Title:

The genomic and molecular basis of response to selection for longer limbs in mice

Short Title:

Genomics of selection response in Longshanks mice

One-sentence summary:

Sequencing of "Longshanks mice" reveals evolutionary principles and *cis*-acting changes in the rapid evolution of a morphological trait

Authors:

João P. L. Castro [1,*], Michelle N. Yancoskie [1,*], Marta Marchini [2], Stefanie Belohlavy [3], Marek Kučka [1], William H. Beluch [1], Ronald Naumann [4], Isabella Skuplik [2], John Cobb [2], Nick H. Barton [3], Campbell Rolian[2,†], Yingguang Frank Chan [1,†]

Affiliations:

1. Friedrich Miescher Laboratory of the Max Planck Society, Tübingen, Germany
2. University of Calgary, Calgary AB, Canada
3. IST Austria, Klosterneuburg, Austria
4. Max Planck Institute for Cell Biology and Genetics, Dresden, Germany

Corresponding author:

Campbell Rolian

Yingguang Frank Chan

* indicates equal contribution

† indicates equal contribution

Abstract:

A major goal in evolutionary biology is to understand how genomes evolve in response to selection. Here we present a genomic dissection of the Longshanks selection experiment, in which mice were selectively bred over 20 generations for longer tibiae relative to body mass,

resulting in 13% longer tibiae. Using population genetic analysis of sequence data, we identified eight significant loci among a genome-wide polygenic response. Although six of these loci were specific to one of two Longshanks replicates, the two loci with the strongest response were selected in parallel, with the single strongest locus *Nkx3-2* accounting for ~9.4% of the overall selection response. Transgenic reporter assays show that both gain- and loss-of-function enhancer variants contributed to the selection response. By integrating multiple lines of evidence, we dissected the selected locus *Nkx3-2* down to individual base-pairs and confirm the critical role of *cis*-regulatory changes in the rapid evolution of a morphological trait.

Main Text:

A major goal in evolutionary biology is to understand how genomes evolve in response to selection. Given similar starting conditions, are changes in the same genetic loci or pathways involved repeatedly to achieve similar phenotypic outcomes (parallelism)? Does rapid selection response of a complex trait proceed through thousands of loci of infinitesimally small effect, or through a few loci of large effect? And if so, what signature of selection should we expect? Does morphological evolution proceed predominantly via regulatory or protein-coding changes (*1*)? A large body of theory exists to describe the birth, rise and eventual fixation of adaptive variants under diverse selection scenarios (*2-7*). However, few empirical datasets capture sufficient detail on the founding conditions and selection regime to allow full reconstruction of the selection response. Selection experiments, which reproduce rapid evolution under controlled conditions, are excellent tools to understand response to selection—and by extension—adaptive evolution in nature.

Here we describe a multi-faceted investigation into an artificial selection experiment, called Longshanks, in which mice were selected for increased tibia length relative to body mass (*8*). The mammalian hind limb is an ideal model to study the dynamics of complex traits under selection. It is both morphologically complex and functionally diverse, reflecting its adaptive

value and importance to the evolutionary success of mammals (*9*). Moreover, limb development has been studied extensively across vertebrates (e.g., mice, birds and fish) as a genetic and evolutionary paradigm (*9*). These studies have simultaneously highlighted deep evolutionary conservation and broad adaptive potential. The Longshanks selection experiment thus offers an unparalleled opportunity to study selection response not only from a quantitative and population genetics perspective, but also from a developmental (*10*) and genomic perspective.

By design, the Longshanks experiment preserves a nearly complete archive of the phenotype (trait measurements) and genotype (via tissue samples) in the pedigree. By sequencing the initial and final genomes, we can study this example of rapid evolution with unprecedented detail and resolution. Using the exact trait measurements, pedigree and genotypes, we can model the predicted selection responses and test the outcome against empirical results. This precision and genetic resolution is important, because although quantitative genetic methods such as genomic selection or the breeder's equation have been tremendously successful in predicting trait values from genotypes or the phenotypic selection response (*11, 12*), they ignore key details such as haplotype structure or molecular pathways and make no attempt to understand the genetic mechanism that underlie the selection response. With the Longshanks experiment, we aim to achieve sufficient statistical and molecular resolution to address major questions such as the genetic architecture of the response to selection on a complex morphological trait in a mammal and its associated selection signature, the prevalence of genetic parallelism, and the contribution of *cis*-acting vs. coding variants to morphological evolution.

Longshanks selection for longer tibiae

At the start of the Longshanks experiment (LS), we established three base populations with 14 pairs each by sampling from a genetically diverse, commercial mouse stock [Hsd:ICR,

76 also known as CD-1; derived from mixed breeding of classical laboratory mice (*13*). In two
77 replicate "Longshanks" lines (LS1 and LS2), we selectively bred mice with the longest tibia
78 relative to the cube root of body mass [within-family indexed selection with 15–20% of
79 offspring selected for breeding; see (*8*) for details]. We kept a third Control line (Ctrl) using an
80 identical breeding scheme, except that breeders were selected at random. In LS1 and LS2, we
81 observed a strong and significant response in tibia length to selection [0.29 and 0.26 Haldane
82 or standard deviations (s.d.) per generation from a selection differential of 0.73 s.d. in LS1 and
83 0.62 s.d. in LS2]. Over 20 generations, selection for longer relative tibia length produced
84 increases of 5.27 and 4.81 s.d. in LS1 and LS2, respectively (or 12.7% and 13.1% in tibia
85 length), with negligible changes in body mass (Fig. 1B & C; fig. S1). By contrast, Ctrl showed
86 no directional change in tibia length or body mass (Fig. 1C; Student's *t*-test, $P > 0.05$). This
87 approximately 5 s.d. change in 20 generations is rapid compared to typical rates observed in
88 nature [(*14*) but see (*15*)] but is in line with responses seen in selection experiments (*12, 16-18*).
89

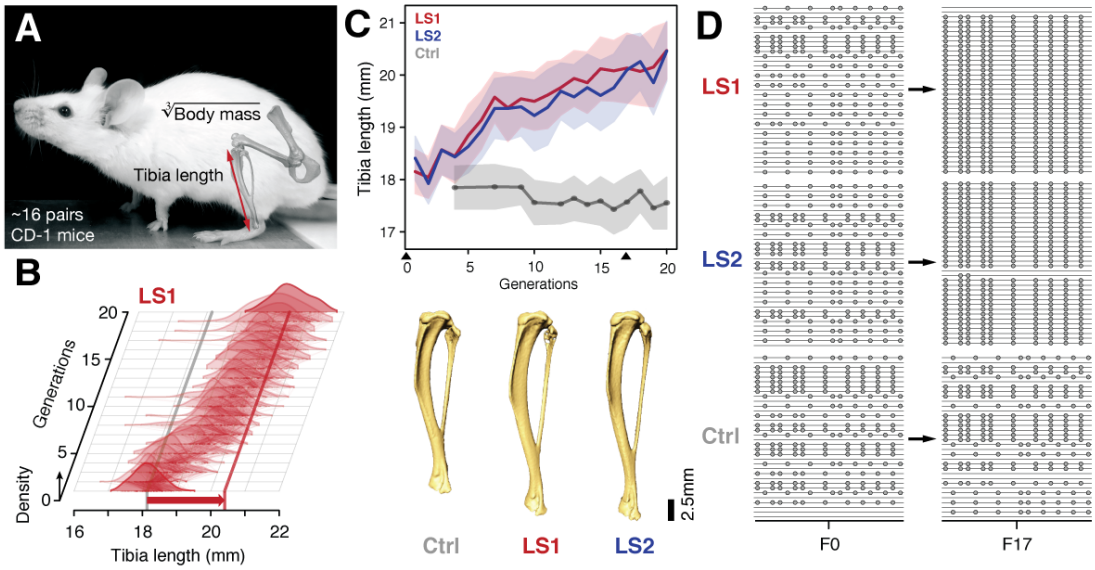


Fig. 1. Selection for Longshanks mice produced rapid increase in tibia length.
(**A** and **B**) Tibia length varies as a quantitative trait among outbred mice derived from the Hsd:ICR (also known as CD-1) commercial stock. Selective breeding for mice with the longest tibiae relative to body mass within families has produced a strong selection response in tibia length over 20 generations in Longshanks mice (13%, red arrow, LS1). (**C**) Both LS1 and LS2 produced replicated rapid increase in tibia length (red and blue; line and shading show mean ± s.d.) compared to random-bred Controls (grey). Arrowheads mark sequenced generations F0 and F17. See fig. S1 for body mass data. Lower panel: Representative tibiae from the Ctrl, LS1 and LS2 after 20 generations of selection. (**D**) Analysis of sequence diversity data (linked variants or haplotypes: lines; variants: dots) may detect signatures of selection, such as selective sweeps (F17 in LS1 and LS2) that result from selection favoring a particular variant (dots), compared to neutral or background patterns (Ctrl).

