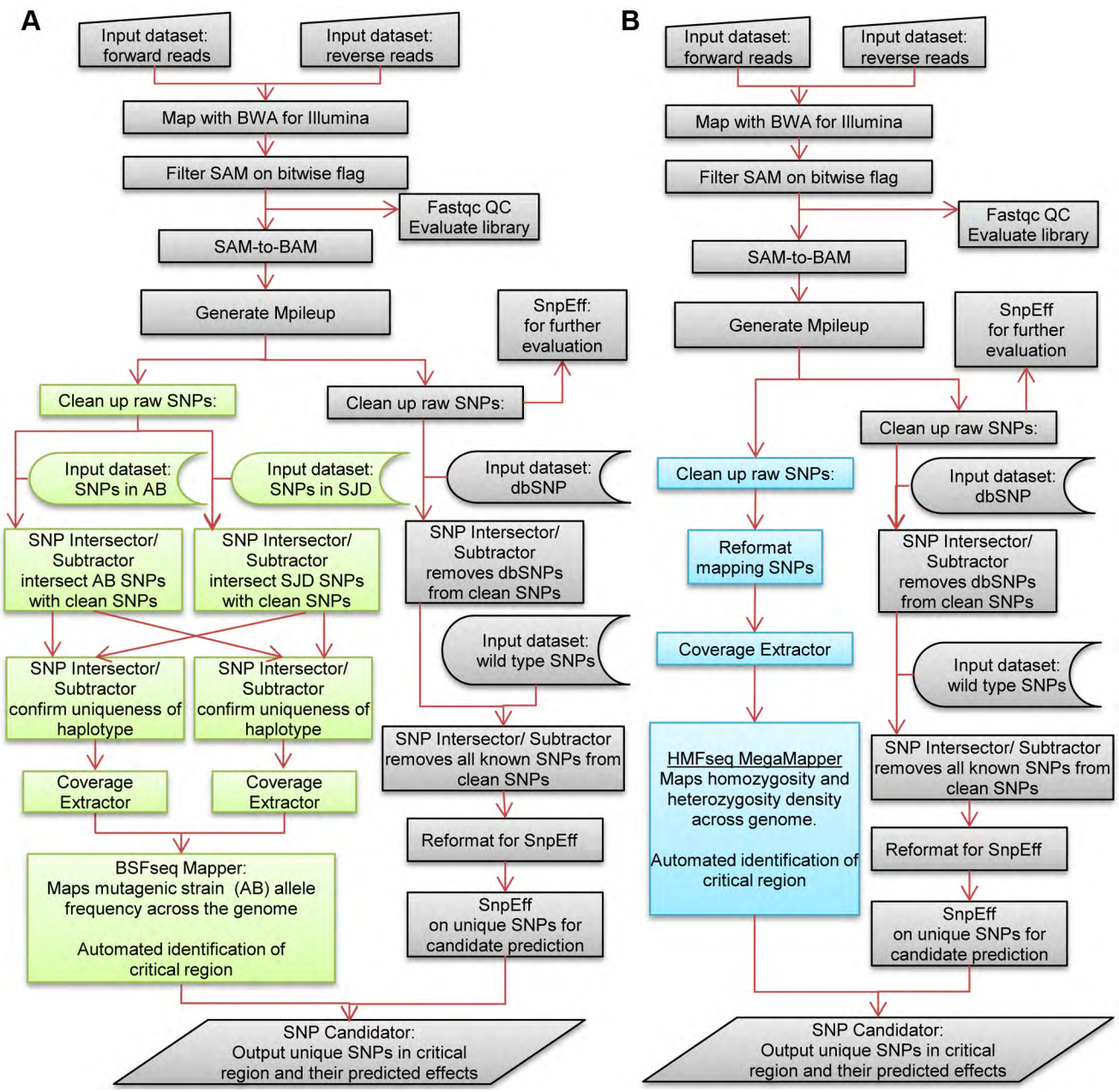

Fig. S1. MegaMapper, a zebrafish community bioinformatics resource for positional cloning by BSFseq or HMFseq. (**A**) BSFSeq flowchart. After read processing and mapping to the reference genome, BSFSeq intersects the candidate SNP library with a strain-specific SNP list to determine haplotype frequency in intersecting positions. (**B**) HMFSeq flowchart. After read processing and mapping to the reference genome, HMFSeq intersects homozygous and heterozygous SNP densities within the candidate library to determine the degree of homozygosity (hom/het ratio) in intersected positions.

