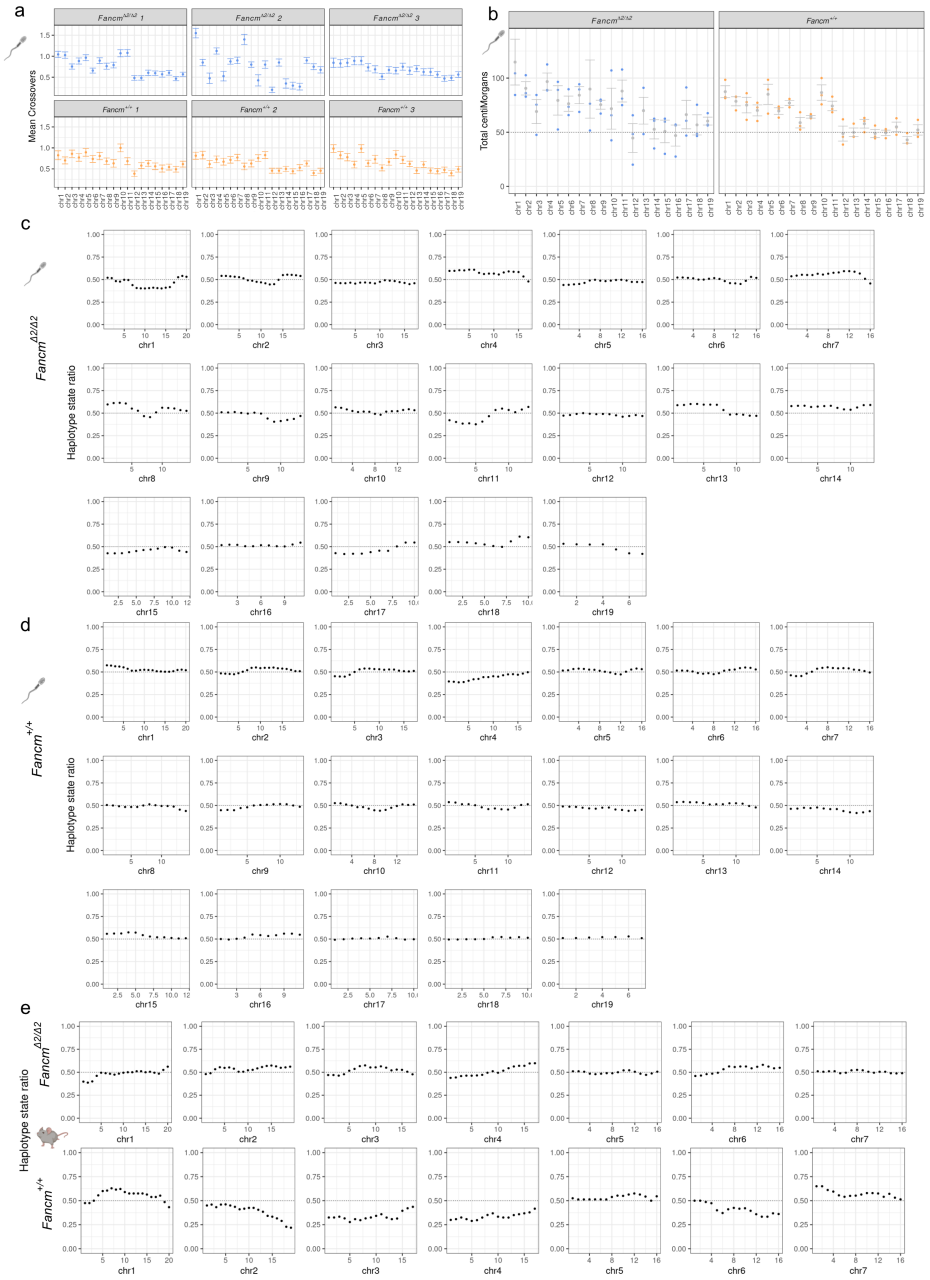
S5 Fig. **The number of crossovers detected in F1 samples and gametes from each genotype group. Related to Figure 1.** a. The number of crossovers detected in F1 samples. b. The number of crossovers detected in single gametes.



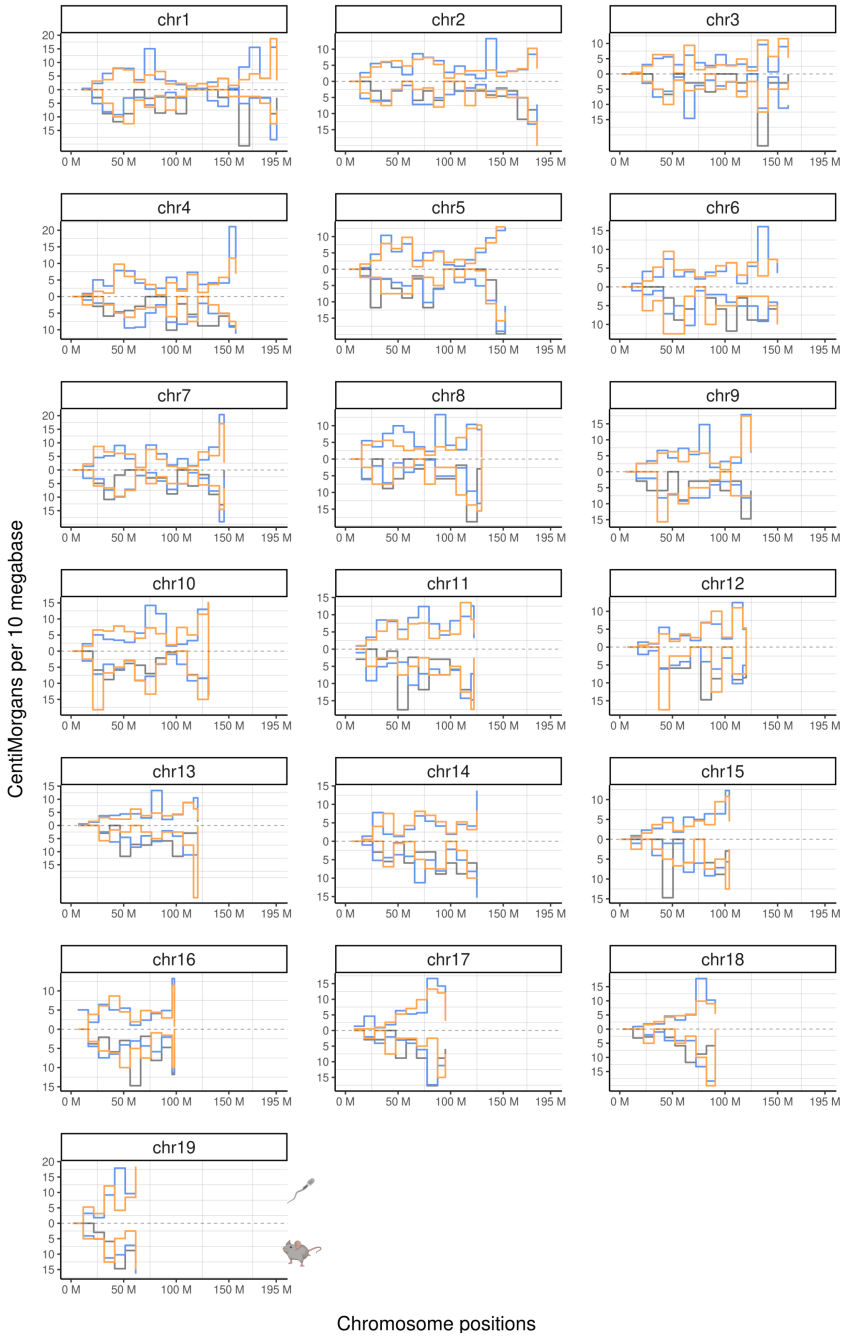
S6 Fig. **Male and female genome-wide genetic distances. Related to Figure 1.** Male and female crossover frequencies detected using bulk sequencing of pedigree samples. Genetic distances in each bin between *Fancm*$^{+/+}$ (orange), *Fancm*$^{\Delta 2/+}$ (black) and *Fancm*$^{\Delta 2/\Delta 2}$ (blue) were tested, and there were no significant differences (permutation testing, B = 1,000).