Simulating selection response: infinitesimal model with linkage

We next developed a simulation that faithfully recapitulates the artificial selection experiment by integrating the trait measurements, selection regime, pedigree and genetic diversity of the Longshanks selection experiment, in order to generate an accurate expectation for the genomic response. Here we cannot use a neutral null model, since we know that strong selection was applied. Using the actual pedigree and trait measurements, we mapped fitness onto the selected traits tibia length $\boldsymbol{T}$ and body mass $\boldsymbol{B}$ as a single composite trait $ln(\boldsymbol{T}\boldsymbol{B}^{\theta})$. We estimated θ from actual data as –0.57, such that the ranking of breeders best matched the actual $\boldsymbol{T}$ and $\sqrt[3]{\boldsymbol{B}}$ composite ranking used in the selection experiment (*8*). We assumed the maximally polygenic genetic architecture using an "infinitesimal model with linkage" (abbreviated here as H_{INF}), under which the trait is controlled by very many loci, each of infinitesimally small effect (see Supplementary Notes for details). Results from simulations seeded with actual genotypes or haplotypes showed that overall, the predicted increase in inbreeding closely matched the observed data (fig. S2A). We tested models with varying selection intensity and linkage disequilibrium (LD), and for each, ran 100 simulated replicates to determine the significance of changes in allele frequency (fig. S2B–D). This flexible quantitative genetics framework allows us to explore possible changes in genetic diversity over 17 generations of breeding, under strong selection.

In simulations, we followed blocks of genome as they were passed down the pedigree. In order to compare with observations, we seeded the initial genomes with single nucleotide polymorphisms (SNPs) in the same number and initial frequencies as the data. We observed much more variation between chromosomes in overall inbreeding (fig. S2A) and in the distribution of allele frequencies (fig. S3B) than expected from simulations in which the ancestral SNPs were initially in linkage equilibrium. This can be explained by linkage disequilibrium (LD) between the ancestral SNPs, which greatly increases random variation.

Therefore, we based our significance tests on simulations that were seeded with SNPs drawn with LD consistent with the initial haplotypes (fig. S2B & D; see Supplementary Notes).

Because our simulations assume infinitesimal effects of loci, allele frequency shifts exceeding this stringent threshold serve as evidence that discrete loci contribute significantly to the selection response. An excess of such loci in either a single LS replicate or in parallel would thus imply a mixed genetic architecture of a few large effect loci amid an infinitesimal background.

Sequencing the Longshanks mice reveals genomic signatures of selection

To detect the genomic changes in the actual Longshanks experiment, we sequenced all individuals of the founder (F0) and 17th generation (F17) to an average of 2.91-fold coverage (range: 0.73–20.6×; n = 169 with <10% missing F0 individuals; Table S1). Across the three lines, we found similar levels of diversity, with an average of 6.7 million (M) segregating SNPs (approximately 0.025%, or 1 SNP per 4 kbp; Table S1; figs S3A & S4). We checked the founder populations to confirm negligible divergence between the three founder populations (across-line F_{ST} on the order of 1×10^{-4}), which increased to 0.18 at F17 (Table S2). This is consistent with random sampling from an outbred breeding stock. By F17, the number of segregating SNPs dropped to around 5.8 M (Table S1). This 13% drop in diversity (0.9M SNPs genome-wide) closely matched our simulation results, from which we could determine that the drop was mostly due to inbreeding, with only a minor contribution from selection (Supplementary Notes, fig. S2A, C).

We conclude that despite the strong selection on the LS lines, there was little perturbation to genome-wide diversity. Indeed, the changes in diversity during the 17 generations were remarkably similar in all three lines, despite Ctrl not having experienced selection on relative tibia length (fig. S3A). Hence, and consistent with our simulation results

(fig. S2A, C), changes in global genome diversity had little power to distinguish selection from neutral drift despite the strong *phenotypic* selection response.

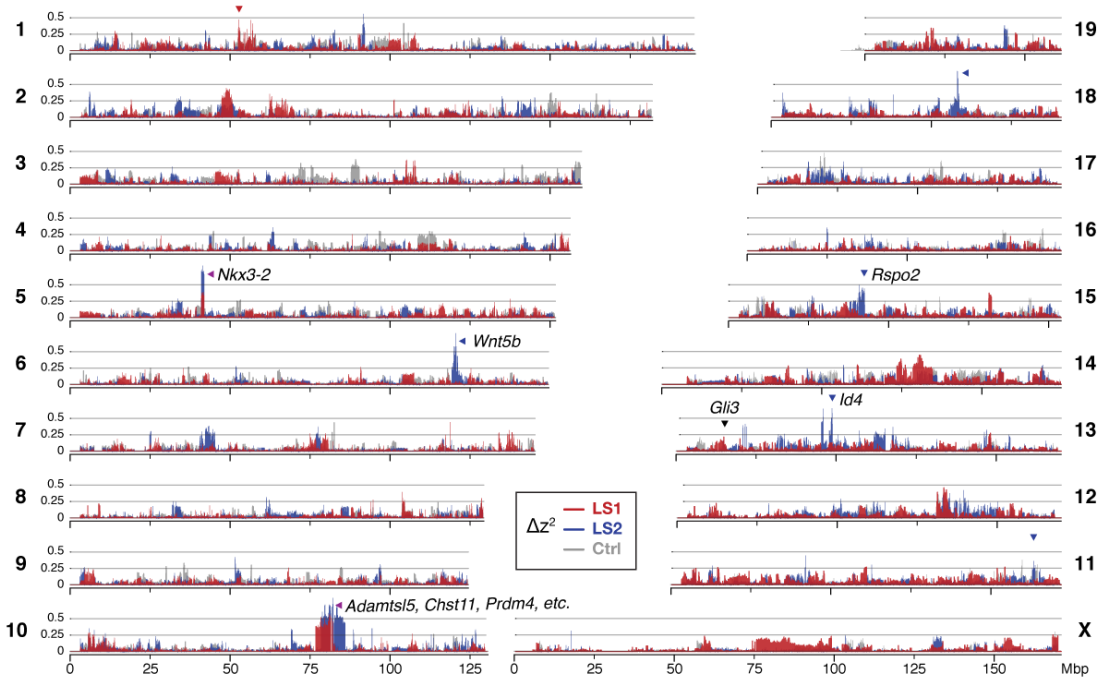


Fig. 2. Widespread genomic response to selection for increased tibia length. Allele frequency shifts between generations F0 and F17 in LS1, LS2 and Ctrl lines are shown as Δz^2 profile across the genome (plotted here as fraction of its range from 0 to π^2). The Ctrl Δz^2 profile (grey) confirmed our expectation from theory and simulation that drift, inbreeding and genetic linkage could combine to generate large Δz^2 shifts even without selection. Nonetheless the LS1 (red) and LS2 (blue) profiles show a greater number of strong and parallel shifts than Ctrl. These selective sweeps provide support for the contribution of discrete loci to selection response beyond a polygenic background (arrowheads, red: LS1; blue: LS2; purple: parallel; see also fig. S4, S5). Candidate genes are highlighted (Table 1).

We next asked whether specific loci may reveal more definitive differences between the LS replicates and Ctrl (and from infinitesimal predictions). We calculated Δz^2, the square of arc-sin transformed allele frequency difference between F0 and F17; this has an expected variance of $1/2N_e$ per generation, independent of starting frequency, and ranges from 0 to π^2. We averaged Δz^2 within 10 kbp windows (see Methods for details), and found 169 windows belonging to eight clusters that had significant shifts in allele frequency in LS1 and/or 2 (corresponding to 9 and 164 clustered windows respectively at $P \leq 0.05$ under $H_{INF,\ max\ LD}$; Fig. 2; figs. S2D, S5, S6; see Methods for details) and in 3 loci in Ctrl (8 windows). The eight loci overlapped between 2 to 179 genes and together contain 11 candidate genes with known roles

in bone, cartilage and/or limb development (e.g., *Nkx3-2* and *Sox9*; Table 1). Four out of the eight loci contain genes with a "short tibia" or "short limb" knockout phenotype (Table 1; $P \leq 0.032$ from 1000 permutations, see Methods for details). Of the broader set of genes at these loci with any limb knockout phenotypes, only *fibrillin 2* (*Fbn2*) is polymorphic for mutations coding for different amino acids, suggesting that for the majority of loci with large shifts in allele frequency, gene regulatory mechanism(s) were likely important in the selection response (fig. S7; Table S3; see Supplementary Notes for further analyses on enrichment in gene functions, protein-coding vs. *cis*-acting changes and clustering with loci affecting human height).

Taken together, two major observations stand out from our genomic survey. One, a polygenic, infinitesimal selection model with strong LD amongst marker SNPs best fits the observed data; and two, we nevertheless find more discrete loci in LS1 and LS2 than in Ctrl beyond the significance threshold set by the infinitesimal model (Fig. 2; fig. S5). Thus we conclude that the genetic basis of the selection response in the Longshanks experiment was mostly widespread and polygenic, but that there was also a significant contribution from discrete loci with major effect.