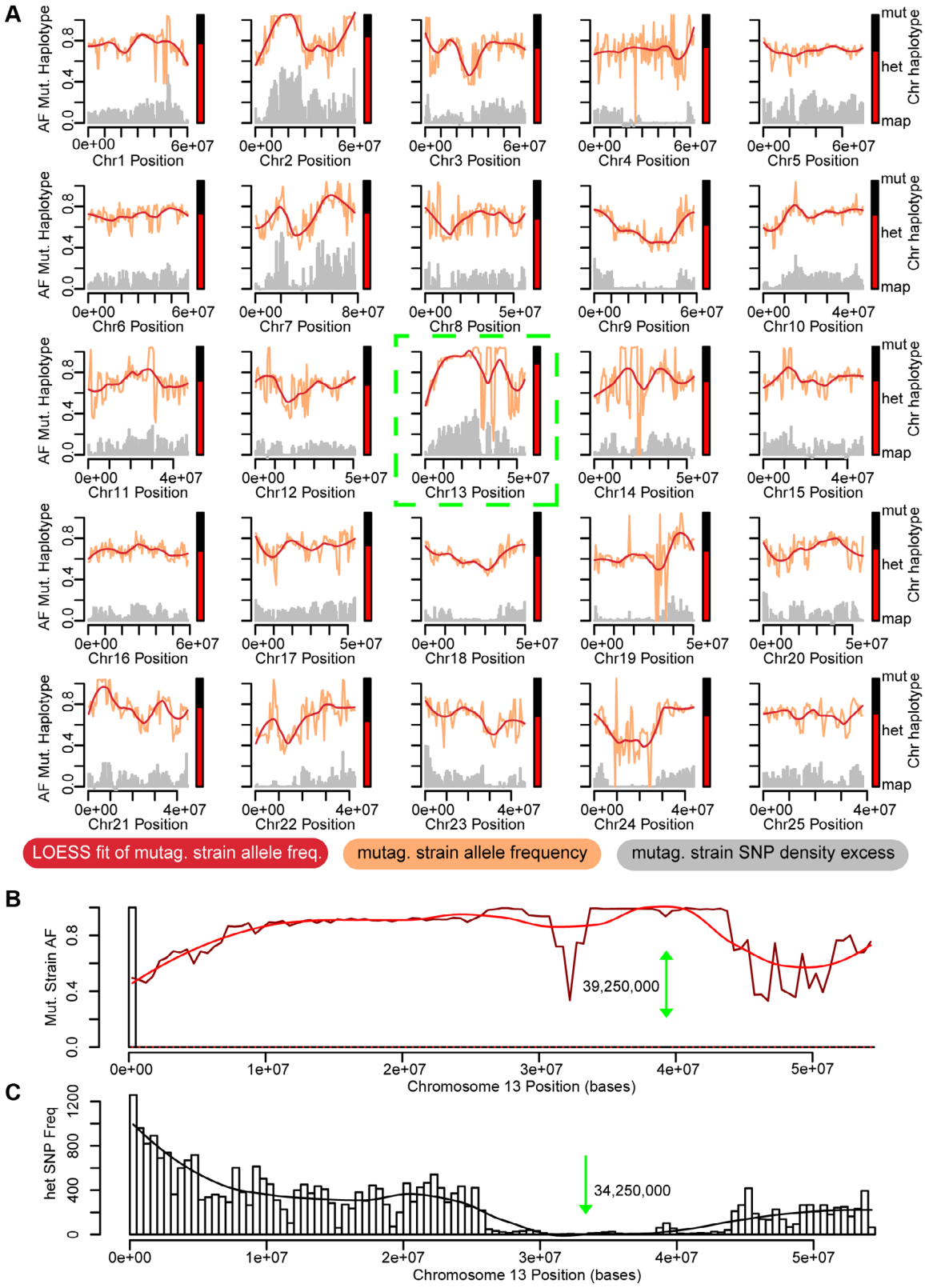

Fig. S2. Bulk segregant linkage mapping and filtering of *jj59/jagged1b*. (**A**) Mapping of regional AB haplotype allele frequencies onto individual chromosomes (AF Mut. Haplotype). Chromosomal position intervals of ~200 kb are plotted in orange. A LOESS fit of the haplotype interval data is plotted in red. Histogram data show the measured local SNP density of AB-haplotype SNPs (gray). Bars next to individual chromosomes show averaged chromosomal mutagenesis strain allele frequency (red, vs mapping strain AF in black). (**B**) Allele frequency scan of the candidate chromosome 13 (highest bulk allele frequency, green box in A) position by AB haplotype allele frequency (dark red) and the corresponding fit (red line). The maximum of the fit is indicated (green arrow). (**C**) The panel shows a SNP density histogram of all detected heterozygous SNPs on chromosome. Loss of heterozygosity is calculated by fitting to the density histogram (black line). The minimum of the fit is indicated (green arrow). After filtering of known wild-type SNPs, only one homozygous SNP remains. The predicted effects of this SNP denote a disrupted SPLICE_DONOR for *jagged1b* (Table 3).