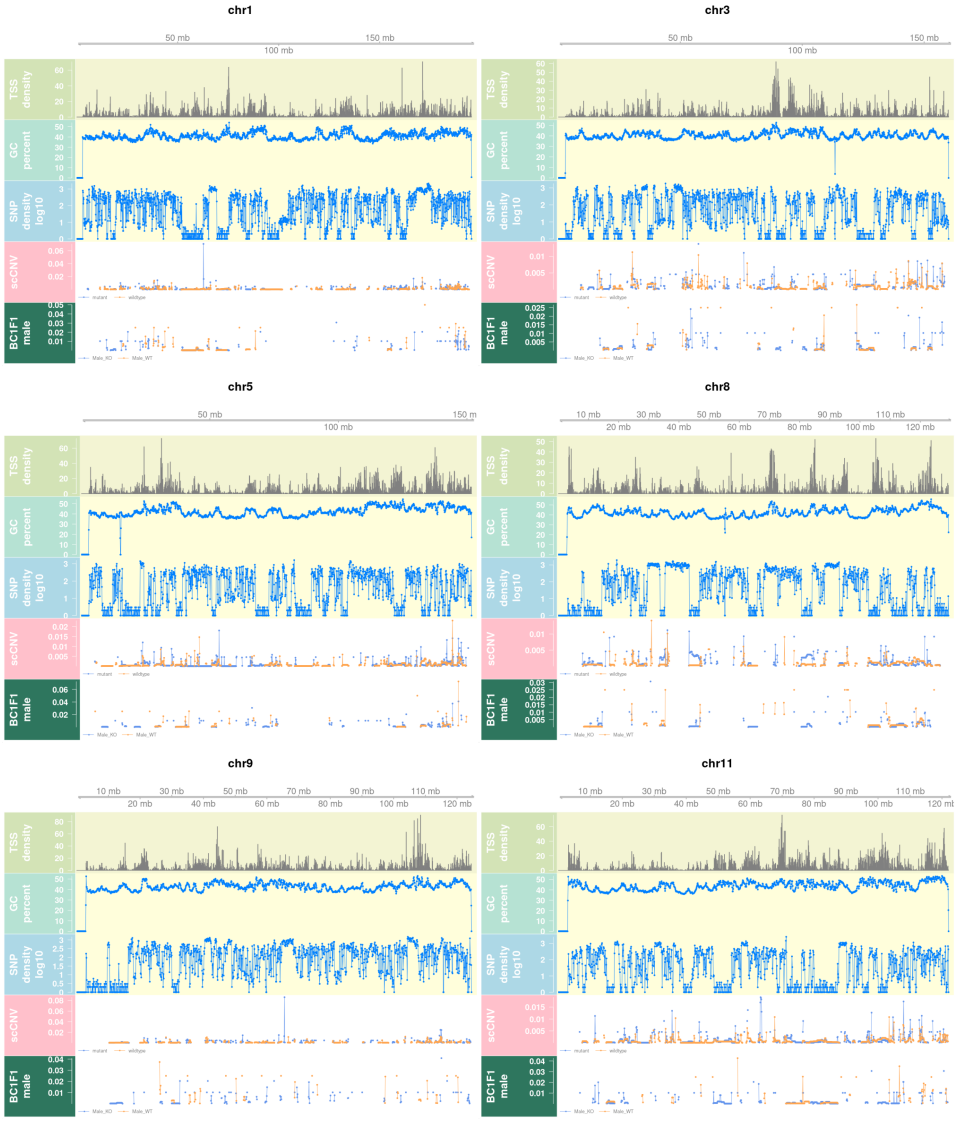
S7 Fig. **Crossover detection with single sperm sequencing. Related to Figure 1.** Droplet-based single sperm sequencing produced datasets that could detect crossovers a) Mean crossovers per chromosome of sperm sequenced from F1 individuals. b) Total centiMorgans per chromosome from observed crossover rates per individual. Number of cells per individual: $Fancm^{\Delta 2/\Delta 2}$: 114, 40, 64. $Fancm^{+/+}$ 57, 70, 63. c) Marker segregation ratio in all gametes from $Fancm^{\Delta 2/\Delta 2}$ F1 individuals. Ratios of counts of cells with the two haplotypes were calculated in chromosome bins (of size 10 megabases). d) Marker segregation ratio in all gametes from $Fancm^{+/+}$ F1 individuals. Hypothesis testing using a binomial test was performed to test if the marker segregation ratio is different from 0.5, no significant differences were observed in any chromosomes. e) Same observation in Bulk BC1F1 pups that no marker segregation distortions were found (only showing selected chromosomes).


a
b
c
d
e
Mean Crossovers
Total centiMorgans
Haplotype state ratio
Fancm Δ2/Δ2
Fancm +/+

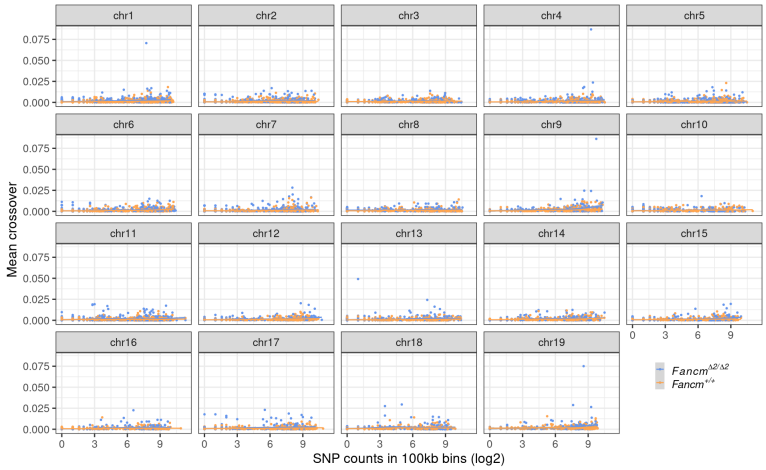
S8 Fig. **F1 single-sperm sequencing and BC1F1 bulk sequencing produce comparable crossover distributions. Related to Figure 1.** Male recombination maps of all 19 mouse autosomes are presented in separate plots of *Fancm*$^{+/+}$ (orange), *Fancm*$^{\Delta 2/+}$ (black) and *Fancm*$^{\Delta 2/\Delta 2}$ (blue). The genetic distances from sperm-sequencing data from F1 mice is flipped to the negative scale. Bin sizes are 10 megabases.


chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
0 M
50 M
100 M
150 M
195 M
CentiMorgans per 10 megabase
Chromosome positions

