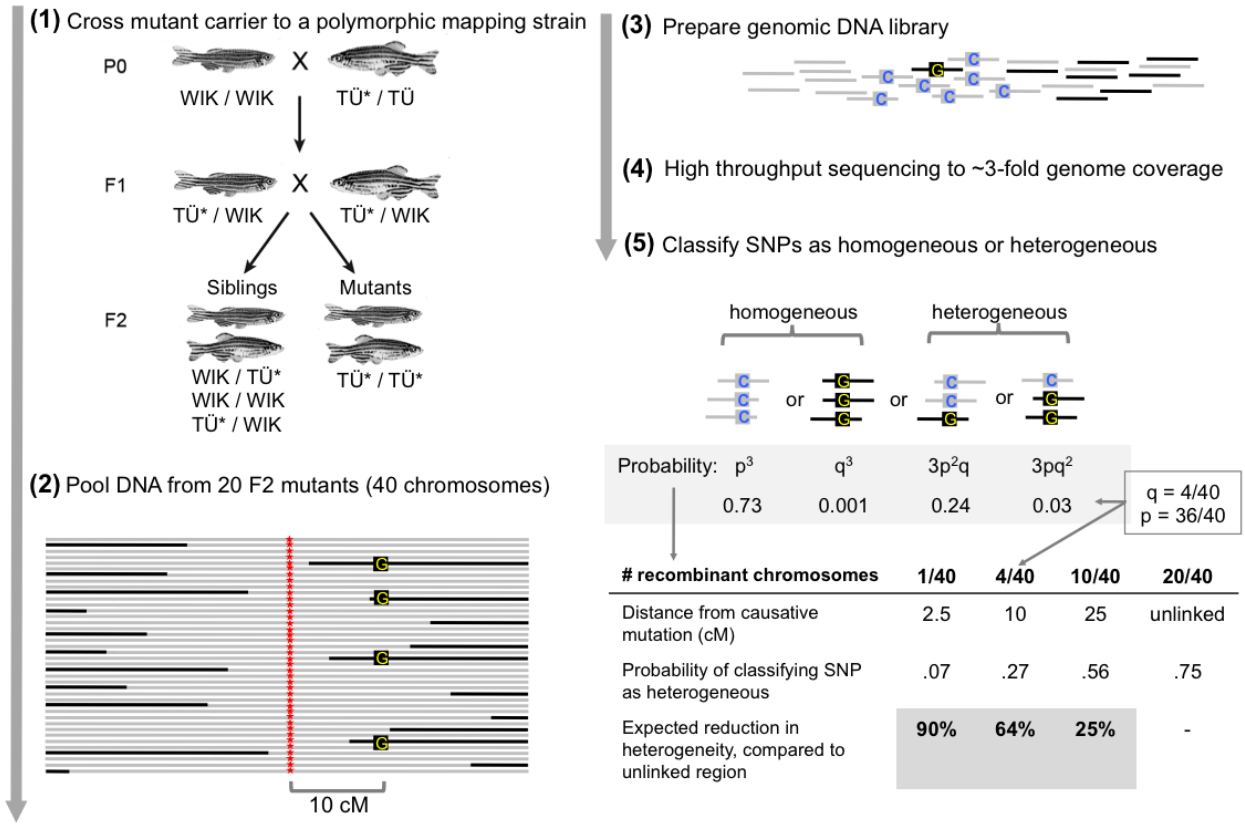

Figure S1 Outline and predicted sensitivity of the approach. (**1**) The generation of fish used in mapping is accomplished by crossing identified mutants carrying a recessive ENU-induced mutation (*) within the Tü background, to a polymorphic mapping strain, (e.g., WIK). Mutant carriers (Tü*/WIK) of the F1 generation are then intercrossed to generate F2 progeny. These F2 fish are sorted based on the presence or absence of the mutant phenotype. (**2**) DNA is prepared from 20 F2 mutant progeny (Tü*/Tü*) and pooled in equal quantities. The diagram depicts the 40 chromosomes containing a phenotype-causing ENU-induced mutation (red asterisk) among the 20 mutant fish. The mutation is linked to genomic sequence originating from the Tü strain used for mutagenesis (grey fragments). Recombinants having sequence originating from the outcross strain (black fragments) can be observed at different distances from the causative mutation as a result of meiotic recombination during meiosis in the F1 generation. In a SNP located ~10 cM from the causative mutation, we expect by definition, 4 of the 40 mutation-containing chromosomes to show a mapping strain allele (G in WIK; black square) as a result of meiotic recombination. (**3**) Physical fractionation of DNA from the 20 mutant fish produces DNA fragments, that contain the aforementioned SNP (boxed C for the Tü and G for the WIK alleles), (**4**) Whole genome sequencing of the fragmented DNA library is performed on a single lane of an Illumina HiSeq platform resulting in ~3x genome coverage. (**5**) Probability for detecting a SNP as being homogenous or heterogeneous in pooled DNA from 20 mutant fish sequenced to 3x coverage. SNPs are classified as homogeneous if all 3 reads covering the SNP represent the same allele (probability = $p^3 + q^3$) and as heterogeneous if both alleles are represented (probability = $3p^2q + 3pq^2$). In an unlinked region, where both alleles are equally represented (q = 0.5; p = 0.5), the probability of a SNP being detected as heterogeneous is 0.75. Likewise, in regions where 10% of the chromosomes are recombinant, as in our example, statistically 4 out of 40 (q = 0.1) reads would show the mapping-strain allele (G), while 36 out of 40 (p = 0.9) would show the reference allele (C; Tü). Thus the probability of detecting the SNP in a heterogeneous state is 0.27. Therefore, the number of heterogeneous SNPs identified in such a region is expected to be ~64% lower than in an unlinked region (0.27/0.75 = 64%). Similarly, a ~90% reduction in heterogeneity is expected for regions containing 1 recombinant chromosome, while a ~25% reduction is expected for regions with 10 recombinant chromosomes. According to this analysis using low genome coverage, it would be of no added benefit to pool larger numbers of fish to increase the resolution of mapping. As the probability of detecting a single recombinant in, for example, 40 fish (80 chromosomes) would be lower than the level of detecting false positive heterogeneous SNPs and thus indistinguishable from noise.