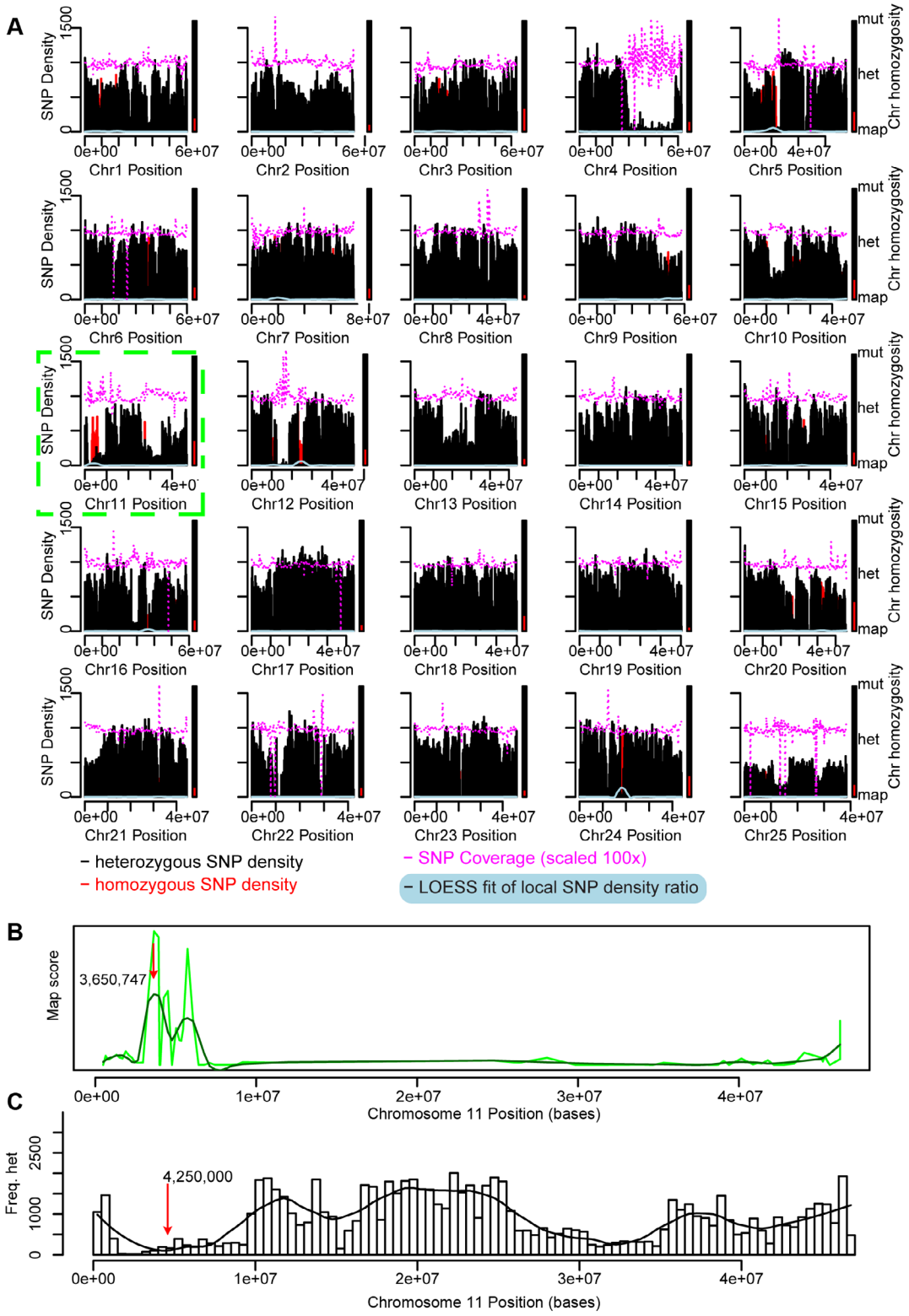

Fig. S3. Homozygosity mapping and filtering of *tm290d/lhfpl5a*. (**A**) Chromosomal SNP density and homozygosity map of *tm290d/lhfpl5a*. Red bars are homozygous SNP counts, black bars are heterozygous SNP counts. Homozygous-to-heterozygous ratios were fitted for trend recognition using local regression (LOESS; light blue). Bars next to individual chromosomes show the degree of homozygosity (red bar, vs heterozygosity in black). Chromosome 11 shows the highest degree of homozygosity and thus probably harbors the mutation. (**B**) Scan of chromosome 11 for a peak in local homozygosity and fit peaks at 3,650,747 (red arrow). (**C**) A SNP density histogram of all detected heterozygous SNPs on chromosome 11. Loss of heterozygosity is calculated by fitting to the density histogram (black line). The minimum of the fit is indicated (red arrow). The average of the resulting values of B and C were used to position the center of the critical interval. Variant effect prediction of all detected homozygous SNPs within the critical interval after filtering reveals a shortlist of only one candidate SNP remains. SNP 11_2815634_T/A introduces a stop codon at residue 80 of Lhfpl5a (Table 3).