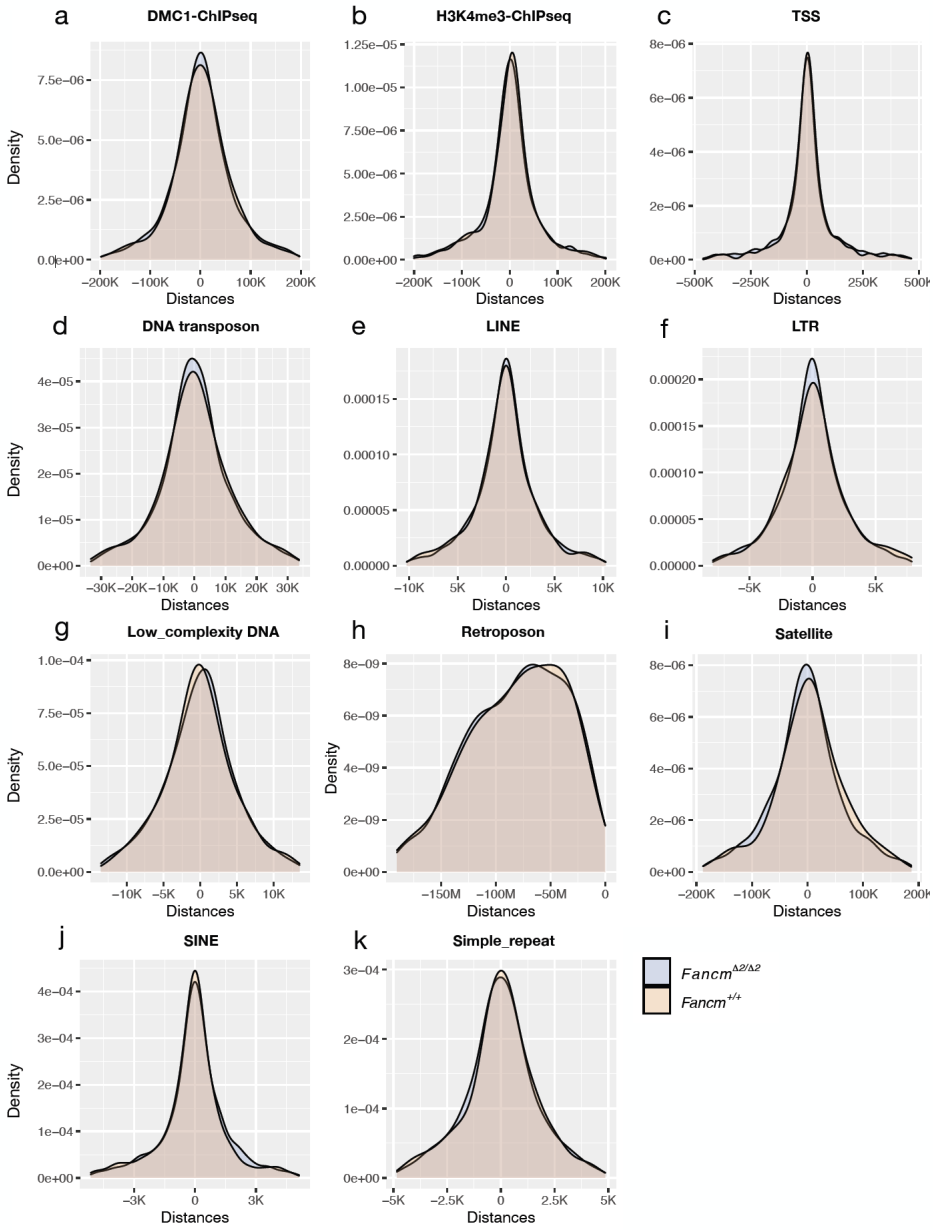
S9 Fig. **The effect of chromosomal features on crossover frequencies. Related to STAR methods.** Visual representation of the effect of SNP density and other chromosomal features on crossover frequency (Size of 100kb physical base pair bins) from selected chromosomes. The bottom two rows of each chromosome track represent crossover densities for the single-gamete sequencing ("scCNV") and pedigree-based ("BC1F1 male") methods. Orange = *Fancm*$^{+/+}$ and blue = *Fancm*$^{\Delta 2/\Delta 2}$.



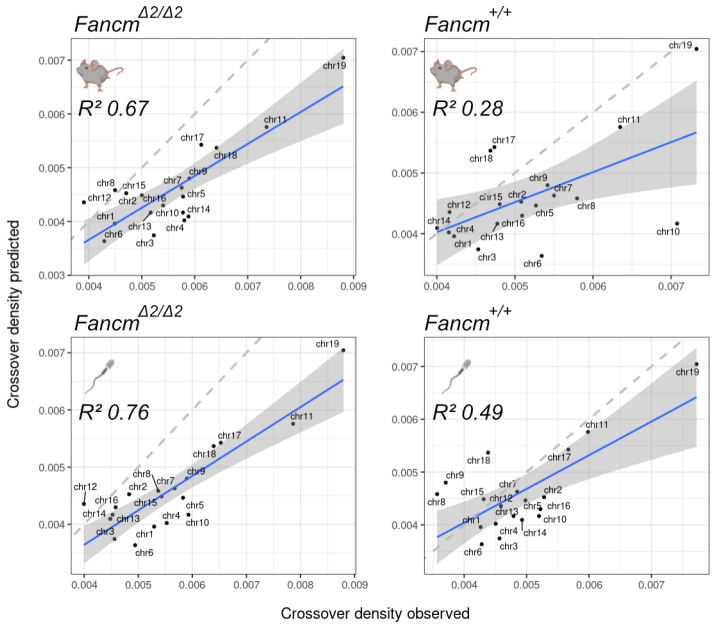
S10 Fig. **Assessment of the relationship between SNP density and crossover frequency. Related to STAR methods.** Visual representation of the effect of SNP density on crossover frequency using a physical bin size of 100kb).



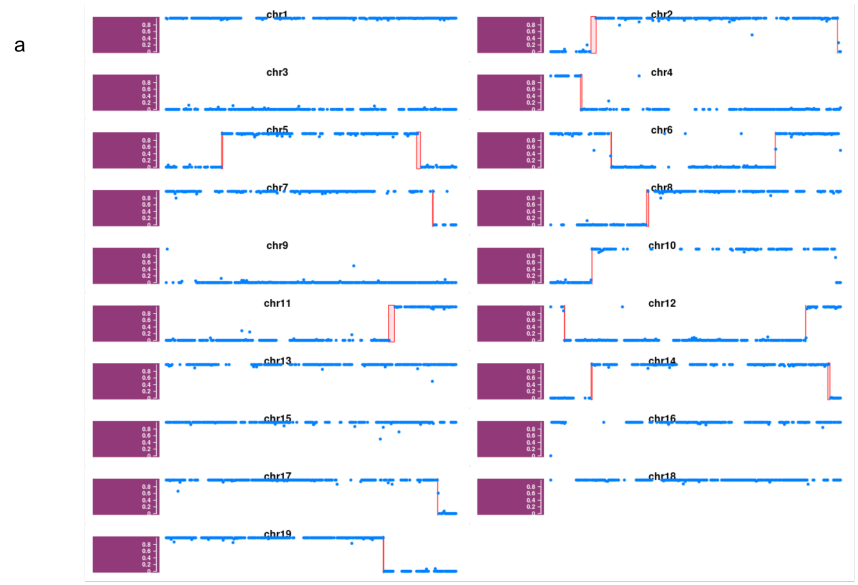
S11 Fig. **The correlation of meiotic DSBs and genomic features on crossover frequency. Related to STAR methods.** Meiotic crossover frequency was compared to DSBs and genomic features to determine if the extra crossovers in *Fancm*-deficient mice display typical patterns of distribution with respect to these variables. a) DMC1-ChIPseq data from [3], which is a proxy for DSBs, b) H3K4me3-ChIPseq data from [3], which is commonly associated with accessible chromatin and crossover occurence, c) transcription start sites, d) DNA transposons, e) long interspersed nuclear elements, f) long terminal repeat retrotransposons, g) low complexity repeats are AT-rich or GC-rich regions, h) retrotransposons, i) satellite repeats, j) short interspersed nuclear elements, k) simple repeats are micro-satellite repeats that have interspersed characters. c,d,e,f,i,j,k genomic annotations are from RepeatMasker [2]

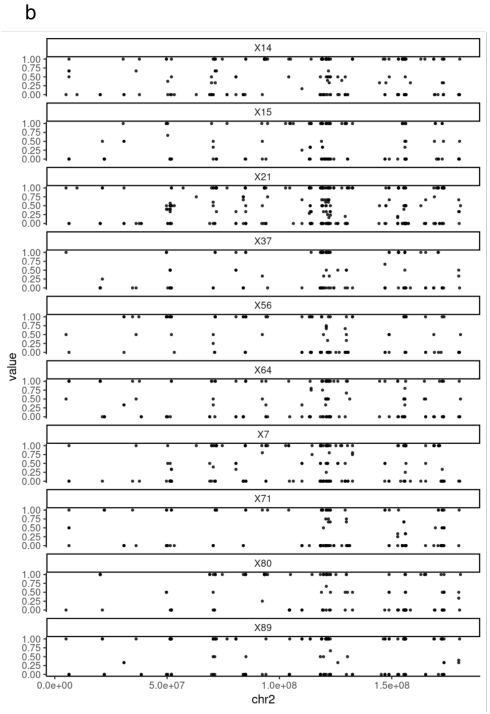

a
DMC1-ChIPseq
b
H3K4me3-ChIPseq
c
TSS
d
DNA transposon
e
LINE
f
LTR
g
Low_complexity DNA
h
Retroposon
i
Satellite
j
SINE
k
Simple_repeat
Density
Distances
Fancm^Δ2/Δ2
Fancm^+/+

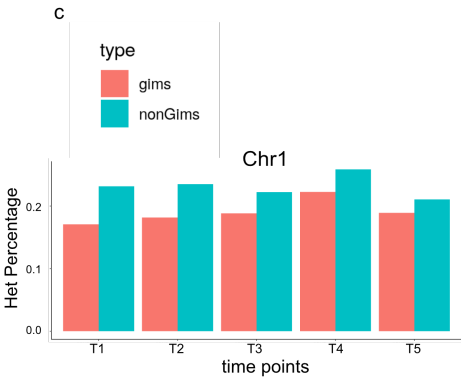
S12 Fig. **Experimental and predicted regression modelling of crossover densities. Related to STAR methods.** The model from [4] was used to predict crossover density for the 19 autosomes. Linear regression was fitted to compare the predicted crossover densities with observed crossover densities from each sample and data group. Dashed lines indicate the diagonal lines with slope 1 with adjusted R squares printed generated using lm() from R (R Core Team, 2021). Grey area is the standard error of the regression.