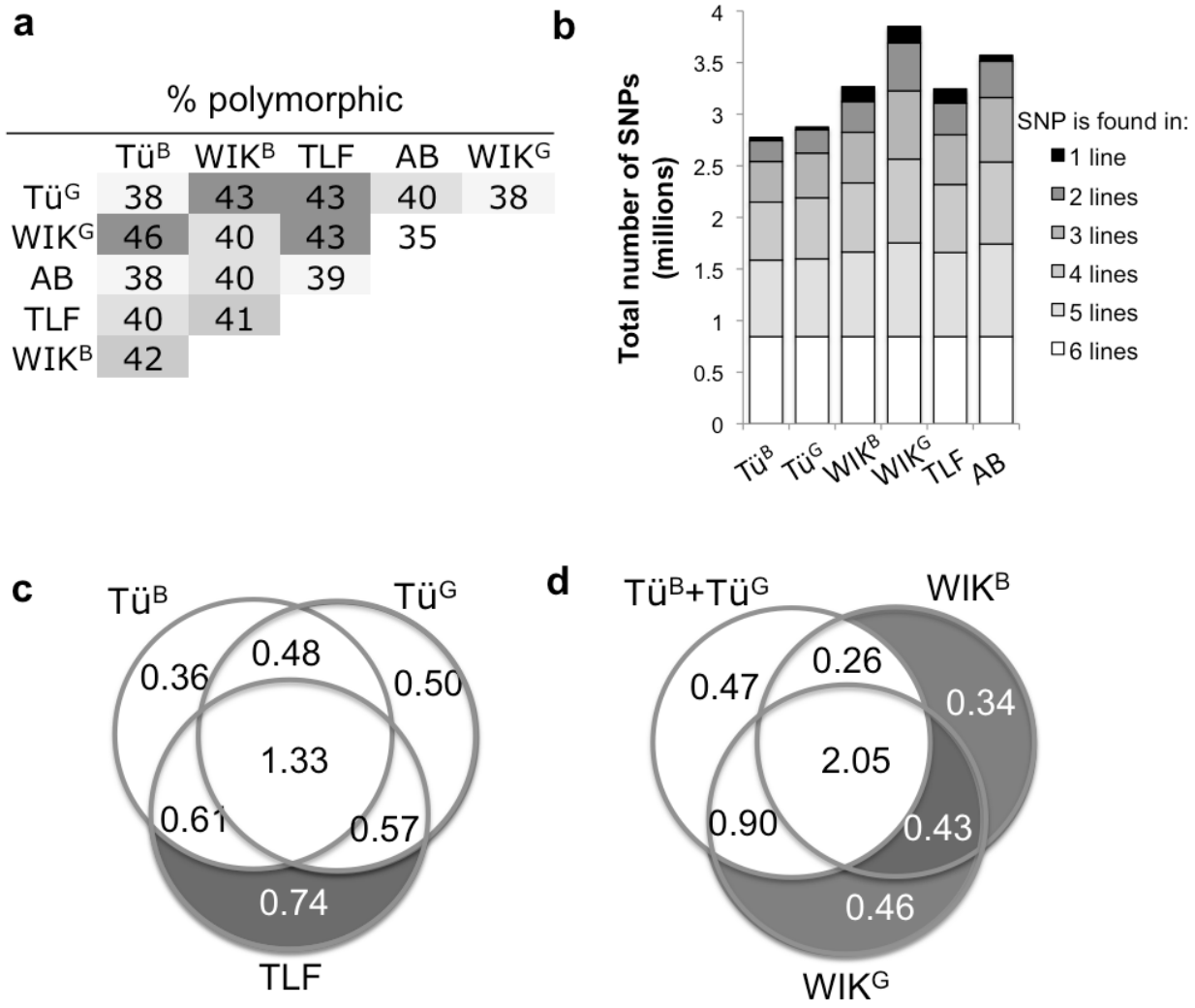

Figure S2 Genetic variation in zebrafish strains detected in low coverage WGS data of pooled DNA from 20 fish. (**a**) Pairwise comparison of SNP genotypes between parental lines showing the percentages of SNPs that are polymorphic and thus could be used to predict the parental origin for mapping based on homozygosity-by-descent. SNPs were classified as polymorphic if only the reference genome allele was observed in one line, while at least one alternate allele was observed in the second line. (**b**) Graph showing the number of polymorphic SNPs (in millions) identified in each parental line, as well as the number of lines with which these SNPs are shared. For (a) and (b), only sites with sequence coverage in all lines were considered (5.2 million sites out of the 7.6 million total SNP sites). (**c, d**) Venn diagrams showing the SNPs that were classified as mapping strain SNPs. This includes 0.74 million SNPs found in TLF but not in either of the Tü lines, and 1.2 million SNPs found in either of the WIK lines but not in either of the Tü lines. For (c) the two Tü lines are shown separately, while in (d) the data from these two lines were combined. Interestingly, due to high levels of intra-strain variation, there is a high number of SNPs that are not shared between the two Tü lines (c), and a similar number of SNPs that are not shared between the two WIK lines (d).