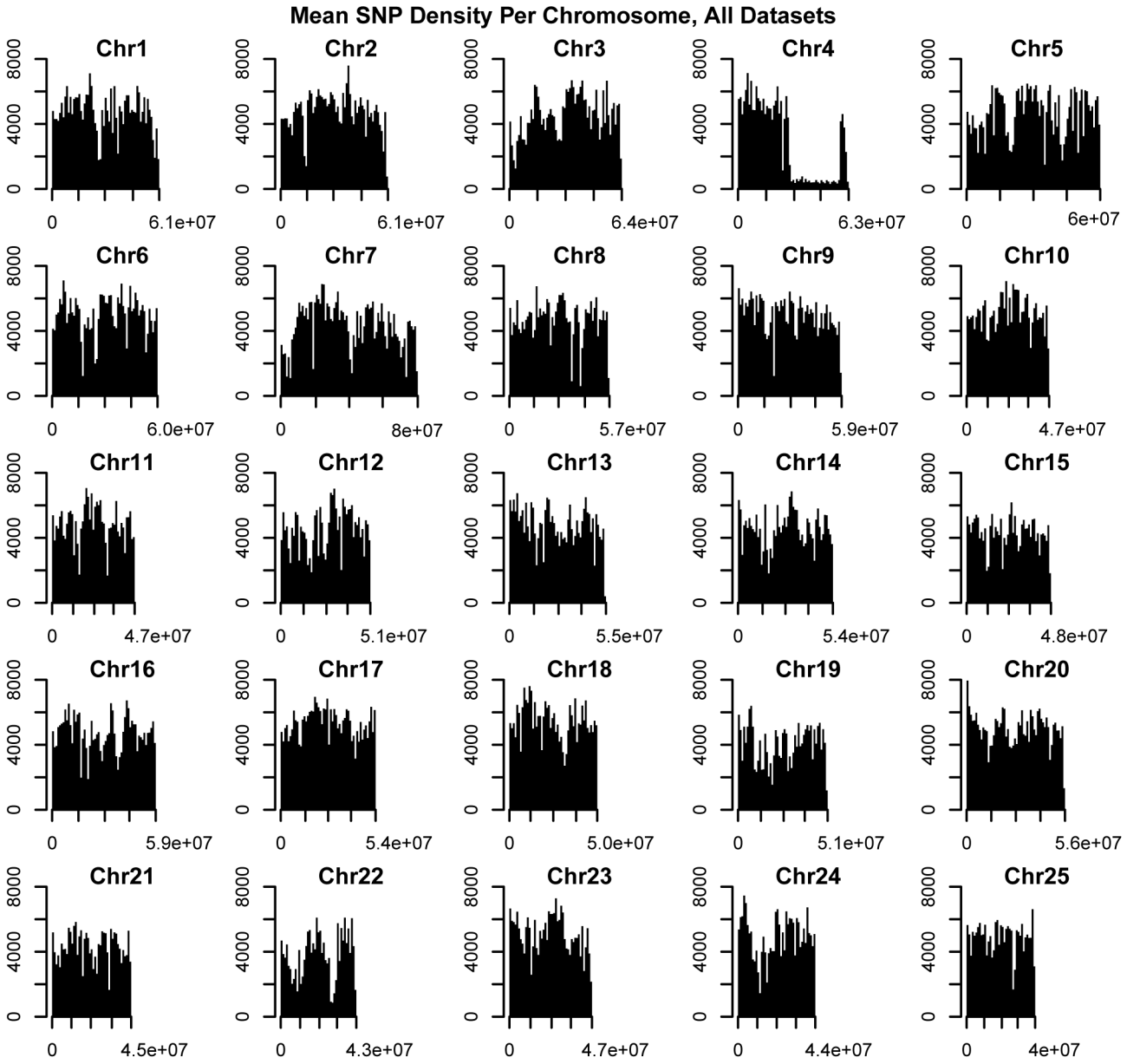

Fig. S4. Mean SNP density per chromosome revealed uneven genome-wide distribution of SNPS. The SNPs from all datasets were combined and plotted across each chromosome as SNP densities, the number of SNPs falling within the bin. Although there is a very large SNP desert on chromosome 4, every chromosome has regions of low SNP density and regions of higher SNP density that probably distort the accuracy of all mapping efforts.