Rk	Chr	Span (Mbp)	Peak	Core (kbp)	TAD (kbp)	Genes	$\Delta z^2/\pi^2$ LS1	$\Delta z^2/\pi^2$ LS2	*s*	Δq LS1	Δq LS2	Δq Ctrl	Type	Candidate genes*
1	**5**	**38.95–45.13**	**41.77**	**900**	**720**	***3***	**0.22**	**0.45**	**0.26**	**0.69**	**0.86**	**-0.14**	**Parallel**	***Nkx3-2***
2	**10**	**77.47–87.69**	**81.07**	**5,360**	**6,520**	***175***	**0.32**	**0.42**	**0.21**	**0.79**	**0.88**	**-0.04**	**Parallel**	***Sbno5, Aes, Adamtsl5*, Chst11*, Cry1, Prdm4****
3	18	53.63–63.50	58.18	220	520	*4*	0.03	0.36	0.18	0.05	**0.78**	-0.06	LS2-specific	-
4	13	35.59–55.21	48.65	70	2,600	*22*	0.06	0.26	0.18	0.24	**0.80**	-0.03	LS2-specific	*Id4*
5	1	53.16–57.13	55.27	10	720	*4*	0.36	0.12	0.14	-0.14	**0.66**	-0.15	LS1-specific	-
6	15	31.92–44.43	41.54	10	680	*3*	0.08	0.27	0.14	-0.23	**0.66**	0.02	LS2-specific	*Rspo2**
7	6	118.65–125.25	120.30	130	1,360	*12*	0.07	0.34	0.13	-0.03	**0.79**	-0.15	LS2-specific	*Wnt5b**
8	11	111.10–115.06	113.42	10	2,120	*2*	0.02	0.36	0.12	**0.65**	0.01	-0.23	LS2-specific	*Sox9**

Rk Rank

Chr Chromosome

Core Span of 10 kbp windows above $H_{INF,\ max\ LD}$ $P \leq 0.05$ significance threshold

TAD Merged span of topologically associating domains (TAD) overlapping the core span. TADs mark segments along a chromosome that share a common regulatory mechanism. Data from (*25*).

s Selection coefficient (see Supplementary Notes and Methods; likely over-estimated due to the loci being outliers)

***** Genes within the TAD span showing "short tibia", "short limbs", "abnormal osteoblast morphology" or "abnormal cartilage morphology" knockout phenotypes are listed, with * marking those with "short tibia".

Table 1. Major loci likely contributing to the selection response. These 8 loci show significant allele frequency shifts in Δz^2 and are ordered according to their estimated selection coefficients. Shown for each locus are the full hitchhiking spans, peak location and their size covering the core windows, the overlapping TAD and the number of genes found in it. The two top-ranked loci show shifts in parallel in both LS 1 & 2, with the remaining six showing line-specific response (LS1: 1; LS2: 5). Candidate genes found within the TAD with limb, cartilage or bone developmental knockout phenotypes functions are shown, with asterisks (*) marking those with a "short tibia" knockout phenotype (see also fig. S6 and Table S3 for full table).

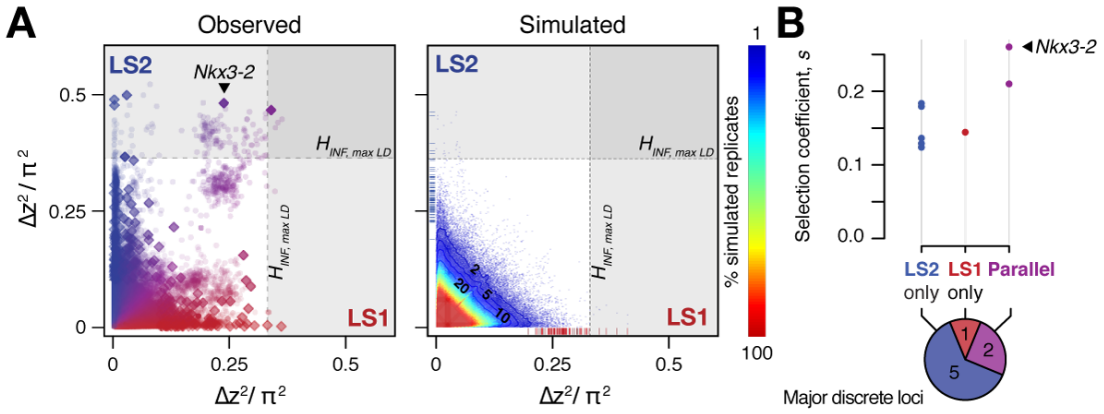


Fig. 3. Selection response in the Longshanks lines was largely line-specific, but the strongest signals occurred in parallel.
(**A**) Genetic differentiation at all sites in the genome showed largely independent changes in LS1 (red) vs. LS2 (blue; left panel; diamonds: peak windows; dots: other 10 kbp windows; see fig. S8 and Supplementary Methods for details). The overall distribution closely matches simulated results under the infinitesimal model with maximal linkage disequilibrium ($H_{INF, max LD}$; right heatmap summarizes the percentage seen in 100 simulated replicates), with most of the windows showing little to no shift (red hues near 0; see also fig. S8 for an example replicate). Tick marks along axes showed genome-wide maximum Δz^2 shifts in each replicate in LS1 (red) and LS2 (blue), from which we derived line-specific thresholds at the $P \leq 0.05$ significance level. In contrast to the simulated results, we observed that many loci showed strong parallel shifts in both LS1 and LS2 (left panel, points along the diagonal; adjacent windows appear as clusters due to hitchhiking). (**B**) Estimated selection coefficients *s* among the candidate discrete loci at the $P \leq 0.05$ under $H_{INF, max LD}$. While six out of eight total loci showed significant shift in only LS1 or LS2, the two top-ranked loci in terms of selection coefficients were likely selected in parallel in both LS1 and LS2 (also see left panel in A).

We next tested the repeatability of the selection response at the gene/locus level using the two LS replicates. If the founding populations shared the same selectively favored variants, we may observe parallelism or co-incident selective sweeps, as long as selection could overcome random drift. Indeed, the Δz^2 profiles of LS 1 & 2 were more similar to each other than to Ctrl (Fig. 2 & 3A; fig. S8; Pearson's correlation in Δz^2 from 10 kbp windows: LS1–LS2: 0.21, vs. LS1–Ctrl: 0.063 and LS2–Ctrl: 0.045). Whereas previous genomic studies with multiple natural or artificial selection replicates focused mainly on detecting parallel loci (*19-23*), here we have the possibility to quantify parallelism and determine the selection value of a given locus. Six out of eight significant loci at the $H_{INF, max LD}$ threshold were line-specific, even though the selected alleles were always present in the F0 generation in both lines. This prevalence of line-specific loci was consistent, even if we used different thresholds. However, the two remaining loci that ranked first and second by selection coefficient were parallel, both with *s* >

0.3. (Fig. 3B; note that as outliers, the selection coefficient may be substantially overestimated, but their rank order should remain the same), supporting the idea that the probability of parallelism can be high among those loci with the greatest selection advantage (*24*). In contrast to changes in global diversity over 17 generations, where we could only detect a slight difference between the LS lines and Ctrl, we found the signature of parallelism to be significantly different between the selected LS1 and LS2 replicates, as opposed to comparisons with Ctrl, or between simulated replicates (fig. S8; χ^2 test, LS1–LS2: $P \leq 1 \times 10^{-10}$; Ctrl–LS1 and Ctrl–LS2, $P > 0.01$ and $P > 0.2$, respectively, both non-significant after correcting for multiple testing; see Supplementary Notes for details). As such, the parallel selected loci between LS1 and LS2 may provide the strongest evidence for the role of discrete major loci, and represent prime candidates for molecular dissection.