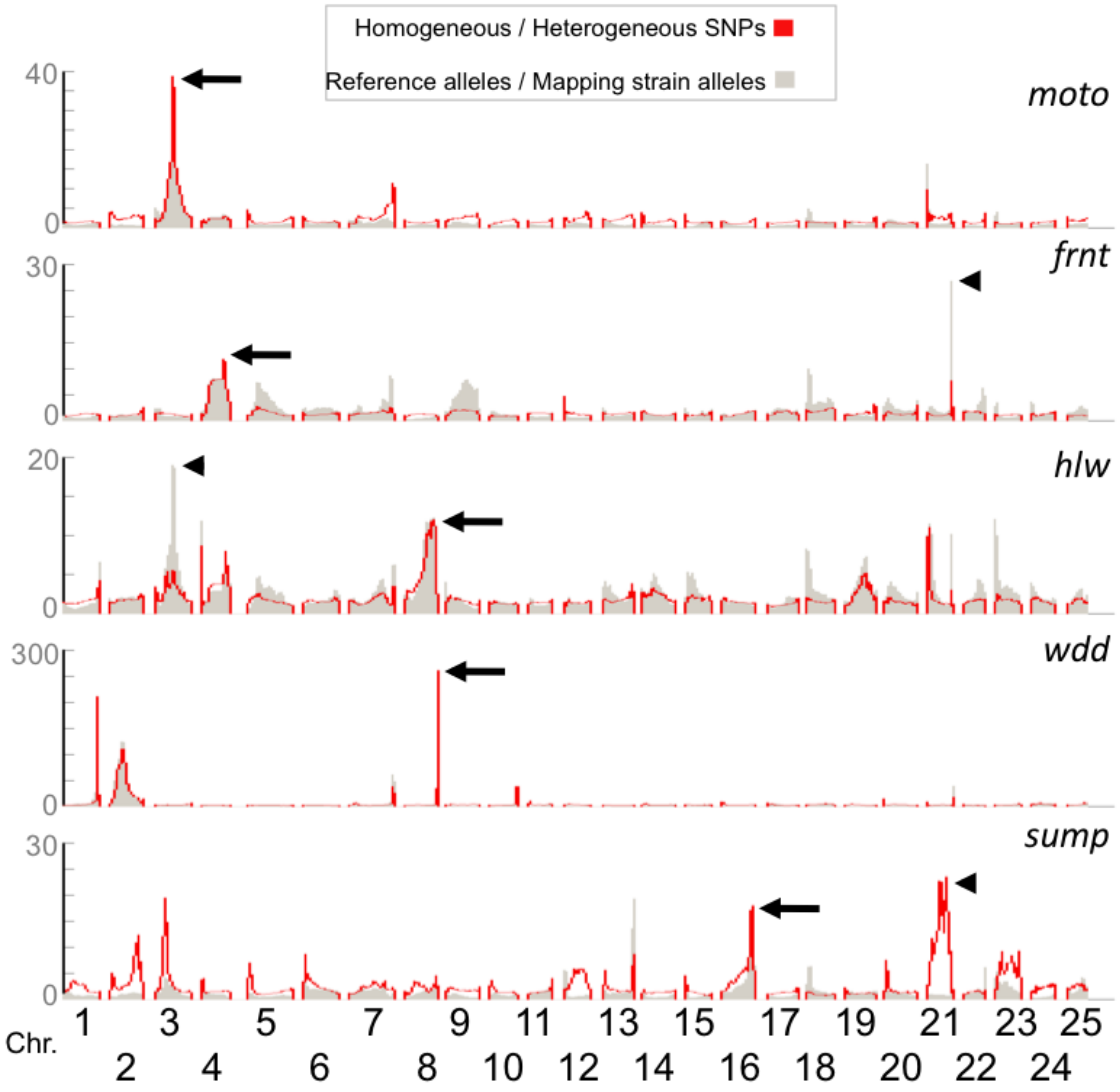

Figure S3 Mapping by combining both the frequency of heterogeneous SNPs and the frequency of mapping strain SNPs helps to eliminate false positives. Graphs show the comparison between the ratio of homogeneous to heterogeneous SNPs (red lines) to the ratio of reference genome alleles to mapping strain alleles (gray bars), for the five mutants analyzed (*moto, frnt, hlw, wdd, sump)*. Ratios were calculated for all 25 chromosomes using sliding windows of 20 cM in size, with an overlap of 19.75 cM between adjacent windows. Genetic distances were defined by the MGH meiotic map. The arrow indicates the linked region for each mutant. For three mutants (*moto, frnt, wdd*), both approaches independently predict the linked region as the region in the genome with the highest ratio. In the *hlw* and *frnt* mutants, other regions show the highest ratio of reference alleles to mapping strain alleles (arrowheads). These regions do not have a high ratio of homogeneous to heterogeneous SNPs. Similarly, for the *sump* mutant, region on Chr21 shows the highest ratio of homogeneous to heterogeneous SNPs, but this region does not have a high ratio of reference alleles to mapping strain alleles (arrowhead). Accordingly, these false positive regions would result in a lower mapping score in our combined analysis and thus would be ranked as less likely to be linked to the mutation.