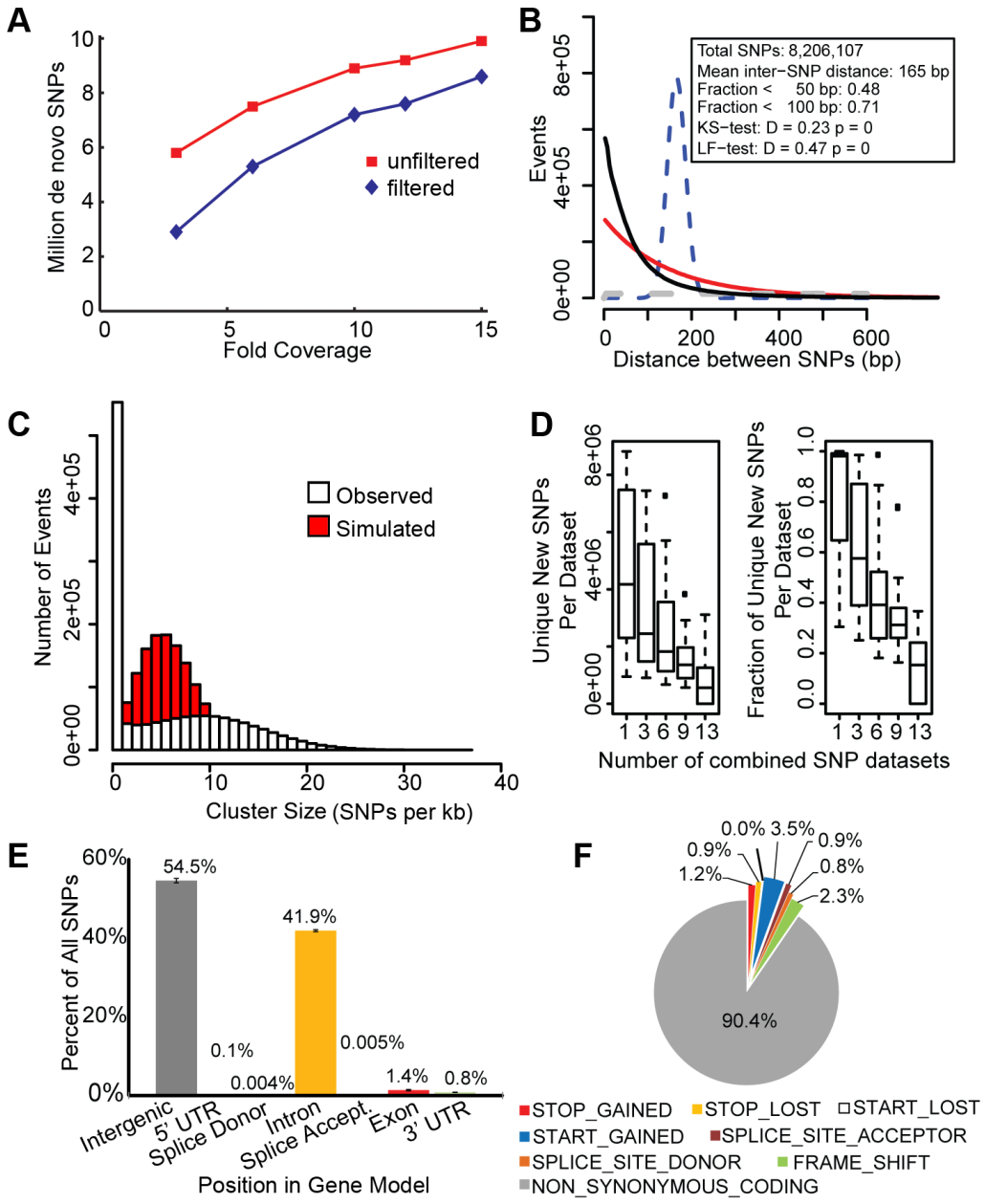

Fig. S5. Towards a saturating analysis of the zebrafish SNP universe. (**A**) Saturation of detected de novo SNPs per library at high coverage (*n*=14 libraries). (**B**) Distribution of inter-SNP distance (ISD) in sequencing data (black line) is skewed towards shorter ISD with a mean of 165 bp. Simulations (*n*=10) of the randomized distribution of an equal number of SNPs (red line) differs from real data (D=0.23, *P*=0), tight normal distribution around the mean of the measured data with an artificial standard deviation of one-eighth of the mean (blue dotted line), normal distribution around 165 bp using the measured s.d. (gray dashed line). (**C**) Histogram of observed and simulated SNP clusters (SNPs per kb) indicates that SNPs cluster more than they would by chance to produce regions of high SNP density and SNP deserts. (**D**) The number of new SNPs per genome decreases as a function of the number of genomes considered, indicating saturation. Left, absolute number of newly discovered SNPs versus number of combined datasets. Right, fraction of newly discovered SNPs versus number of combined datasets. (**E**) The distribution of detected SNPs is biased against exons and their splice site, supporting Zv9 gene annotation and SNP ascertainment (see also supplementary material Fig. S6). (**F**) Average distribution (*n*=10 datasets) of predicted non-silent SNP effects shows a small but reproducible number of deleterious effects in wild-type and mutant genomes alike according to Zv9 gene annotations.

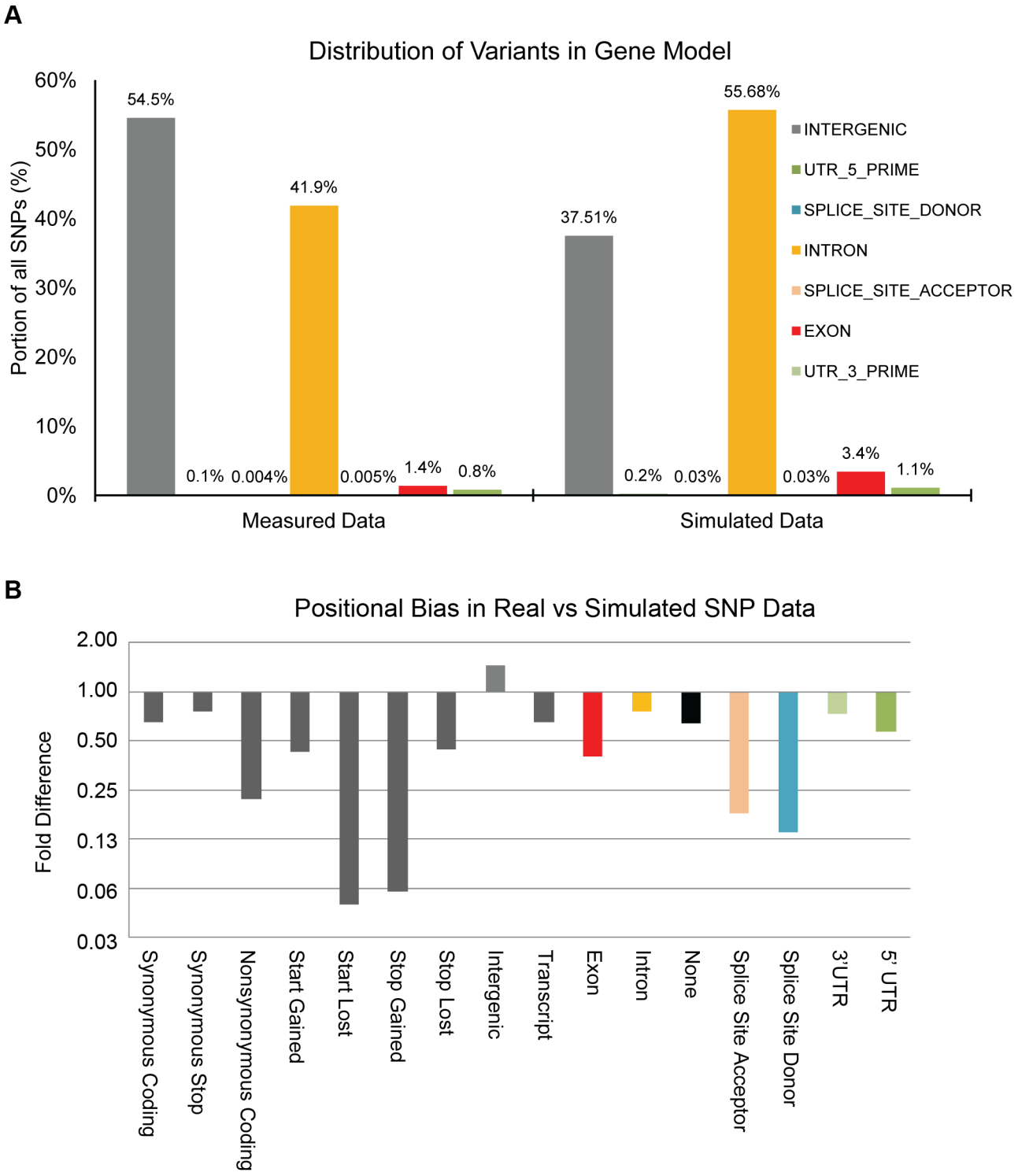

Fig. S6. A bias for intergenic regions in the distribution of variants with respect to gene architecture. We compared the distribution of the total observed SNPs to a size matched distribution of randomly simulated loci. (**A**) We observed a greater than random portion of SNPs within intergenic regions of the genome. By contrast, all genic regions were under-represented compared with our simulated set of loci. (**B**) The non-random distribution SNPs in our datasets produced a negative enrichment for variants that would alter gene expression or function.