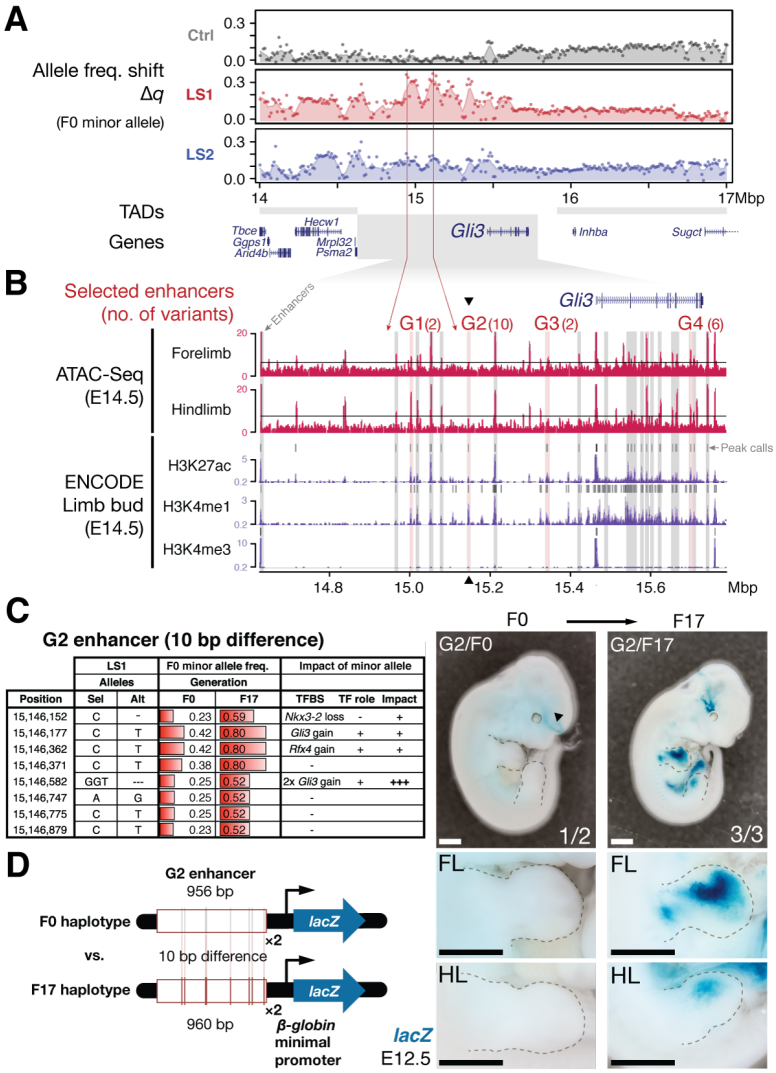


Fig. 4. A novel enhancer variant in LS1 boosts *Gli3* expression during limb bud development. (**A**) LS1 showed elevated Δ*q* (peak windows: arrows) in the intergenic region in a topologically associated domain (TAD, grey shaded boxes) containing *Gli3*. (**B**) Putative limb enhancers (grey bars) were identified through peaks from ATAC-Seq (top; ref. *26*) and histone modifications (bottom tracks, data from ENCODE project) in E14.5 limb buds. Four of the enhancers contain sequence variants (in parentheses) with large allele frequency shifts between F0 and F17 in LS1 (red shading, titled G1–4). One of the enhancers close to the peak Δ*q* signal (G2, arrowhead) containing 10 bp differences was chosen for *lacZ* transgenic reporter assay. (**C**) In G2, individual variants showed an average increase of 0.33 in allele frequency, with 4 sites affecting predicted transcription factor binding in the *Gli3* pathway (including 3 additional copies of *Gli3* binding site), all of which are predicted to boost the G2 enhancer activity. (**D**) The F17 G2 enhancer variant drove robust and consistent *lacZ* reporter gene expression at E12.5, recapitulating *Gli3* expression in the developing fore- and hindlimb buds (FL and HL; right; fig. S9A). Fractions indicate number of embryos showing similar *lacZ* staining out of all transgenic embryos. Substitution of 10 positions (F0 haplotype) led to little observable expression in the limb buds (left). These G2 enhancer gain-of-function mutations (contrasting the major allele between F0 and F17) may confer an advantage under selection for increased tibia length. Scale bar: 1 mm for both magnifications.

We next performed an in-depth molecular dissection of specific loci, in order to investigate the particular mechanisms by which selected SNPs may contribute to the selection response. We first focused on a candidate gene *Gli3*, an important early limb developmental regulator on chromosome 13 (Chr13; Fig. 4A). This locus showed a substantial shift in minor allele frequency of up to 0.42 in LS1 (Δq, 98th quantile genome-wide, but below the $H_{INF,\ max\ LD}$ threshold to qualify as a discrete major locus). We performed functional validation of *Gli3*, given its limb function (*27*) and considering that *Gli3* could be among the many minor loci in the polygenic background contributing to the selection response in LS1. For the second functional validation, we chose amongst the two major, parallel loci. We chose the locus on chromosome 5 (Chr5) at 41–42 Mbp because it showed the strongest estimated selection coefficient and both Longshanks lines shared a clear and narrow selection signature. Crucially, it contains only three genes, including *Nkx3-2*, a strong candidate with known roles in limb development (also known as *Bapx1*; Fig. 2 & 5A)(*28*).

Gain-of-function cis*-acting changes at the limb activator* Gli3

At the *Gli3* locus we could only find conservative amino acid changes (D1090E and I1326V) that are unlikely to impact protein function. Because the signal in LS1 was stronger in the 5' flanking intergenic region, we examined the *Gli3 cis*-regulatory topologically associating domain (TADs, which mark chromosome segments with shared gene regulatory logic) (*25*) and identified putative enhancers using chromatin modification marks from the ENCODE project and our own ATAC-Seq data (Fig. 4B)(*26*, *29*). Four putative enhancers carried SNPs with large allele frequency changes. Among them, an upstream putative enhancer G2 (956 bp) carried 6 SNPs along with two 1- and 3-bp insertion/deletion ("indel") with putative functional impact due to predicted gain or loss in transcription factor binding sites (Fig. 4C). We tested the G2 putative enhancer in a transgenic reporter assay by placing its sequence as a tandem duplicate upstream of a *lacZ* reporter gene (see Methods for details). Here, we were not only interested

in whether the sequence encoded for enhancer activity, but specifically whether the SNPs would affect the activity. We found that only the F17 LS1 allele was able to drive consistent *lacZ* expression in the developing limb buds (Fig. 4D). Importantly, this enhancer was active not only in the shaft of the limb bud but also in the anterior hand/foot plate, a major domain of *Gli3* expression and function (fig. S9A). Furthermore, substitution of the enhancer sequence with the F0 allele (10 differences out of 956 or 960 bp) abolished *lacZ* expression (Fig. 4D). This showed that 10 or fewer changes within this novel enhancer sequence were sufficient to convert the inactive F0 allele into an active limb enhancer corresponding to the selected F17 allele ("gain-of-function"), suggesting that a standing genetic variant of the F17 allele may have been selectively favored because it drove stronger expression of *Gli3*, a gene essential for tibia development (*30*) [but see (*31*)].