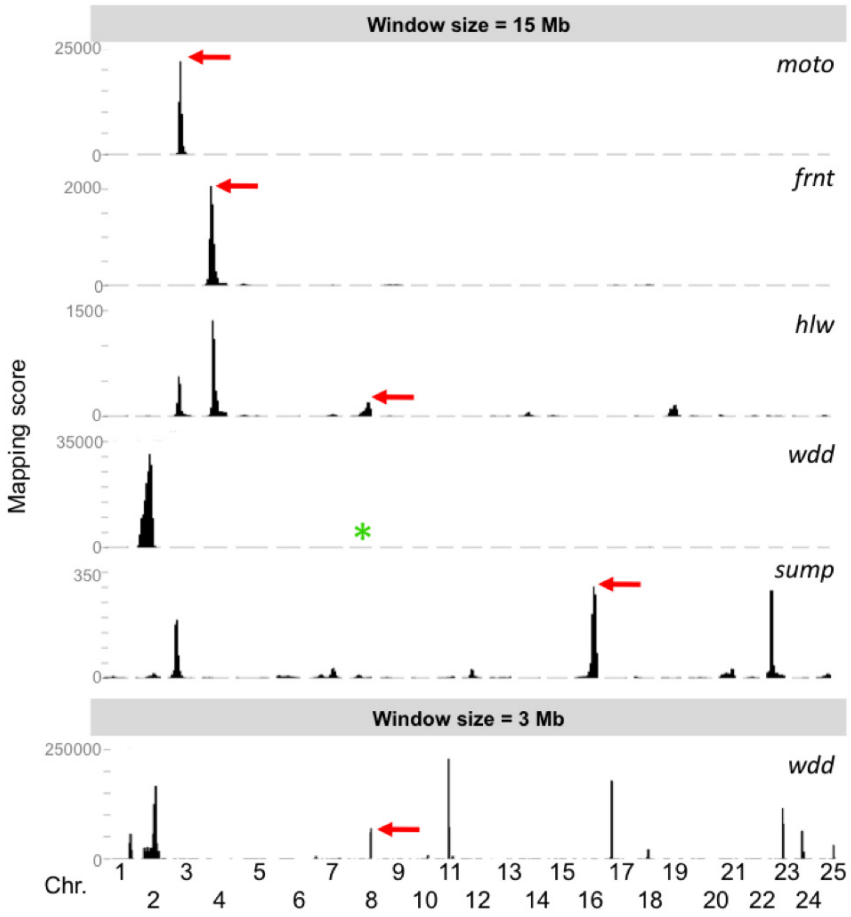

Figure S4 The sensitivity and specificity of mapping is affected by the size of the window used to calculate the mapping score. Graphs showing the genome-wide mapping scores using a sliding window of 15 Mb in size for all mutants (above), or a sliding window of 3 Mb in size for *wdd* (below), rather than the 20 cM windows used in our analysis. When 15 Mb sliding windows are used, in only three of the five mutants (*moto, frnt, sump*) the linked region is contained within the window with the highest mapping score in the genome (red arrows). In *hlw* the linked region is contained within the peak with the third highest mapping score (red arrow). In *wdd,* the linked region is not detected by an increase in the mapping score (asterisk), because the linked interval on Chr8 spans only 4 Mb. When a 3 Mb window was used for *wdd*, which should be small enough to detect the linked region, a mapping score peak appears at the linked interval (red arrow), but it is only the 5th highest peak.