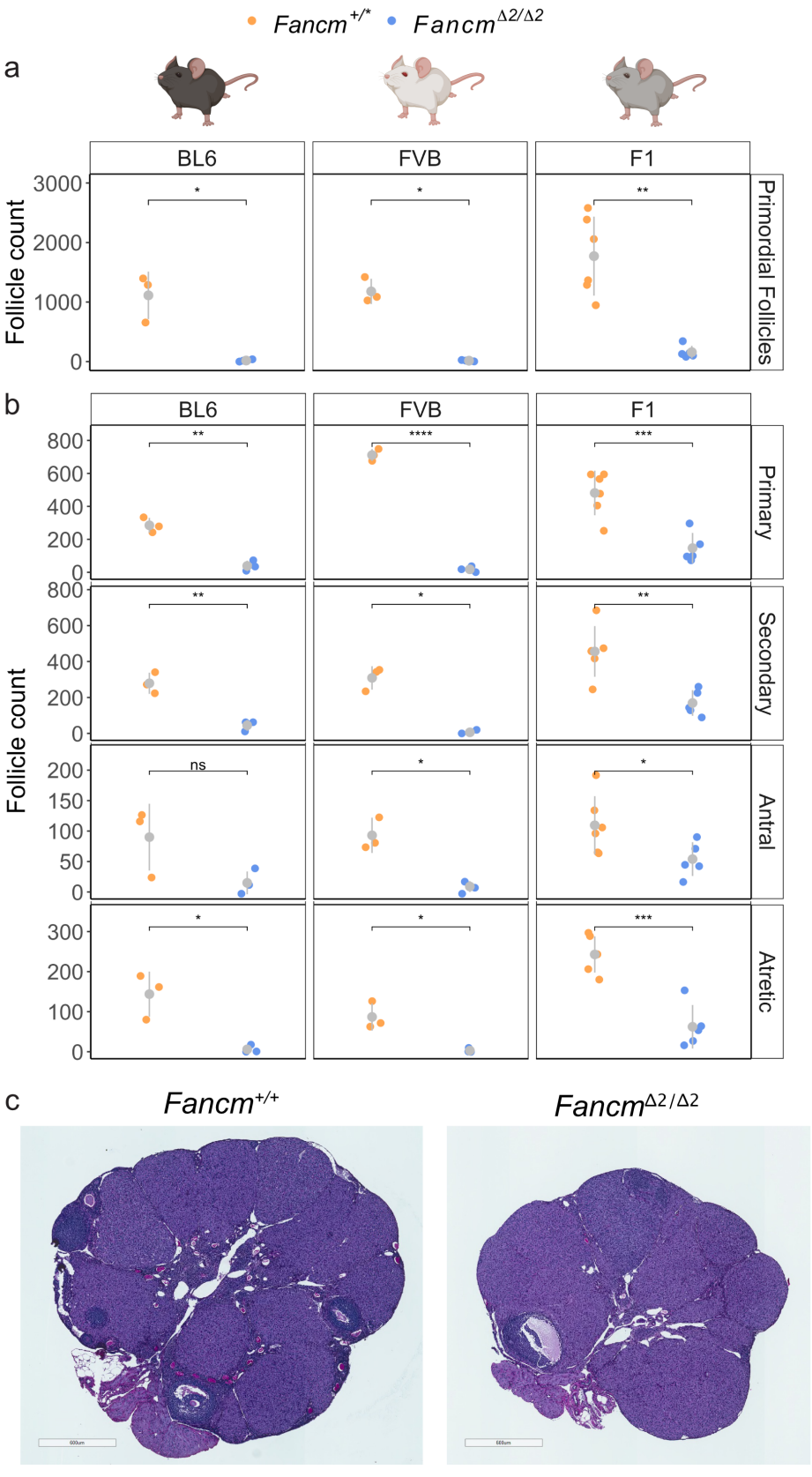
S13 Fig. **Mendelian inheritance of DNA, but not RNA, in haploid sperm. Related to Figure 1.** a) Single sperm DNA-sequencing methods produce data that is sufficient for crossover detection, b) whereas scRNA-sequencing methods do not. c) However, selected transcripts (GIMs from [5]) are more representative of the haploid nuclear allelic content, but still show significant transcript sharing, which is reflected by detection of heterozygosity. 100 cells from each time points are selected and the SNPs were grouped by whether falling in GIM or non GIMs regions. The percentage of the SNPs were called with heterogzygous (alternative allele frequencies were between 0 (C57BL/6J) to 1 (FVB/N) were plotted).


a
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19

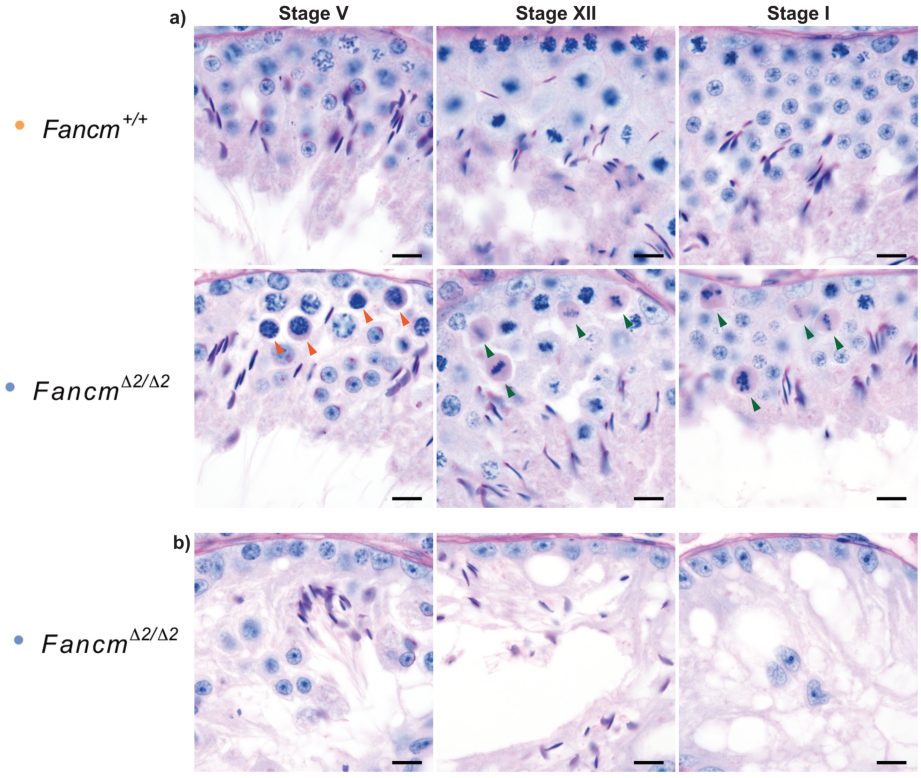

b
X14
X15
X21
X37
X56
X64
X7
X71
X80
X89
value
1.00
0.75
0.50
0.25
0.00
0.0e+00
5.0e+07
1.0e+08
1.5e+08
chr2