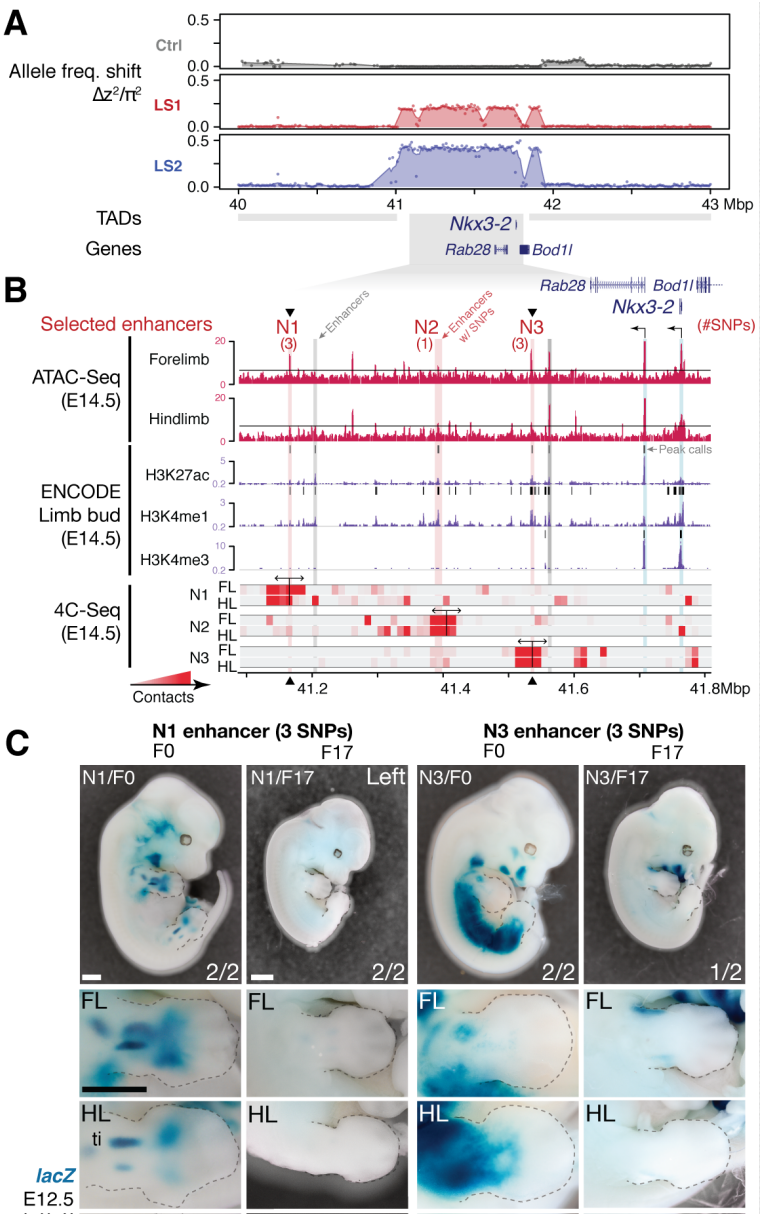


Fig. 5. Strong parallel selection response at the bone maturation repressor *Nkx3-2* locus was associated with decreased activity of two enhancers.
(**A**) Δz^2 in this region showed strong parallel differentiation spanning 1 Mbp in both Longshanks but not in the Control line, which contains three genes *Nkx3-2*, *Rab28* and *Bod1l* (whose promoter lies outside the TAD boundary, shown as grey boxes). In LS2, an originally rare allele had almost swept to fixation (fig. S10A & B). (**B**) Chromatin profiles (ATAC-Seq, red, ref. *26*; ENCODE histone modifications, purple) from E14.5 developing limb buds revealed five putative limb enhancers (grey and red shading) in the TAD, three of which contained SNPs showing significant frequency shifts. Chromosome conformation capture assays (4C-Seq) from E14.5 limb buds from the N1, N2 and N3 enhancer viewpoints (bi-directional arrows) showed significant long-range looping between the enhancers and sequences around the *Nkx3-2* promoter (heat-map with red showing increased contacts; Promoters are shown with black arrows and blue vertical shading). (**C**) Transient transgenic reporter assays of the N1 and N3 enhancers showed that the F0 alleles drove robust and consistent expression at centers of future cartilage condensation (N1) and broader domains of *Nkx3-2* expression (N3) in E12.5 fore- and hindlimb buds (FL, HL; ti: tibia). Fractions indicate number of embryos showing similar *lacZ* staining out of all transgenic embryos. Substituting the F17 allele (replacing 3 positions each in N1 and N3) led to little observable limb bud

expression in both the N1/F17 and the N3/F17 enhancers, suggesting that selection response for longer tibia involved de-repression of bone maturation through a loss-of-function regulatory allele of *Nkx3-2* at this locus. Scale bar: 1 mm for both magnifications.

Loss-of-function cis*-acting changes at the bone repressor* Nkx3-2

At the parallel selected Chr5 locus, the pattern of variation resembles a selective sweep spanning 1 Mbp (Fig. 5A). Close examination of the haplotypes suggests that a rare allele in the region was swept to high frequency in LS1 & 2 but remained unchanged in Ctrl (fig. S10). Comparison between F0 and F17 individuals revealed no recombinant in this entire region (fig. S10A, top panel), precluding fine-mapping using recombinants. Thus, we identified the underlying functional elements using the same strategy as before. First, we determined that no coding changes existed for either of the two genes located within the TAD, *Rab28* or *Nkx3-2*. We then performed *in situ* hybridization and detected robust expression of *Nkx3-2* and *Rab28* in the developing fore- and hindlimb buds of Ctrl, LS1 and LS2 E12.5, in a domain broadly overlapping the presumptive zeugopod, the region including the tibia (fig. S9B). A third gene, *Bod1l*, straddled the TAD boundary with its promoter located in the neighboring TAD, making its regulation by sequences in the selected locus unlikely. Accordingly, *Bod1l* showed only weak or undetectable expression in the developing limb bud (fig. S9A). We next combined ENCODE chromatin profiles and our ATAC-Seq data to identify limb enhancers in the focal TAD. Here we found 3 novel enhancer candidates (N1, N2 and N3) carrying 3, 1 and 3 SNPs respectively, all of which showed significant allele frequency shifts in LS1 & 2 (range: 0.60–0.70 in LS1 and 0.82–0.85 in LS2; Fig. 5B; fig. S10A; Table S4). Chromosome conformation capture assays showed that the N1–N3 sequences formed long-range looping contacts with the *Nkx3-2* promoter—a hallmark of enhancers—despite nearly 600 kbp of intervening sequence (Fig. 5B). We next used transgenic reporter assays to determine if these sequences could drive limb expression and if so, whether the selected variants affect its enhancer activity. Whereas we found that the F0 alleles of the N1 and N3 enhancers (3 SNPs each in about 1 kbp) drove

336 robust and consistent *lacZ* expression in the developing limb buds as well as in other expanded
337 expression domains at E12.5, transgenic reporters carrying the selected F17 alleles of N1 and
338 N3 from the Chr 5: 41 Mbp locus showed consistently weak, nearly undetectable *lacZ*
339 expression (Fig. 5C). This is opposite to our observed direction of effect at the G2 enhancer at
340 the *Gli3* locus. At *Nkx3-2*, switching from the F0 to the F17 enhancer alleles led to a nearly
341 complete loss in activity ("loss-of-function"). The apparent paradox of a selectively favored
342 loss-of-function enhancer allele driving an *increase* in bone length may be explained by the role
343 of *Nkx3-2* as a repressor in bone formation (*28*). We hypothesize that the F17 allele causes *de-*
344 *repression* of bone formation by reducing enhancer activity and *Nkx3-2* expression. Crucially,
345 the F0 N1 enhancer showed activity that presages future long bone cartilage condensation in
346 the limb (Fig. 5D). This pattern recalls previous results that suggest that undetected early
347 expression of *Nkx3-2* may mark the boundaries and size of limb bone precursors, including the
348 tibia (*32*) (fig. S9C). Conversely, over-expression of *Nkx3-2* has been shown to cause
349 shortened tibia (even loss) in mice (*33, 34*). Thus, we have characterized two novel *Nkx3-2*
350 enhancers, N1 and N3. Notably, at the N3 enhancer, the DNA sequence of the F0 founding
351 allele was predicted to contain *Nkx3-1* and *Nkx3-2* binding sites, both of which were disrupted
352 by the selected F17 SNPs [Table S4; UNIPROBE database (*35*)]. Our results suggest that up to
353 6 total SNPs are sufficient to greatly modulate *Nkx3-2* enhancer activity in the developing limb
354 bud, presumably via disruptions of an auto-feedback loop. Just as phenotypes from amino
355 acid replacements may be more subtle than in knockout mutants, our *in situ* hybridization data
356 did not reveal differences in *Nkx3-2* expression domains between Ctrl or LS embryos (fig. S9B).
357 Finally, we connect the molecular changes to the selective value of this locus by reconstruct
358 the changes in allele frequencies at the N3 enhancer in all three lines in 1569 individuals across
359 20 generations. The trajectories show gradual changes and are consistent with our estimates
360 of the selection coefficient *s* of ~0.24 ± 0.12 at this locus (fig. S10C & D; see Supplementary

Notes section on "Estimating selection coefficient"). By combining population genetics, functional genomics and developmental genetic techniques, we were able to dissect a megabase-long locus and present data supporting the identification of up to 6 candidate quantitative trait nucleotides (QTNs). In the mouse, this represents a rare example of genetic dissection of a trait to the base-pair level.

Discussion

A central goal (and challenge) in studying evolution is to estimate the key factors that influence adaptive evolution and describe evolutionary change through time. Here we analyze the genomic changes in the Longshanks experiment, which was conducted under replicated and controlled conditions, allowing us to address several major evolutionary questions on how the genome of a complex vertebrate such as the mouse responds to selection.

An important conclusion from the Longshanks experiment is that tibia length increased readily and repeatedly in response to selection, even in an extremely bottlenecked population with as few as 14-16 breeding pairs. This is because the lines were founded with enough standing variation, and generation 17 was still only a fraction of the way to the characteristic time for the selection response at ~$2N_e$ generations (*36*), estimated here to be ~90 (fig. S2A; see Supplementary Notes on simulation). Our results underscore the importance of standing genetic variation in rapid adaptive response to a changing environment, a recurrent theme in natural adaptation (*20, 37, 38*) and breeding (*39*). By combining pedigree records with sequencing of founder individuals, here our data had sufficient details to allow precise modeling of trait response with predicted shifts in allele frequency distribution that closely matched our results, and with specific loci that we functionally validated. Our results imply a mixed genetic architecture with few discrete loci of large effect amid an infinitesimal background. This finding highlights another advantage of evolve-and-resequence (E&R) experiments over quantitative trait loci (QTL) mapping crosses, in that by sampling a much broader pool of alleles and