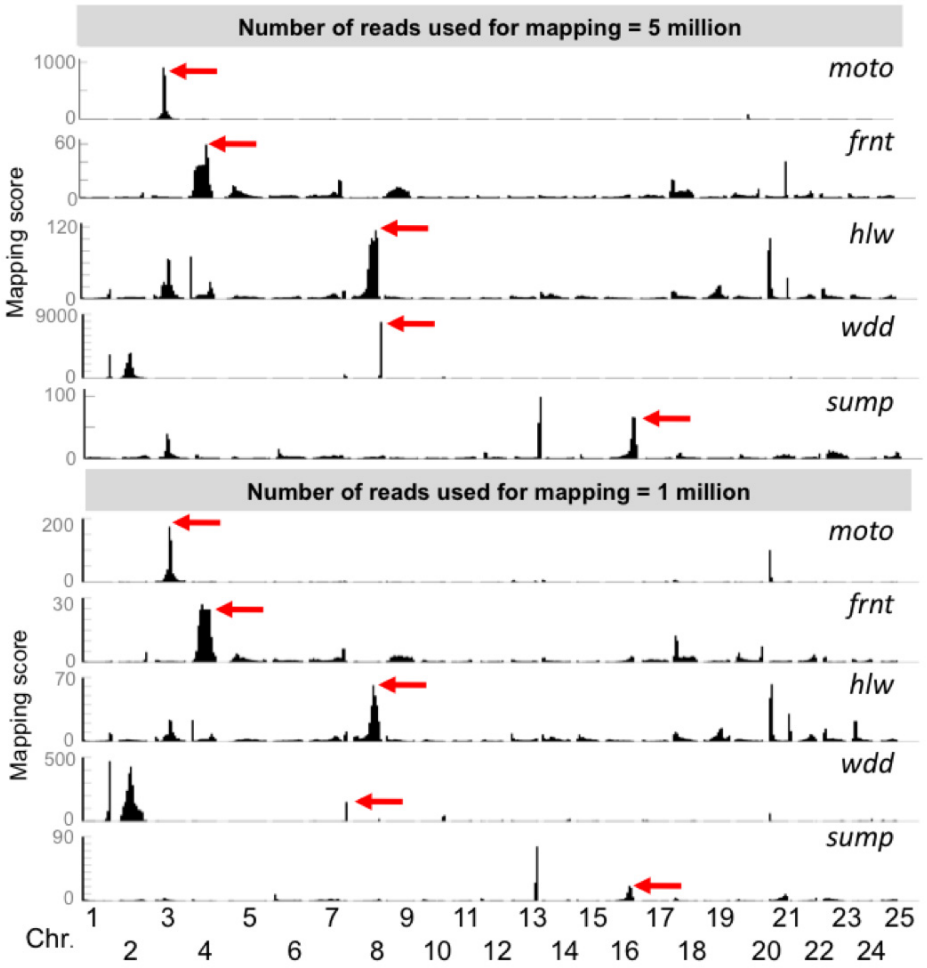

Figure S5 Minimum coverage needed for efficient mapping of zebrafish mutants. Graphs depicting the genome-wide mapping scores calculated for each mutant in 20 cM sliding windows, using either only 5 million (top) or 1 million (bottom) randomly selected Illumina sequencing reads. The actual map positions for each mutant are indicated (red arrows). When 5 million reads are used, the mapping score plots are not significantly different from those generated using all reads (>60 million) (Figure 1). The only exception is that, for *sump*, the linked region has only the 2nd highest mapping score. Even when only 1 million reads are used, in three mutants (*moto, frnt, hlw*) the linked region has the highest mapping score in the genome. For two other mutants (*wdd, sump*), the relative heights of the false positive peaks are significantly increased.

Table S1 Whole genome sequencing of pooled DNA from wild-type zebrafish strains

Wild-type pool	# reads[a] (millions)	Genome coverage	# SNPs[b] (per kb)	% het[c]
Tü^{B}	97	5.1x	2.6	88
Tü^{G}	81	4.1x	2.5	89
WIK^{B}	96	4.1x	2.9	64
WIK^{G}	81	4.0x	3.5	73
AB	90	4.6x	3.3	79
TLF	91	3.8x	2.9	56

[a]Number of 100 bp reads obtained by Illumina single end sequencing. [b]Number of positions at which at least one read representing an alternate allele was observed. Only positions at the 7.6 million SNP sites identified in this study were considered. [d]Percentage of SNPs that were heterogeneous (i.e., both reference-genome and alternate alleles were represented)

Table S2 Classifying SNPs identified by whole genome sequencing of pooled DNA from zebrafish mutants

Mutant pool	SNP genotype[a] (average per kb)				Parental origin of alleles[b] (average per kb)	