c
type
gims
nonGims
Chr1
Het Percentage
0.2
0.1
0.0
T1
T2
T3
T4
T5
time points

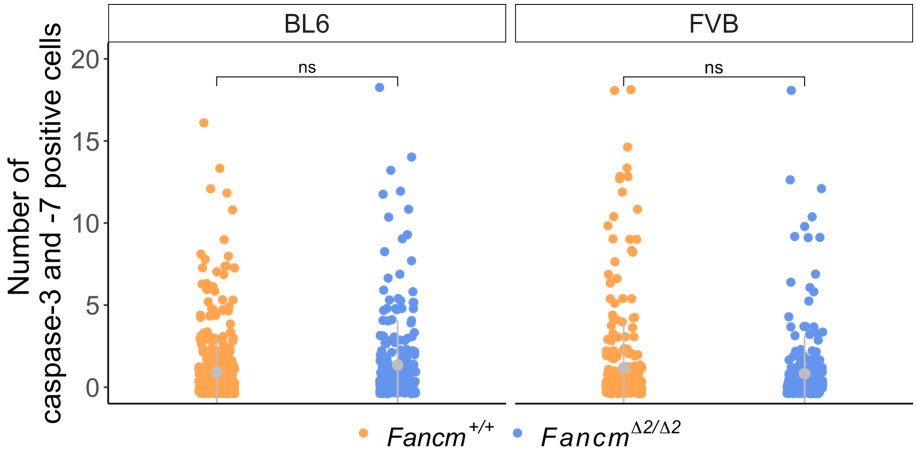
S14 Fig. **Folliculogenesis is perturbed *Fancm*-deficient mice. Related to Figure 4.** Follicle numbers in 3 month old $Fancm^{+/*}$ and $Fancm^{\Delta 2/\Delta 2}$ in C57BL/6J, FVB/N and F1 strains. a) Primordial follicle number are significantly reduced in FVB.$Fancm^{\Delta 2/\Delta 2}$, B6.$Fancm^{\Delta 2/\Delta 2}$ and F1.$Fancm^{\Delta 2/\Delta 2}$ ovaries. b) Primary, secondary and atretic follicle numbers are significantly lower in FVB.$Fancm^{\Delta 2/\Delta 2}$, B6.$Fancm^{\Delta 2/\Delta 2}$ F1.$Fancm^{\Delta 2/\Delta 2}$ ovaries. c) Representative images of middle sections of FVB.$Fancm^{+/+}$ and FVB.$Fancm^{\Delta 2/\Delta 2}$ ovaries. Scale is 600 μm. $Fancm^{+/*}$ indicates $Fancm^{+/+}$ or $Fancm^{+/\Delta 2}$. Data shown are mean $\pm$ s.d. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$.


Fancm+/+
Fancm Δ2/Δ2
a
BL6
FVB
F1
Follicle count
Primordial Follicles
b
Primary
Secondary
Antral
Atretic
ns
c
Fancm+/+
Fancm Δ2/Δ2

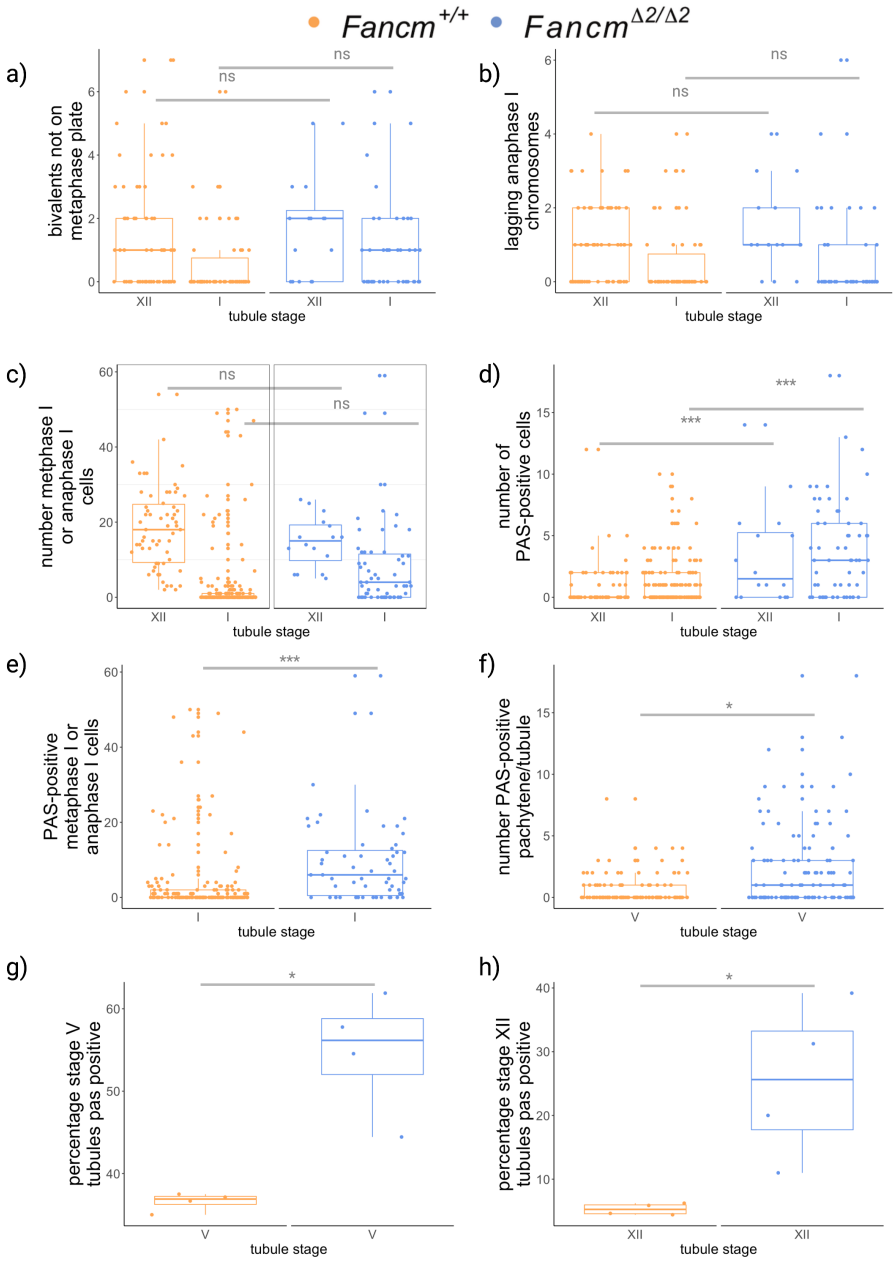
S15 Fig. **Representative images of histological defects observed in B6.*Fancm*$^{\Delta 2/\Delta 2}$ seminiferous tubules. Related to Figure 5.** a) Orange arrows indicate PAS-positive pachytene spermatocytes, green arrows indicate PAS-positive spermatocytes in stage XII and I. b) Representative images that depict tubules where spermatocytes appear to be the first population to be lost. On the left panel spermtogonia, Sertoli cells, spermatids (round and elongated) were observed but no spermatocytes. The middle and far right panels show progression to a Sertoli cell-only phenotype. Scale bar = 10 microns.



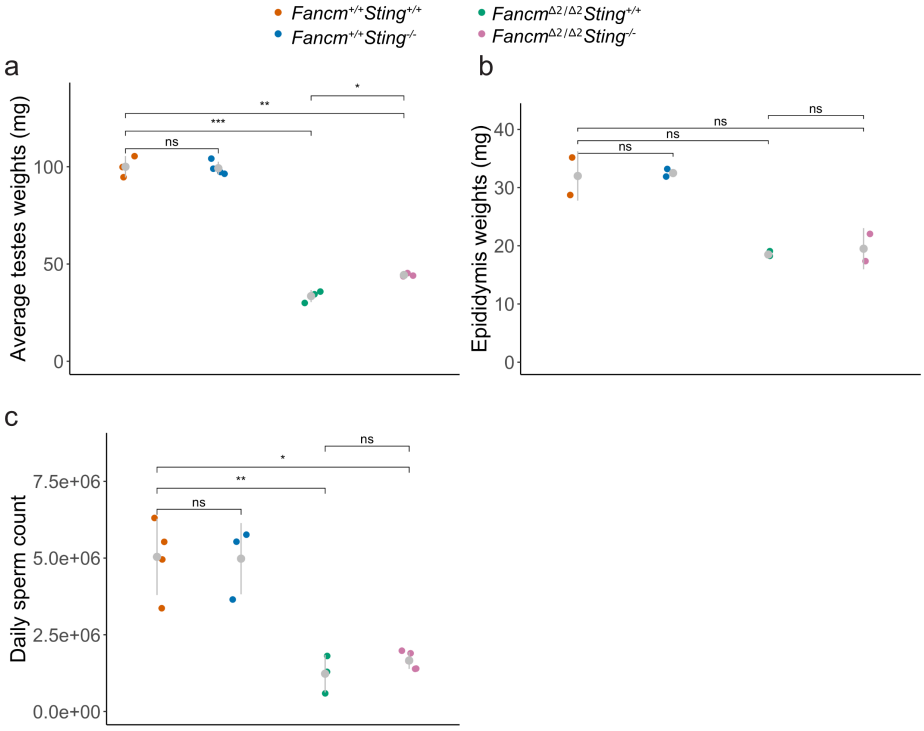
S16 Fig. **Germ cell apoptosis analysis in *Fancm*-deficient seminiferous tubules. Related to Figure 3.** Caspase-3 and -7 positive cells $Fancm^{\Delta 2/\Delta 2}$ in BL6 and FVB seminiferous tubules are not significantly different from $Fancm^{+/+}$. Generalised linear mixed (GLM) models used as to compare the number of caspase-positive cells per tubule between genotypes.