386 continually competing them against each other, the inferred genetic architecture and
387 distribution of effect sizes are more likely to be representative of the population at large.
388 Unlike other random evolutionary forces like demography, drift or recombination,
389 selection acts to favor specific variants. Accordingly, parallel evolution is often seen as a
390 hallmark for detecting selection (*21, 40-42*). We investigated the factors favoring parallelism by
391 contrasting the two Longshanks replicates against the Control line. We observed little to no
392 parallelism between selected lines and the Ctrl lines, or between simulated replicates of
393 selection, even though the simulated haplotypes were sampled directly from actual founders.
394 This underscores that parallelism depends on both shared selection pressure (absent in Ctrl)
395 *and* the availability of large effect loci that confer a substantial selection advantage (absent
396 under the infinitesimal model; Fig. 3A; fig. S8). Indeed, we estimate that the single *Nkx3-2*
397 locus contributes 9.4% of the total selection response (limits 3.6 – 15.5%; see Supplementary
398 Notes section "estimating selection coefficient" for details). This fraction is similar to effect size
399 estimates commonly cited in QTL mapping studies, except that here it is inferred not directly
400 from individual trait measurements, but indirectly, via changes in frequency of SNPs, under a
401 known selection regime. Overall, our results are consistent with theory, which shows that the
402 probability of parallel evolution for mutations of large effect can be high (*24*), even though
403 parallel loci may initially contribute a minority of the selection response. Beyond the top two
404 parallel loci, our broader parallelism results also agree with other empirical work on parallel
405 evolution (*42*) or E&R experiments showing correlated shifts in allele frequencies across
406 selection replicates (*19, 22*). Over a sufficiently long timescale with consistent selection
407 pressure, such standing variants may come to dominate the overall responses (*20*).
408 By systematically analyzing the impact of SNPs and functional enrichment, we found a
409 broad trend supporting a pivotal role of *cis*-regulatory changes in this example of morphological
410 evolution (fig. S7; see Supplementary Notes on genome-wide analysis on enrichment of

411 regulatory vs. coding mutations). This is further bolstered by our molecular dissection at two
412 specific loci, *Gli3* and its upstream regulator *Nkx3-2*. Both *Gli3* and *Nkx3-2* are broadly
413 expressed pleiotropic transcription factors, which are lethal when knocked out (*27, 43*). We
414 found only conservative (*Gli3*) or no amino acid changes (*Nkx3-2*). Rather, modulation of
415 tissue-specific expression by multiple enhancers likely played a more important role. We also
416 note that *Nkx3-2*, the locus with far stronger selection signature, acts upstream of *Gli3*, possibly
417 increasing the phenotypic impact [Fig. 4C and Table S4; (*34*)]. Our results here agree with
418 findings from another study across laboratory mouse strains that although coding changes tend
419 to have greater effects individually, across the whole genome the overall phenotypic impact is
420 minor compared to the vast number of regulatory variants (*44*). In fact, the genetic basis of a
421 moderate trait difference (typical for quantitative traits) is more likely to be due to regulatory
422 than to coding changes (*44*). Here, we rely on both genome-wide patterns and functional
423 testing of specific SNPs in enhancers to show the key role of regulatory changes in the
424 Longshanks mice. These results provide further support that the *cis*-regulatory hypothesis of
425 morphological evolution may also apply at the intra-specific level (*1*).

426 ***Conclusion***

427 Using the Longshanks selection experiment, we have detailed the genomic response to
428 selection at high resolution to reveal the underlying evolutionary and molecular mechanisms.
429 We show that parallelism may mark discrete, large effect loci in the selection response of a
430 polygenic trait and by combining a fully pedigreed experimental design with modern
431 sequencing and functional techniques, it is possible to identify some of the individual SNPs that
432 cause the response to selection on morphology. Further work should focus on dissecting the
433 mechanisms behind the dynamics of selective sweeps and/or polygenic adaptation by re-
434 sequencing the entire selection pedigree; testing how the selection response may depend on
435 the genetic architecture; and the extent to which linked variants within haplotypes may limit

436 inference of selection. Improved understanding in these areas may have broad implications for
437 conservation, rapid adaptation to climate change and quantitative genetics in medicine,
438 agriculture and in nature.

439 **References and Notes:**

440 1. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: a genetic theory of
441 morphological evolution. *Cell* **134**, 25-36 (2008), doi: 10.1016/j.cell.2008.06.030.
442 2. J. M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23-35
443 (1974), doi: 4407212.
444 3. N. H. Barton, Linkage and the limits to natural selection. *Genetics* **140**, 821-41 (1995).
445 4. S. P. Otto, N. H. Barton, Selection for recombination in small populations. *Evolution* **55**, 1921-
446 31 (2001), doi: 10.1111/j.0014-3820.2001.tb01310.x.
447 5. D. B. Weissman, N. H. Barton, Limits to the rate of adaptive substitution in sexual populations.
448 *PLoS Genet.* **8**, e1002740. (2012), doi: 10.1371/journal.pgen.1002740.
449 6. J. F. Crow, M. Kimura, Evolution in sexual and asexual populations. *Am. Nat.* **99**, 439-50
450 (1965).
451 7. W. G. Hill, A. Robertson. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**,
452 269-94 (1966).
453 8. M. Marchini, L. M. Sparrow, M. N. Cosman, A. Dowhanik, C. B. Krueger, B. Hallgrimsson, C.
454 Rolian, Impacts of genetic correlation on the independent evolution of body mass and skeletal
455 size in mammals. *BMC Evol. Biol.* **14**, 258. (2014), doi: 10.1186/s12862-014-0258-0.
456 9. F. Petit, K. E. Sears, N. Ahituv, Limb development: a paradigm of gene regulation. *Nat. Rev.*
457 *Genet.* **18**, 245-58 (2017), doi: 10.1038/nrg.2016.167.
458 10. M. Marchini, C. Rolian, Artificial selection sheds light on developmental mechanisms of limb
459 elongation. *Evolution* **72**, 825-37 (2018), doi: 10.1111/evo.13447.
460 11. T. H. Meuwissen, B. J. Hayes, M. E. Goddard, Prediction of total genetic value using
461 genome-wide dense marker maps. *Genetics* **157**, 1819-29 (2001).
462 12. D. S. Falconer, T. F. Mackay, *Introduction to quantitative genetics (4th Ed.)* (Pearson,
463 London, 1996).
464 13. B. Yalcin, J. Nicod, A. Bhomra, S. Davidson, J. Cleak, L. Farinelli, M. Østerås, A. Whitley, W.
465 Yuan, X. Gan, M. Goodson, P. Klenerman, A. Satpathy, D. Mathis, C. Benoist, D. J. Adams, R.
466 Mott, J. Flint, Commercially available outbred mice for genome-wide association studies. *PLoS*
467 *Genet.* **6**, e1001085. (2010).
468 14. A. P. Hendry, M. T. Kinnison, Perspective: the pace of modern life: measuring rates of
469 contemporary microevolution. *Evolution* **53**, 1637-53 (1999), doi: 10.1111/j.1558-
470 5646.1999.tb04550.x.
471 15. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches.
472 *Science* **296**, 707-11 (2002), doi: 10.1126/science.1070315.
473 16. W. Pitchers, J. B. Wolf, T. Tregenza, J. Hunt, I. Dworkin, Evolutionary rates for multivariate
474 traits: the role of selection and genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**,
475 20130252. (2014), doi: 10.1098/rstb.2013.0252.
476 17. L. Bünger, U. Renne, R. C. Buis, in *Encyclopedia of Genetics*, E. C. R. Reeve, Eds. (Fitzroy
477 Dearborn Publishers, Chicago, 2001), pp. 337-360.
478 18. P. D. Gingerich, Rates of evolution on the time scale of the evolutionary process. *Genetica*
479 **112-113**, 127-44 (2001).