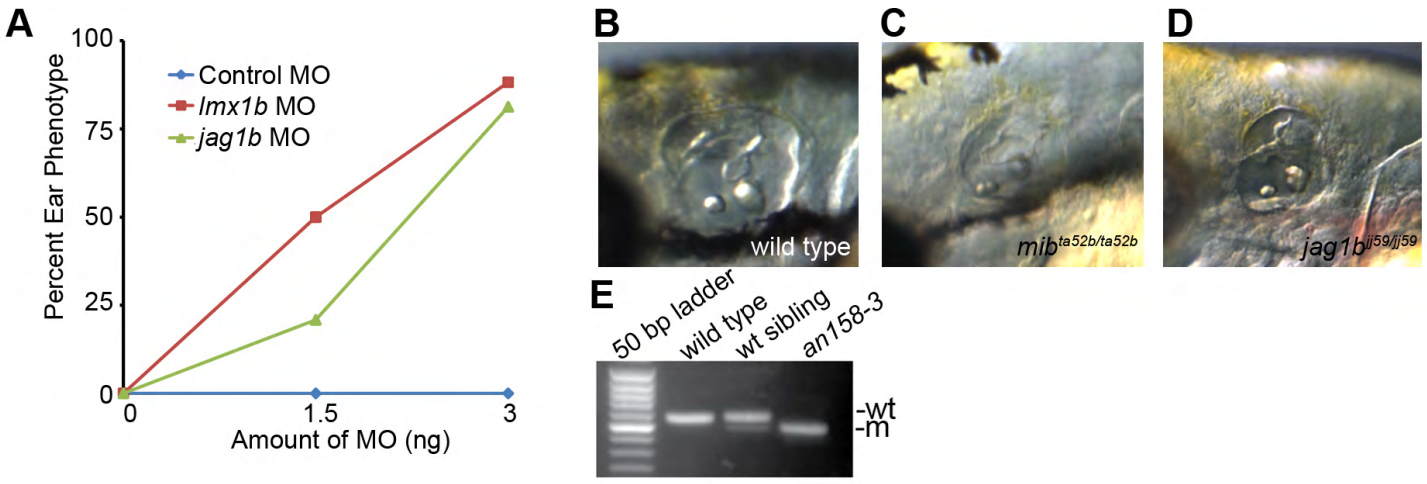

Fig. S7. Additional validation of mutants supported their computed identities. (**A**) We observed a dose-dependent phenocopy of the mutant phenotypes with morpholinos against *lmx1b* and *jag1b*. (**B-D**) Compared with wild type (B), the ear phenotypes of *mib*$^{ta52b/ta52b}$ (C) *and jag1b*$^{j59/jj59}$ (D) resemble one another and support the overlapping roles for both gene products in Notch signaling. (**E**) RT-PCR of *cdh23* transcript in wild type, sibling and mutant reveals mis-splicing in *an158-3(cdh23nl9)* transcripts.