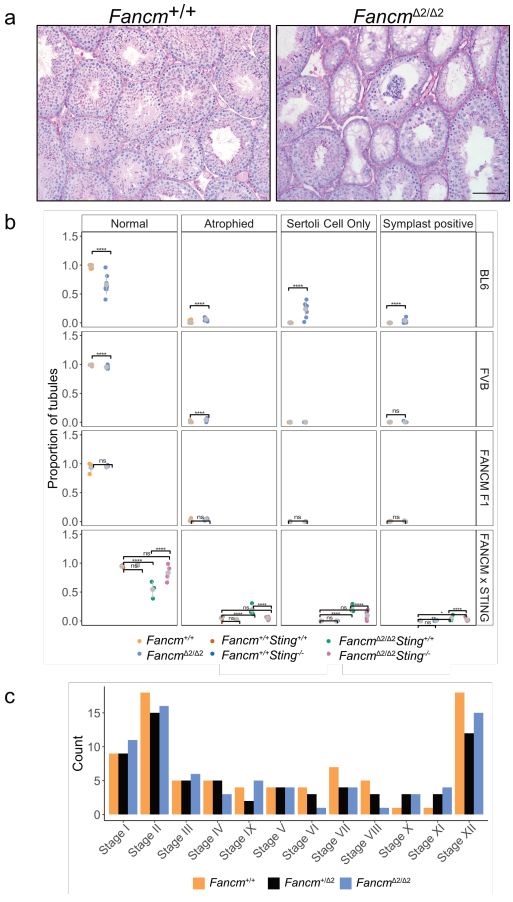
S17 Fig. **Increased histological defects *Fancm* seminiferous tubules. Related to Figure 3.** PAS-stained histological sections of seminiferous tubules were used for quantification of abnormalities. a) Bivalents align on the metaphase plate and b) segregate at anaphase I comparably in B6.*Fancm*$^{\Delta 2/\Delta 2}$ and B6.*Fancm*$^{+/+}$. c) The number cells in metaphase I and anaphase I in stage XII and stage I seminiferous tubules was comparable between genotypes. d) An increase in PAS-positive or pyknotic cells was observed in both stage XII and stage I tubules in B6.*Fancm*$^{\Delta 2/\Delta 2}$. e) PAS-positive pachytene and f) metaphase I/anaphase I were increased in B6.*Fancm*$^{\Delta 2/\Delta 2}$. g) The percentage of stage V tubules containing PAS-positive pachytene spermatocytes. h) The percentage of metaphase I-anaphase I cells spermatocytes that were PAS-Positive in Stage XII. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$.


Fancm+/+
Fancm Δ2/Δ2
a)
bivalents not on metaphase plate
ns
ns
XII
I
XII
I
tubule stage
b)
lagging anaphase I chromosomes
ns
ns
tubule stage
c)
number metphase I or anaphase I cells
ns
ns
tubule stage
d)
number of PAS-positive cells
***
***
tubule stage
e)
PAS-positive metaphase I or anaphase I cells
***
I
I
tubule stage
f)
number PAS-positive pachytene/tubule
*
V
V
tubule stage
g)
percentage stage V tubules pas positive
*
V
V
tubule stage
h)
percentage stage XII tubules pas positive
*
XII
XII
tubule stage

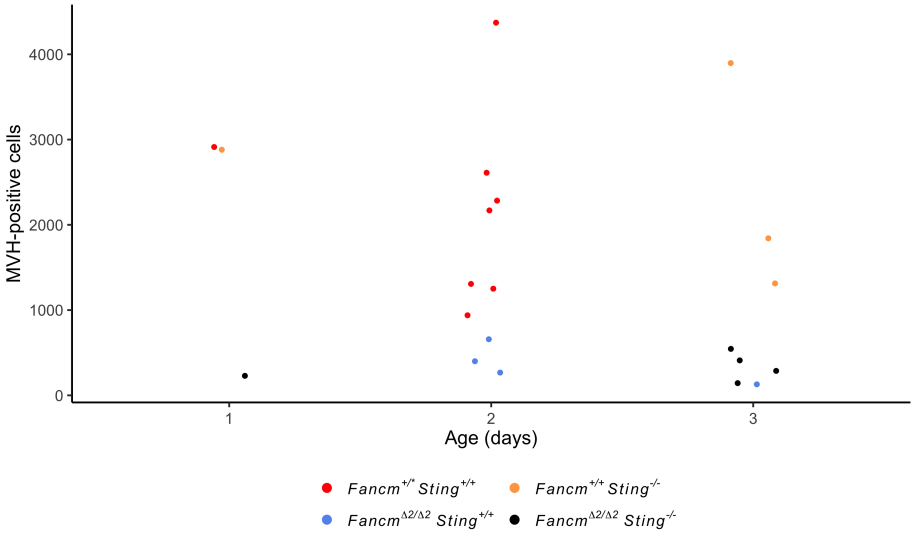
S18 Fig. **Sting depletion partially rescues gonad weights in *Fancm*-deficient mice. Related to Figure 5.** a) Testicular weight exhibited a subtle partial but significant rescue in $Fancm^{\Delta 2/\Delta 2}Sting^{-/-}$ compared to $Fancm^{\Delta 2/\Delta 2}$. b) epididymal weight and c) daily sperm production were not significantly different between $Fancm^{\Delta 2/\Delta 2}Sting^{+/+}$ and $Fancm^{\Delta 2/\Delta 2}Sting^{-/-}$. Due to limited epididymis sample numbers, there is low statistical power to detect difference. * indicates $p \leq 0.05$, ** indicates $p \leq 0.01$, *** indicates $p \leq 0.001$ and *** indicates $p \leq 0.0001$ (one-sided t-test with multiple testing correction "fdr").



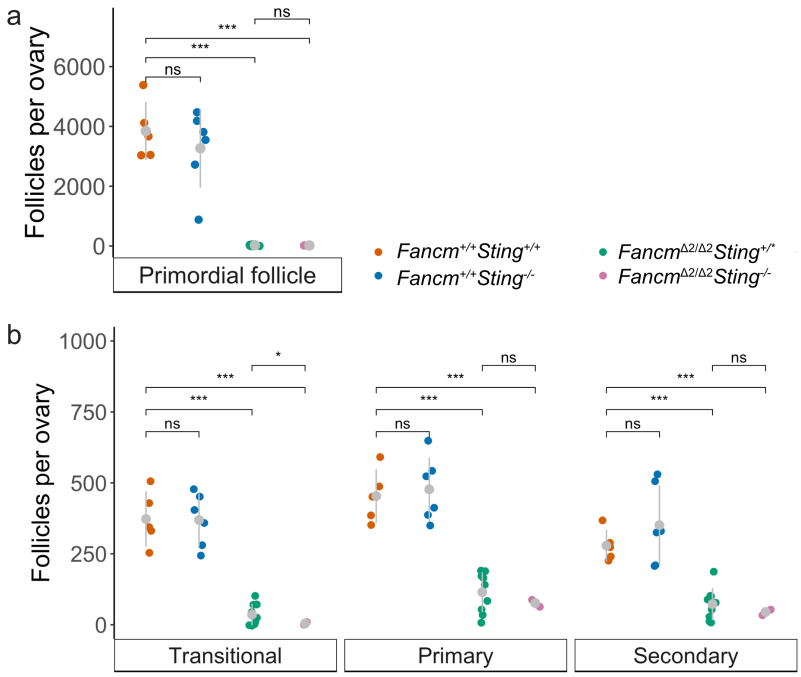
S19 Fig. ***Sting*** **depletion partially rescues histological defects in the seminiferous tubules in** ***Fancm*****-deficient mice. Related to Figure 5.** a) PAS stain of seminiferous tubules in B6.$Fancm^{+/+}$ and B6.$Fancm^{\Delta 2/\Delta 2}$ mice. Scale bar = 100 μm. b) Quantification of the PAS-labelled slides. Wild-type tubules are relatively replete in their appearance. Mutant mice have a heterogeneous reduction in germ cell content of the tubules. A mixture of normal, atrophied, symplast-positive and Sertoli cell-only tubules were be observed. c) Categorisation and quantification of seminiferous tubules of different spermatogenic stages in adult mice. **** indicates p-value ≤ 0.0001, ** indicates p-value ≤ 0.01, * indicates p-value ≤ 0.05 (pair-wised proportion test with multiple testing correction "fdr"). 127.625 ± 19.87 B6.$Fancm^{\Delta 2/\Delta 2}$ (8 mice), 174.13 ± 15.62 B6.$Fancm^{+/+}$ (8 mice), 109 ± 31.39 FVB.$Fancm^{\Delta 2/\Delta 2}$ (6 mice), 121 ± 12.68 FVB.$Fancm^{+/+}$ (6 mice), 168 F1.$Fancm^{\Delta 2/\Delta 2}$ (4 mice), 161 ± 82.76 F1.$Fancm^{+/+}$ (4 mice), 145 ± 15.56 $Fancm^{+/+}Sting^{+/+}$ (2 mice), 158.5 ± 9.19 $Fancm^{+/+}Sting^{-/-}$ (2 mice), 159.67 ± 17.5 $Fancm^{\Delta 2/\Delta 2}Sting^{-/-}$ (3 mice), 151.25 ± 10.34 $Fancm^{\Delta 2/\Delta 2}Sting^{-/-}$ (4 mice) tubules counted.