	Het	Hom		n/d	Mapping strain allele	Reference genome allele
		Non-ref	Ref			
moto	1.6	0.9	1.8	1.3	0.6	0.6
wdd	1.0	0.4	2.1	2.2	0.3	0.3
hlw	1.7	0.4	2.4	1.0	0.4	0.8
frnt	2.1	0.6	2.1	0.7	0.6	0.7
sump	1.8	1.1	2.0	0.6	0.7	0.6

[a]Calculated for the 7.6 million SNP sites identified in this study. SNPs were defined as heterogeneous (Het) for sites at which both a reference genome allele (Zv9) and an alternate allele were observed in the WGS of pooled DNA. SNPs were defined as homogeneous (Hom) for sites at which only the alternate allele (non-ref) or the reference genome allele (ref) were observed; sites that were covered by less than 2 sequencing reads were deemed uninformative (n/d). [b]Calculated for all SNP sites at which an alternate allele was present in the TLF or WIK mapping strain, but not in the Tü strain (0.7 million and 1.2 million sites respectively); Mapping strain allele = at least one read representing the alternate allele was observed in a mutant pool; Reference genome allele = all reads in a mutant pool represented the reference genome allele. Note that the reference genome is based on the Tü strain.

Table S3 SSLP and SNP marker linkage data

Zebrafish mutant	Chr[a]	Region of homogeneity (Mb)[b]	Marker[c]	Position[d] (Mb)	Recombinants/meiosis[e]
moto	3	19-36.6	z9964	34.3	2/90
wdd	8	51.8-55	*bmp1a*	53.5	0/66
hlw	8	43-47	z25210	43	2/44
			z7130	45.6	1/46
frnt	4	15-27.8	z11538	23.1	3/44
			z20450	24.4	2/44
sump	16	41.8-51.2	z8819	41.3	2/56
			z15739	43.9	0/56
			z4670	50.8	1/40
			z6854	50.6	1/40

[a]Chromosome showing the highest mapping score. [b]Interval on the chromosome showing homogeneity. [c]Marker used to confirm linkage. [d]Marker position on the Chromosome in Mb. [e]Recombinants identified for each marker in the number of meiosis tested.