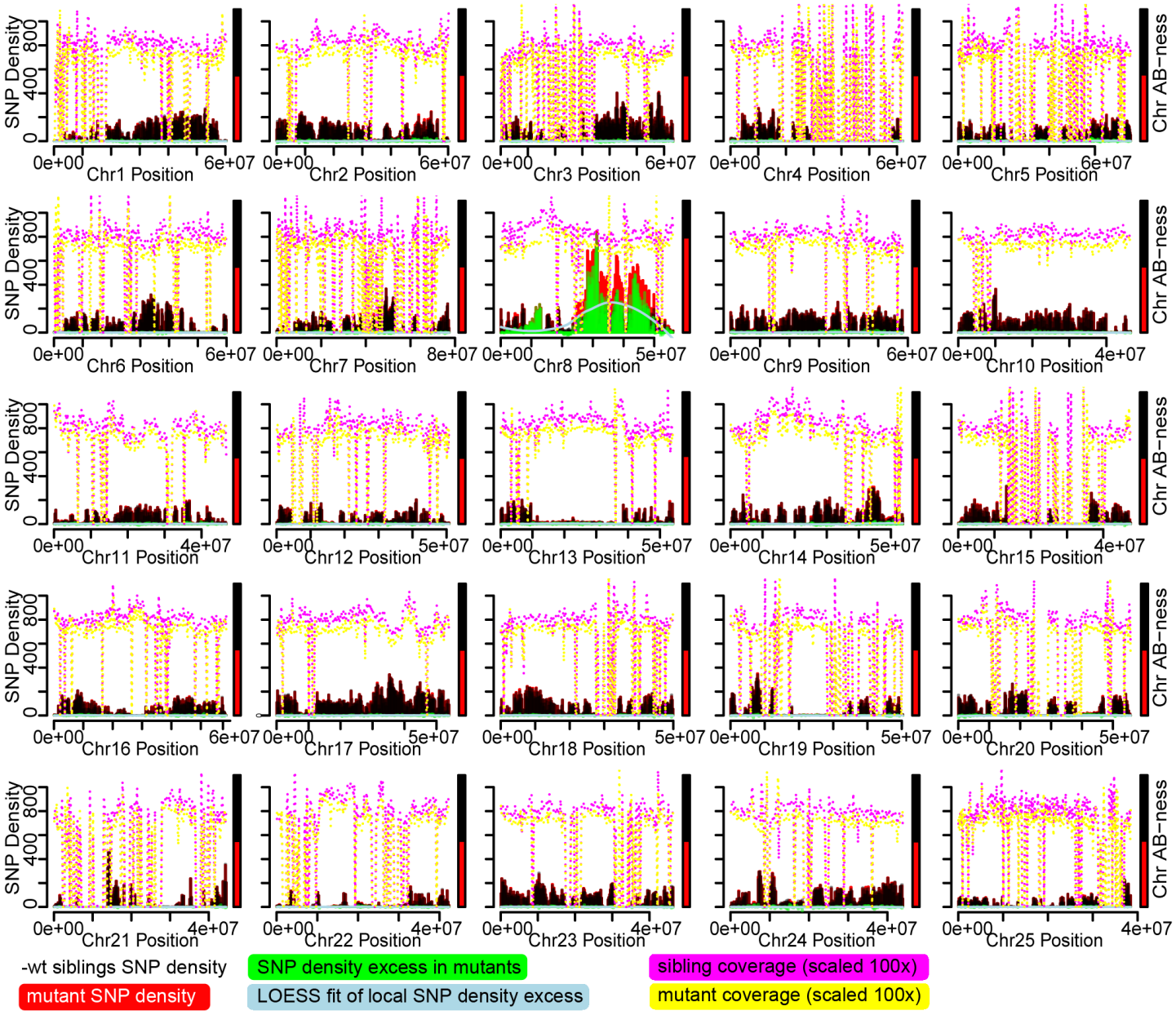

Fig. S8. Subtracting SNPs of wild-type siblings from the *jj410* mutant dataset cleaned up the mapping noise for *jj410*, but was not necessary. SNP density was plotted across the chromosomes for wild-type (SJD, black) and mutant (AB, red) SNP density. After subtraction, only chromosome 8 had a total AB density much higher than 50% (red and black bars to the right of each chromosomal plot, compare to Fig. 2A). Additionally, SNP density excess of AB (green) was only significant on chromosome 8. We also included plots of wild-type sibling coverage (magenta) and mutant coverage (yellow) to demonstrate that both libraries shared similar coverage biases. For instance, chromosome 4, which contains a large SNP desert, also has an overlapping region of intermittent coverage. The reason(s) for this are unknown, but may have something to do with biases against heterochromatin and/or elevated representation of repeats or difficult to map sequence.

Table S1. Morpholino and DNA oligo sequences.

Description	Sequence (5′ to 3′)
Standard control morpholino	CCTCTTACCTCAgTTACAATTTATA
lmx1b.1 splice site morpholino	TTGAAGGACTTACCGAGCATAACTC
jag1b splice site morpholino	AATAGTCTTTCTTACGGGAGTGGC
lmx1b.1 in situ probe forward	CCATCGACCGACTCTACTCCATGCAG
lhfpl5a ATG morpholino	CAGATAGCATTTTCGCCATGTTTGC
lhfpl5a splice site morpholino	GCTGGAGATGAAAAACACACTCAAT
lmx1b.1 in situ probe reverse	TAATACGACTCACTATAGGGAGACATCGTCTTTCAGAAGGGCTAAAC
jag1b in situ probe forward	CGAGGGCAAGAACTCCATCATTGC
jag1b in situ probe reverse	TAATACGACTCACTATAGGGAGACTCCTTTCCGACAATTGCTGTGGTG
lhfpl5a in situ probe forward	CCTGAAGATTTAGGTGACACTATAGAAGAGAAATGCTATCTGCCC AAGAGGCTGCCAAGA
lhfpl5a in situ probe reverse	GCCTGAATAATACGACTCACTATAGGGAGAACCGTACAGTTGCCC AGCGTGTATTTGTCT
lhfpl5a RT-PCR/ genotyping F	AAATGCTATCTGCCCAAGAGGCTGCCAAGA
lhfpl5a RT-PCR/ genotyping R	ACCGTACAGTTGCCCAGCGTGTATTTGTCT
an158-3/cdh23nl9 genotyping F	TGGTGCTGTAACAATGGCCCTTCA
an158-3/cdh23nl9 genotypingR	AGCTCTGCATTTAGACCCGCGTCATT
an158-3/cdh23nl9 RT-PCR For	TTCCTATTGGATCGTCTGTGTTCAGGGTGCAAGTCC
an158-3/cdh23nl9 RT-PCR R	GCCCGCTGGACCGTTATCGATGGCTTC
lmx1b.1 genotyping forward	GAAGGCTCGTCTCTGCTGTGTGGTG
lmx1b.1 genotyping reverse	CGTTATGGATGCGCTGAGACTGAATACC
jag1b genotyping forward	GACATAGATGACTGCAGCTTGAACC
jag1b genotyping reverse	GCATGTAGCTTCGTCACACTGGC

Table S1. Morpholino and DNA oligo sequences.