S20 Fig. **Sting depletion does not affect germ cell numbers in newborn *Fancm*-deficient male mice. Related to Figure 5.** Total MVH-positive cells detected in one testis per unique mouse of the indicated genotype. *Fancm*$^{+/*}$ indicates pooling of *Fancm*$^{+/+}$ and *Fancm*$^{+/\Delta 2}$ data as the number of germ cells per tubule was not different between the genotypes.



S21 Fig. **Sting depletion does not affect follicle numbers in *Fancm*-deficient female mice. Related to Figure 6.** a) Primordial and b) growing follicles were assayed in 10-day old female mice deficient for *Fancm*, *Sting* or both. Data shown are mean ± s.d. * indicates $p \leq 0.05$, *** indicates $p \leq 0.001$.



S22 Fig. **Ibuprofen does not improve gametogenesis, hypogonadism or genomic instability in *Fancm*-deficient mice. Related to Figure 5.** a) Quantification of H and E stain of seminiferous tubules in ibuprofen-treated B6.$Fancm^{\Delta 2/\Delta 2}$ mice and controls b) Average testes weights of 8 week old ibuprofen-treated and untreated control mice c) Average ovary weights of 8 week old ibuprofen-treated and untreated control mice d) Genomic instability was assayed by quantification of micronucleated red blood cells of ibuprofen-treated and untreated control mice. **** indicates p-value ≤ 0.0001, ** indicates p-value ≤ 0.01, * indicates p-value ≤ 0.05 (pair-wised proportion test with multiple testing correction "fdr"). 16 $Fancm^{+/+}$ and 24 $Fancm^{\Delta 2/\Delta 2}$ control mice. 8 $Fancm^{+/+}$ and 16 $Fancm^{\Delta 2/\Delta 2}$ mice used.

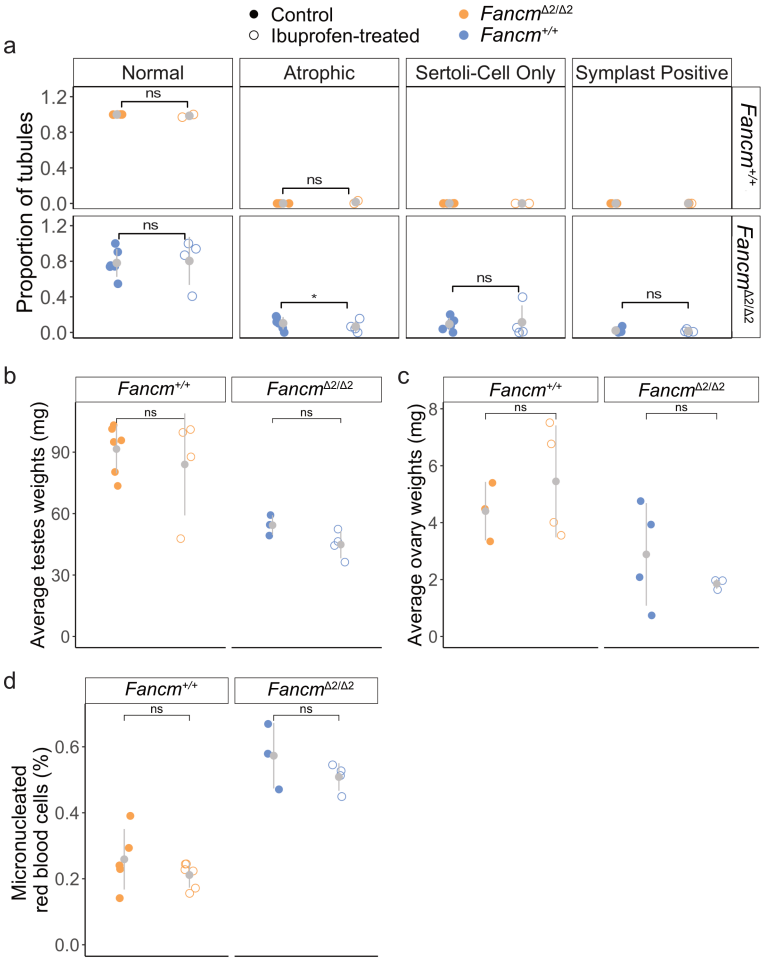


Table S1. Offspring of FVB.*Fancm*$^{\Delta 2/+}$ x FVB.*Fancm*$^{\Delta 2/+}$ intercrosses. Related to Table 1.

	Male		Female	
	Observed	Expected	Observed	Expected
Fancm$^{+/+}$	84	77	68	77
Fancm$^{\Delta 2/+}$	161	153	149	153
Fancm$^{\Delta 2/\Delta 2}$	70	77	81	77
Total	315		298	

Table S2. Offspring of FVB.*Fancm*$^{\Delta 2/+}$ x B6.*Fancm*$^{\Delta 2/+}$ intercrosses. Related to Table 1.

	Male		Female	
	Observed	Expected	Observed	Expected
Fancm$^{+/+}$	58	53	57	53
Fancm$^{\Delta 2/+}$	115	106	92	106
Fancm$^{\Delta 2/\Delta 2}$	56	53	45	53
Total	229		194	

Table S3. Meiotic prophase substaging with SYCP3 is unchanged in F1.*Fancm*$^{\Delta 2/\Delta 2}$ compared to F1.*Fancm*$^{+/+}$. Data is from two mice per genotype. Related to Figure 2.

	leptotene	zygotene	pachytene	diplotene	diakinesis
Fancm$^{+/+}$	1	35	160	31	8
Fancm$^{\Delta 2/\Delta 2}$	9	24	170	36	11

Table S4. Meiotic metaphase univalent counts is unchanged in F1.*Fancm*$^{\Delta 2/\Delta 2}$ compared to F1.*Fancm*$^{+/+}$. Data is from seven and eight mice for the wild type and mutant respectively. Related to Figure 3.

	balanced metaphase I	two unbalanced univalents
Fancm$^{+/+}$	184	7
Fancm$^{\Delta 2/\Delta 2}$	273	2

Table S5. B6.$Fancm^{\Delta 2/\Delta 2}$ have absent populations of pachytene cells, which is not detected in B6.$Fancm^{+/+}$. Related to Figure 4. Data is from four mice of each genotype. X-squared = 19.314, df = 1, p-value = 1.109e-05

	pachytene present I	pachytene absent
$Fancm^{+/+}$	106	0
$Fancm^{\Delta 2/\Delta 2}$	99	22

Calling crossovers using Hidden Markov Model

HMM was applied to infer strain of origins for the list of informative SNP markers (hetSNPs, SNPs with different alleles between the two mouse strains) and call crossovers in each mouse sample or sperm cell. Different parameter settings were applied for the bulk samples and sperm cells.