19. M. K. Burke, J. P. Dunham, P. Shahrestani, K. R. Thornton, M. R. Rose, A. D. Long, Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature* **467**, 587-90 (2010), doi: 10.1038/nature09352.
20. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell *et al.*, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55-61 (2012), doi: 10.1038/nature10944.
21. Y. F. Chan, F. C. Jones, E. McConnell, J. Bryk, L. Bünger, D. Tautz, Parallel selection mapping using artificially selected mice reveals body weight control loci. *Curr. Biol.* **22**, 794-800 (2012), doi: 10.1016/j.cub.2012.03.011.
22. V. Soria-Carrasco, Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi, P. Nosil, Stick insect genomes reveal natural selection's role in parallel speciation. *Science* **344**, 738-42 (2014), doi: 10.1126/science.1252136.
23. J. K. Kelly, K. Hughes, An examination of the evolve-and-resequence method using *Drosophila simulans*. *bioRxiv* (2018), doi: 10.1101/337188.
24. H. A. Orr, The probability of parallel evolution. *Evolution* **59**, 216-20 (2005).
25. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012), doi: 10.1038/nature11082.
26. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213-8 (2013), doi: 10.1038/nmeth.2688.
27. D. Büscher, B. Bosse, J. Heymer, U. Rüther, Evidence for genetic control of *Sonic hedgehog* by *Gli3* in mouse limb development. *Mech. Dev.* **62**, 175-82 (1997).
28. S. Provot, H. Kempf, L. C. Murtaugh, U.-I. Chung, D.-W. Kim, J. Chyung, H. M. Kronenberg, A. B. Lassar. *Nkx3.2/Bapx1* acts as a negative regulator of chondrocyte maturation, *Development* **133**, 651-62 (2006), doi: 10.1242/dev.02258.
29. Y. Shen, F. Yue, D. F. McCleary, Z. Ye, L. Edsall, S. Kuan, U. Wagner, J. Dixon, L. Lee, V. V. Lobanenkov, B. Ren. A map of the *cis*-regulatory sequences in the mouse genome, *Nature* **488**, 116-20 (2012), doi: 10.1038/nature11243.
30. R. Akiyama, H. Kawakami, J. Wong, I. Oishi, R. Nishinakamura, Y. Kawakami, *Sall4-Gli3* system in early limb progenitors is essential for the development of limb skeletal elements. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 5075-80 (2015), doi: 10.1073/pnas.1421949112.
31. L. Koziel, M. Wuelling, S. Schneider, A. Vortkamp, *Gli3* acts as a repressor downstream of Ihh in regulating two distinct steps of chondrocyte differentiation. *Development* **132**, 5249-60 (2005), doi: 10.1242/dev.02097.
32. V. Sivakamasundari, H. Y. Chan, S. P. Yap, X. Xing, P. Kraus, T. Lufkin. New *Bapx1*(*Cre-*EGFP) mouse lines for lineage tracing and conditional knockout studies, *Genesis* **50**, 375-83 (2012), doi: 10.1002/dvg.20802.
33. Y. Bren-Mattison, M. Hausburg, B. B. Olwin, Growth of limb muscle is dependent on skeletal-derived *Indian hedgehog*, *Dev. Biol.* **356**, 486-95 (2011), doi: 10.1016/j.ydbio.2011.06.002.
34. C. Tribioli, T. Lufkin, *Bapx1* homeobox gene gain-of-function mice show preaxial polydactyly and activated *Shh* signaling in the developing limb, *Dev. Dyn.* **235**, 2483-92 (2006), doi: 10.1002/dvdy.20867.
35. M. F. Berger, G. Badis, A. R. Gehrke, S. Talukder, A. A. Philippakis, L. Peña-Castillo, T. M. Alleyne, S. Mnaimneh, O. B. Botvinnik, E. T. Chan, F. Khalid, W. Zhang, D. Newburger, S. A. Jaeger, Q. D. Morris, M. L. Bulyk, T. R. Hughes. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences, *Cell* **133**, 1266-76 (2008), doi: 10.1016/j.cell.2008.05.024.
36. A. Robertson, A theory of limits in artificial selection. *Proc. Roy. Soc. Lond. B Biol. Sci.* **153**, 234-49 (1960).

37. Heliconius Genome Consortium, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94-8 (2012), doi: 10.1038/nature11041.
38. A. M. Hancock, D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin, R. Sukernik, G. Utermann, J. K. Pritchard, G. Coop, A. Di Rienzo, Adaptations to climate-mediated selective pressures in humans. *PLoS Genet.* **7**, e1001375. (2011), doi: 10.1371/journal.pgen.1001375.
39. Z. Sheng, M. E. Pettersson, C. F. Honaker, P. B. Siegel, Ö. Carlborg, Standing genetic variation as a major contributor to adaptation in the Virginia chicken lines selection experiment. *Genome Biol.* **16**, 219. (2015), doi: 10.1186/s13059-015-0785-z.
40. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, Parallel evolution and inheritance of quantitative traits. *Am. Nat.* **163**, 809-22 (2004), doi: 10.1086/383621.
41. Y. F. Chan, M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, D. M. Kingsley, Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**, 302-5 (2010), doi: 10.1126/science.1182213.
42. A. Martin, V. Orgogozo, The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evolution* **67**, 1235-50 (2013), doi: 10.1111/evo.12081.
43. H. Akazawa, I. Komuro, Y. Sugitani, Y. Yazaki, R. Nagai, T. Noda, Targeted disruption of the homeobox transcription factor *Bapx1* results in lethal skeletal dysplasia with asplenia and gastroduodenal malformation. *Genes Cells* **5**, 499-513 (2000).
44. T. M. Keane, L. Goodstadt, P. Danecek, M. A. White *et al.*, Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* **477**, 289-94 (2011), doi: 10.1038/nature10413.

The following references appeared in Material and Methods only

45. A. Cox, C. L. Ackert-Bicknell, B. L. Dumont, Y. Ding, J. T. Bell, G. A. Brockmann, J. E. Wergedal, C. Bult, B. Paigen, J. Flint, A new standard genetic map for the laboratory mouse. *Genetics* **182**, 1335-44 (2009).
46. S. Picelli, A. K. Björklund, B. Reinius, S. Sagasser, G. Winberg, R. Sandberg, *Tn5* transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033-40 (2014), doi: 10.1101/gr.177881.114.
47. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-20 (2014), doi: 10.1093/bioinformatics/btu170.
48. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010), doi: 10.1093/bioinformatics/btp698.
49. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-303 (2010), doi: 10.1101/gr.107524.110.
50. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly. A framework for variation discovery and genotyping using next-generation DNA sequencing data, *Nat. Genet.* **43**, 491-8 (2011), doi: 10.1038/ng.806.
51. B. L. Browning, S. R. Browning, Genotype imputation with millions of reference samples, *Am. J. Hum. Genet.* **98**, 116-26 (2016).
52. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin. 1000 Genomes Project

579 Analysis Group. The variant call format and VCFtools, *Bioinformatics* **27**, 2156-8 (2011), doi:
580 10.1093/bioinformatics/btr330.
581 53. R. C. Team, *R: A language and environment for statistical computing* (2015).
582 54. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic
583 features. *Bioinformatics* **26**, 841-2 (2010).
584 55. FANTOM5 Consortium. A promoter-level mammalian expression atlas, *Nature* **507**, 462.
585 (2014).
586 56. J. H. Finger, C. M. Smith, T. F. Hayamizu, I. J. McCright *et al.*, The mouse Gene Expression
587 Database (GXD): 2017 update. *Nucleic Acids Res.* **45**, D730-6 (2017).
588 57. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M.
589 Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms,
590 SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*
591 **6**, 80-92 (2012).
592 58. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution
593 rates on mammalian phylogenies. *Genome Res.* **20**, 110-21 (2010).
594 59. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson,
595 J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W.
596 Miller, D. Haussler. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast
597 genomes, *Genome Res.* **15**, 1034-50 (2005).
598 60. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler.
599 The human genome browser at UCSC, *Genome Res* **12**, 996-1006 (2002).
600 61. A. R. Wood, T. Esko, J. Yang, S. Vedantam *et al.*, Defining the role of common variation in
601 the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173-86 (2014),
602 doi: 10.1038/ng.3097.
603 62. S. D. M. Brown, P. Chambon, M. H. de Angelis, EMPReSS: standardized phenotype screens
604 for functional annotation of the mouse genome. *Nat. Genet.* **37**, 1155- (2005).
605 63. P. Milani, R. Escalante-Chong, B. C. Shelley, N. L. Patel-Murray, X. Xin, M. Adam, B.
606 Mandefro, D. Sareen, C. N. Svendsen, E. Fraenkel. Cell freezing protocol suitable for ATAC-Seq
607 on motor neurons derived from human induced pluripotent stem cells, *Sci. Rep.* **6**, 25474.
608 (2016).
609 64. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**,
610 357-9 (2012).
611 65. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R.
612 M. Myers, M. Brown, W. Li, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.
613 (2008).
614 66. T. Sexton, S. Kurukuti, J. A. Mitchell, D. Umlauf, T. Nagano, P. Fraser, Sensitive detection of
615 chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat.*
616 *Protoc.* **7**, 1335-50 (2012), doi: 10.1038/nprot.2012.071.
617 67. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads.
618 *EMBnet. J.* **17**, pp-10 (2011).
619 68. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R.
620 Durbin, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format
621 and SAMtools. *Bioinformatics* **25**, 2078-9 (2009), doi: 10.1093/bioinformatics/btp352.
622 69. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human
623 genome. *Nature* **489**, 57-74 (2012), doi: 10.1038/nature11247.
624 70. R. J. DiLeone, G. A. Marcus, M. D. Johnson, D. M. Kingsley, Efficient studies of long-
625 distance *Bmp5* gene regulation using bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. U.*
626 *S. A.* **97**, 1612-7 (2000).
627 71. D. P. Mortlock, C. Guenther, D. M. Kingsley, A general approach for identifying distant
628 regulatory elements applied to the *Gdf6* gene. *Genome Res.* **13**, 2069-81 (2003).