Table S2. Strain-specific SNP markers

Genome	SNPs vs REF	HOM SNPs vs REF	Unique (all)	Unique (hom)	Coverage (mean)
AB	13.6 E+06	9.0 E+06	4.1 E+06	2.5 E+06	12
TU	9.3 E+06	2.1 E+06	1.8 E+06	0.5 E+06	5
WIK	13.8 E+06	4.4 E+06	4.5 E+06	1.0 E+06	5
SJD	3.0 E+06	2.3 E+06	0.7 E+06	0.4 E+06	4
TLF	6.9 E+06	3.4 E+06	1.6 E+06	0.6 E+06	5
COMBINED	25.4 E+06	14.8 E+06	NA	NA	9
UNIVERSAL	0.5 E+06	0.4 E+06	NA	NA	NA
jj59	6.0 E+06	2.3 E+06	1.0 E+06	0.4 E+06	7
*jj410*_WT	9.2 E+06	2.7 E+06	1.6 E+06	0.4 E+06	8
*jj410*_Mut	8.9 E+06	2.70E+06	1.5 E+06	0.4 E+06	8
a158-3	4.2 E+06	1.6 E+06	0.6 E+06	0.3 E+06	5
tm290D	7.5 E+06	1.1 E+06	1.3 E+06	0.2 E+06	8
Average	8.3 E+06				
Standard deviation	3.5 E+05				

NA, not applicable.

Table S2. Strain-specific SNP markers. This table summarizes the SNPs that we identified in the wild-type and mutant strains that we sequenced and/or analyzed. SNPs were identified relative to the Zv9 genome assembly from Sanger, which is based on TU. Homozygous SNPs were defined as having at least three mapped sequences at the loci and all sequences having the same variant. Unique variants were defined as SNPs present only in that particular dataset and absent from all other dataset. The mean coverage is an effective coverage calculated by the Snpeff tool.

Table S3. Number of detected SNPs per genome

Library	DP3.filter.SNPs	ab.initio.calls	PF.Reads	cov
*AB*_SANGER	8,637,221	9,855,280	270,000,000	15
*AB*_HARRIS	7,579,548	10,141,300	90,000,000	4.6
*TU*_SANGER	2,378,304	3,946,057	186,000,000	11
*TUB*_HARRIS	5,550,176	7,284,290	97,000,000	5.1
*TUG*_HARRIS	3,954,060	6,035,650	81,000,000	4.1
*TLF*_HARRIS	6,845,445	9,862,938	91,000,000	3.8
*WKB*_HARRIS	7,720,936	10,757,717	96,000,000	4.1
*WKG*_HARRIS	7,565,379	10,725,557	81,000,000	4
*SJD*_MEGASON	3,184,219	8,481,074	97,670,816	6
*jj410_sjd*_sib_MEGASON	9,224,092	11,808,490	141,622,193	8
*jj410_sjd*_mut_MEGASON	8,926,680	11,593,392	131,531,446	8
*jj59*_MEGASON	6,017,281	11,843,236	117,941,618	7
*tm290d*__MEGASON	7,520,301	9,748,539	129,792,188	7
*a158-3*__MEGASON	4,167,428	8,260,138	126,796,756	7

Table S3. Number of detected SNPs per genome. This table summarizes each dataset used in the study. DP3 filter SNPs were sequenced at least three times in the particular library. The number of SNPs prior to filtering for number of times sequenced (depth) was the ab.initio.calls. PF reads passed our sequencing and mapping quality filters. The coverage (cov) is an effective coverage calculated by the Snpeff tool.

Table S4. Effect of 10% contamination of wild-type sequences in the mutant pool

Sample	Map mode	Chr	Mapping position	Distance from lesion
Control	Het_Valley	8	36,000,000	–1,870,889
Control	Density_ratio	8	34,726,299	–597,188
Control	Compromise	8	35,363,150	–1,234,039
Trial 1	Het_Valley	8	37,000,000	–2,870,889
Trial 1	Density_ratio	8	34,726,299	–597,188
Trial 1	Compromise	8	35,863,150	–1,734,039
Trial 2	Het_Valley	8	36,250,000	–2,120,889
Trial 2	Density_ratio	8	34,726,299	–597,188
Trial 2	Compromise	8	35,488,150	–1,359,039
Trial 3	Het_Valley	8	36,750,000	–2,620,889
Trial 3	Density_ratio	8	34,726,299	–597,188
Trial 3	Compromise	8	32,738,150	1,390,961

Original measurement (no contamination)				
Map Mode	**POS**	**STDEV**	**DIST.F.LESION**	**SHIFT (bp)**
Het_Valley	36,250,000	NA	–1,870,889	NA
Density_ratio	34,726,299	NA	–597,188	NA
Compromise	34,707,856	NA	–1,234,039	NA
10% contamination, simulated (three trials)				
Map Mode	**AVG**	**STDEV**	**DIST.F.LESION**	**SHIFT (bp)**
Het_Valley	36,666,667	381,881	–2,537,556	–666,667
Density_ratio	34,726,299	0	–597,188	0
Compromise	34,696,483	1,706,300	–567,372	666,667

SNP effect

Percentage contamination	Number	Chr	Pos	REF	ALT	Qual	DP	AF	Type
0	0	8	34129111	T	A	222	11	1.00	hom
10	1	8	34129111	T	A	224	13	0.85	het
10	2	8	34129111	T	A	222	11	1.00	hom
10	3	8	34129111	T	A	225	17	0.71	het

Table S4. Effect of 10% contamination of wild-type sequences in the mutant pool. We computationally simulated mis-sorting of embryos by taking adding 10% additional randomly chosen paired end reads from the wild-type *jj410* library into the mutant *jj410* library. Simulation was repeated three times.

Table S5. SNP effects by category and genome. This table summarizes the effects of all homozygous SNPs for each genome library we analyzed. **Download**