1. BC1F1 bulk samples
   - `Observations`. The allele specific counts (c_f, c_b) across the informative SNP markers (hetSNPs) were abstracted from the variant calling result for each chromosome in each sample. Filters were applied to include sites with `DP > 2 & DP < 30 & GT=="0/1"| GT=="1/1" & MQ>30`.
   - `States`. At each SNP site s_i, there are two hidden states $s_i = 0$ or $s_i = 1$ corresponding to two genotypes, homozygous alternative (FVB/N) / (FVB/N) and heterozygous (C57BL/6J) / (FVB/N). $s_i = 0$ corresponds to (FVB/N) / (FVB/N) while $s_i = 1$ corresponds to (C57BL/6J) / (FVB/N).
   - `Emission probabilities`. Two binomial distributions were used for modelling the observed FVB/N allele counts across the SNP sites. Conditioning on different hidden states, for each site i

$$c_n = c_f + c_b$$

$$c_f|s = 0 \sim Binomial(c_n, \theta_0)$$

$$c_f|s = 1 \sim Binomial(c_n, \theta_1)$$

     Intuitively, the values for θ_0 and θ_1 can be set as 1 and 0.5. For this project, high coverage DNA sequencing (20X) were performed on an inbred FVB/N mouse and a F1 hybird mouse sample which were used to estimate the θs for each site. FVB/N allele count rates calculated from FVB/N mouse were used for estimating $theta_0$ while FVB/N allele count rates calculated from the F1 mouse were used for estimating θ_1. SNP sites were further filtered if the sites with DP >= 100, or their estimated θ_0 were smaller than 0.9 or their estimated θ_1 larger than 0.8. A beta prior $Beta(8, 2)$ was applied to eliminate extreme values for estimated $theta_0$s. The MAP (Maximum a Posteriori) estimates were used finally.

- `Transition Probabilities`. A distance-dependent transition probability model were applied that corresponds to an average of 0.5 cM per 1M base pairs[3]. p_{ij} is the transition probability of transitioning to a different state at SNP j from SNP i.

$$p_{ij} = 1 - exp(-d_{ij} * 0.5 * 10^{-8})$$

- `Initial probabilities`. The initial probabilities for the two hidden states were set to be both 0.5 since they were equally likely to happen.

2. Sperm cells
   - `Observations`. The allele read counts across the informative SNP markers (hetSNPs) were counted for each chromosome in each sperm cell.
   - `States`. Sperm cells have haploid genomes. There are also two hidden states each corresponding to a C57BL/6J or FVB/N segment in the haploid genome and shifts between the two states indicate meiotic crossovers from the F1 donor. At each SNP sites i, there are two hidden states, $s_i = 0$ corresponds to FVB/N sequence while $s_i = 1$ corresponds to C57BL/6J sequence.
   - `Emission probabilities`. Similar to bulk BC1F1 samples, two binomial distributions were used for modelling the emission probabilities.

$$c_n = c_f + c_b$$

$$c_f|s = 0 \sim Binomial(c_n, \theta_0')$$

$$c_f|s = 1 \sim Binomial(c_n, \theta_1')$$

   θ_0' and θ_1' has been set as 0.1 and 0.9 respectively.
3. `Transition Probabilities`. A smaller transition probability was applied to data from single sperm cells compared to the BC1F1 samples to deal with the larger uncertainty and more noise in the single sperm cell data set. This transition probability was configured using *–cmPmb* option via *sgcocaller*.
4. `Initial probabilities`. Equal probabilities (0.5) for two states were used.

The HMM was applied to F1 single sperm sequencing data using the customized command line tool sgcocaller with options: *–cmPmb 0.0001 –maxDP 6 –maxTotalDP 30* for each F1 donor.

Crossover analysis

The hidden state sequences obtained from applying HMM model were further analyzed in R (R Core Team, 2021) with *comapr*[6], to filter out potential false positive crossovers, find crossover intervals, calculate crossover rates, calculate genetic distances and visualize data. To exclude false crossovers, we defined a

quantitative measure for crossover calling confidence *logllRatio*, which was calculated and assigned to each crossover introducing hidden state sequence. The value was derived by taking the log ratio of probabilities for SNPs of the assigned haplotype state versus the their probabilities if their states were reversed. For BC1F1 samples, the *logllRatio* threshold was set to 150, and crossovers that were not supported by 30 SNPs (*minSNP*) from either end, or the distance to the next crossover staring site were shorter than 1,000 base pairs (*bpDist*) were excluded. Samples with chromosomes that had coverage of less than 1,000 SNPs were filtered out (only one sample was excluded using this criteria). With *logll-Ratio* the threshold was set to 150, and changing (*bpDist*) to 20, 1k or 100k bp, or changing *minSNP* to 30, 100, 200, or 300 did not change the crossover distributions observed in BC1F1 samples.

For F1 single sperm sequencing dataset, cells with chromosomes that had coverage of fewer than 500 SNPs were removed, or with chromosomes that had been called 10 or more crossovers; an indication of doublet cells, abnormal chromosomes, or poor read quality. Crossovers were filtered out if they were introduced by state sequence with *logllRatio* $<= 50$, or not supported by more than 10 SNPs from either side of the crossover, or the base pair distance to the starting site of the next crossover were shorted than 100k bp.

Permutation testing for testing genetic distances in genotype groups

Permutation testing (with replacement) were performed by randomly permuting the group labels of samples to test the difference in genetic distances including total genetic distances across the genome, total genetic distances per chromosome or genetic distances in physical chromosome bins.

Permutation for testing differences in total genetic distances in $Fancm^{\Delta 2/\Delta 2}$ versus the $Fancm^{+/*}$ was done by calling the `permuteDist` function from `comapr`. It implements the following steps[1]:

1. Find the current observed difference in the two groups, d_{obs}.
2. Take group labels vector and permute the group labels by randomly assigning the labels across all samples and calculate the new difference with the newly generated grouping.
3. Repeat step 2 for B times (i.e $B = 1,000$).
4. Calculate the corrected P value by using `permp` function from `statmod`[1].

Multiple testing correction (when testing multiple chromosomes or multiple chromosome bins) was applied using *p.adjust* with method “fdr" function from *stats* in R (4.1.0) (R Core Team, 2021) .

Generating null distributions for inter-distances of double crossovers

The null distribution for inter-distances of double crossovers was generated by permuting the observed crossovers across samples.

For each genotype group, chromosomes with exactly two crossovers were found and the inter-crossover distances were calculated by finding the base pair

distances from the mid points of the two crossover intervals. To generate the expected distribution of inter-crossover distances without interference within each genotype group, the crossover positions were permuted across the samples for each chromosome for 1,000 times. Histograms of the expected and observed inter-crossover distances were plotted using R (R Core Team, 2021).

Bibliography

[1] Phipson, B. and Smyth, G. K. (2010). Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *Statistical Applications in Genetics and Molecular Biology*, 9(1):Article 39. 10.2202/1544-6115.1585.

[2] Smit, A., Hubley, R., and Green, P. (2013-2015). Repeatmasker open-4.0.

[3] Hinch, A. G., Zhang, G., Becker, P. W., Moralli, D., Hinch, R., Davies, B., Bowden, R., and Donnelly, P. (2019). Factors influencing meiotic recombination revealed by whole-genome sequencing of single sperm. *Science*, 363(6433):eaau8861. 10.1126/science.aau8861.

[4] Pratto, F., Brick, K., Cheng, G., Lam, K.-W. G., Cloutier, J. M., Dahiya, D., Wellard, S. R., Jordan, P. W., and Camerini-Otero, R. D. (2021). Meiotic recombination mirrors patterns of germline replication in mice and humans. *Cell*. 10.1016/j.cell.2021.06.025.

[5] Bhutani, K., Stansifer, K., Ticau, S., Bojic, L., Villani, A.-C., Slisz, J., Cremers, C. M., Roy, C., Donovan, J., Fiske, B., et al. (2021). Widespread haploid-biased gene expression enables sperm-level natural selection. *Science*, 371(6533). 10.1126/science.abb1723.

[6] Lyu, R., Tsui, V., Crismani, W., Liu, R., Shim, H., and McCarthy, D. J. (2022). sgcocaller and comapr: personalised haplotype assembly and comparative crossover map analysis using single-gamete sequencing data. *Nucleic acids research*, 50(20):e118. 10.1093/nar/gkac764.