629 72. P. Orozco-terWengel, M. Kapun, V. Nolte, R. Kofler, T. Flatt, C. Schlötterer. Adaptation of
630 Drosophila to a novel laboratory environment reveals temporally heterogeneous trajectories of
631 selected alleles, *Mol. Ecol.* **21**, 4931-41 (2012) doi: 10.1111/j.1365-294X.2012.05673.x.
632

633 **Acknowledgements**

634 We thank Felicity Jones for input into experimental design, helpful discussion and for improving
635 the manuscript. We thank the Rolian, Barton, Chan and Jones Labs members for support,
636 insightful scientific discussion and improving the manuscript. We thank the Rolian lab
637 members, the Animal Resource Centre staff at the University of Calgary and MPI Dresden
638 Animal Facility staff for animal husbandry. We thank Derek Lundberg for help with library
639 preparation automation. We thank Christa Lanz, Rebecca Schwab and Ilja Bezrukov for
640 assistance with high-throughput sequencing and associated data processing; Andre Noll for
641 high-performance computing support; the MPI Tübingen IT team for computational support.
642 pRS16 was a gift from François Spitz. We thank Mirna Marinič for creating an earlier version of
643 the transgenic reporter plasmid. We are indebted to Gemma Puixeu Sala, William G. Hill, Peter
644 Keightley for input and discussion on data analysis and simulation. We are also indebted to
645 Stefan Mundlos, Przemko Tylzanowki, Weikuan Gu and David M. Kingsley for suggested
646 experiments and sharing unpublished data. J.P.L.C. is supported by the International Max
647 Planck Research School “From Molecules to Organisms”. S.B. and N.B. are supported by IST
648 Austria. C.R. is supported by Discovery Grant #4181932 from the Natural Sciences and
649 Engineering Research Council of Canada and by the Faculty of Veterinary Medicine at the
650 University of Calgary. Y.F.C. is supported by the Max Planck Society.

651 **Author Contributions**

652 C.R. designed and initiated the Longshanks selection experiment. M.M. and C.R. performed
653 the selection, phenotyping and collected tissue samples for sequencing. Y.F.C. and C.R.
654 designed the sequencing strategy. W.H.B., M.K., prepared the samples and performed
655 sequencing. S.B., N.B. performed simulations and analyzed data. M.M., S.B., N.B., C.R.,

Y.F.C. analyzed the pedigree data. J.P.L.C., M.N.Y., M.K., I.S., J.C., C.R. and Y.F.C. designed, performed and analyzed results from functional experiments. J.P.L.C., R.N. and Y.F.C. planned and performed and analyzed the mouse transgenic experiments. S.B., M.N.Y., C.R., N.B. and Y.F.C analyzed the genomic data. All authors discussed the results and implications, wrote and commented on the manuscript at all stages.

Competing Financial Interests

The authors declare no competing interests. The Max Planck Society, IST Austria, the Natural Sciences and Engineering Research Council of Canada, and the Faculty of Veterinary Medicine of the University of Calgary provide funding for the research but no other competing interests.

Materials and Methods:

Animal Care and Use
All experimental procedures described in this study have been approved by the applicable University institutional ethics committee for animal welfare at the University of Calgary (HSACC Protocols M08146 and AC13-0077); or local competent authority: Landesdirektion Sachsen, Germany, permit number 24-9168.11-9/2012-5.

Reference genome assembly
All co-ordinates in the mouse genome refer to *Mus musculus* reference mm10, which is derived from GRCm38.

Code and data availability
Sequence data have been deposited in the GEO database under accession number [X]. Non-sequence data have been deposited at Dryad under accession number [Y]. Analytical code and additional notes have been deposited in the following repository: https://github.com/evolgenomics/Longshanks .

Pedigree data
Tibia length and body weight phenotypes were measured as previously described *(8)*. A total of 1332 Control, 3054 LS1, and 3101 LS2 individuals were recorded. Five outlier individuals with a skeletal dysplasia of unknown etiology were removed from LS2 and excluded from further analysis. Missing data in LS2 were filled in with random individuals that best matched the pedigree. Trait data were analyzed to determine response to selection based on the measured traits and their rank orders based on the selection index.

Simulations
Simulations were based on the actual pedigree and selection scheme, following one chromosome at a time. Each chromosome was represented by a set of junctions, which recorded the boundaries between genomes originating from different founder genomes; at the end, the SNP genotype was reconstructed by seeding each block of genome with the appropriate ancestral haplotype. This procedure is much more efficient than following each of the very large number of SNP markers. Crossovers were uniformly distributed, at a rate equal to the map length (*45*). Trait value was determined by a component due to an infinitesimal background (V_g); a component determined by the sum of effects of 10^4 evenly spaced discrete loci (V_s); and a Gaussian non-genetic component (V_e). The two genetic components had variance proportional to the corresponding map length, and the heritability was estimated from the observed trait values (see Supplementary Notes under "Simulations"). In each generation, the actual number of male and female offspring were generated from each breeding pair, and the male and female with the largest trait value were chosen to breed.

SNP genotypes were assigned to the founder genomes with their observed frequencies. However, to reproduce the correct variability requires that we assign founder *haplotypes*. This is not straightforward, because low-coverage individual genotypes cannot be phased reliably, and heterozygotes are frequently mis-called as homozygotes. We compared three procedures, which were applied within intervals that share the same ancestry: assigning haplotypes in linkage equilibrium (LE, or "no LD"); assigning heterozygotes to one or other genome at random, which minimises linkage disequilibrium, given the diploid genotype ("min LD"); and assigning heterozygotes consistently within an interval, which maximises linkage disequilibrium ("max LD") (fig. S2B). For details, see Supplementary Notes.

Significance thresholds
To obtain significance thresholds, we summarize the genome-wide maximum Δz^2 shift for each replicate of the simulated LS1 and LS2 lines, averaged within 10kb windows, and grouped by

the selection intensity and extent of linkage disequilibrium (LD). From this distribution of genome-wide maximum Δz^2 we obtained the critical value for the corresponding significance threshold (typically the 95th quantile or $P = 0.05$) under each selection and LD model (Fig. 3A; fig. S2D). This procedure controls for the effect of linkage or hitchhiking, line-specific pedigree structure, and selection strength.

Sequencing, genotyping and phasing pipeline

Sequencing libraries for high-throughput sequencing were generated using TruSeq or Nextera DNA Library Prep Kit (Illumina, Inc., San Diego, USA) according to manufacturer's recommendations or using equivalent *Tn5* transposase expressed in-house as previously described (*46*). Briefly, genomic DNA was extracted from ear clips by standard Protease K digestion (New England Biolabs GmbH, Frankfurt am Main, Germany) followed by AmpureXP bead (Beckman Coulter GmbH, Krefeld, Germany) purification. Extracted high-molecular weight DNA was sheared with a Covaris S2 (Woburn, MA, USA) or "tagmented" by commercial or purified *Tn5*-transposase according to manufacturer's recommendations. Each sample was individually barcoded (single-indexed as N501 with N7XX variable barcodes; all oligonucleotides used in this study were synthesized by Integrated DNA Technologies, Coralville, Iowa, USA) and pooled for high-throughput sequencing by a HiSeq 3000 (Illumina) at the Genome Core Facility at the MPI Tübingen Campus. Sequenced data were pre-processed using a pipeline consisting of data clean-up, mapping, base-calling and analysis based upon `fastQC v0.10.1` (*52*); `trimmomatic v0.33` (*47*); `bwa v0.7.10-r789` (*48*); `GATK v3.4-0-gf196186` modules `BQSR`, `MarkDuplicates`, `IndelRealignment` (*49, 50*). Genotype calls were performed using the `GATK HaplotypeCaller` under the `GENOTYPE_GIVEN_ALLELES` mode using a set of high-quality SNP calls made available by the Wellcome Trust Sanger Centre (Mouse Genomes Project version 3 dbSNP v137 release (*44*), after filtering for sites segregating among inbred lines that may have contributed to the original 7 female and 2 male CD-1 founders, namely 129S1/SvImJ, AKR/J, BALB/cJ, BTBR *T*$^{+}$ *Itpr3*^{tf}/J, C3H/HeJ, C57BL/6NJ, CAST/EiJ, DBA/2J, FVB/NJ, KK/HiJ, MOLF/EiJ, NOD/ShiLtJ, NZO/HlLtJ, NZW/LacJ, PWK/PhJ and WSB/EiJ based on (*13*). We consider a combined ~100x coverage sufficient to recover any of the 18 CD-1 founding haplotypes still segregating at a given locus. The raw genotypes were phased with `Beagle v4.1` (*51*) based on genotype posterior likelihoods using a genetic map interpolated from the mouse reference map (*45*) and imputed from the same putative CD-1 source lines as the reference panel. The site frequency spectra (SFS) were evaluated to ensure genotype quality (fig. S3A).

Population genetics summary statistics

Summary statistics of the F0 and F17 samples were calculated genome-wide [Weir–Cockerham F_{ST}, π, heterozygosity]; in adjacent 10 kbp windows (Weir–Cockerham F_{ST}, π, allele frequencies p and q), or on a per site basis (Weir–Cockerham F_{ST}, π, p and q) using `VCFtools v0.1.14` (*52*). The summary statistics Δz^2 was the squared within-line difference in arcsine square root transformed MAF q; it ranges from 0 to π^2. The resulting data were further processed by custom bash, Perl and `R v3.2.0` (*53*) scripts.

Peak loci and filtering for hitchhiking windows

Peak loci were defined by a descending rank ordering of all 10 kbp windows, and from each peak signals the windows were extended by 100 SNPs to each side, until no single SNP rising above a Δz^2 shift of 0.2 π^2 was detected. A total of 810 peaks were found with a Δz^2 shift ≥ 0.2 for LS1 & 2. Following the same procedure, we found 766 peaks in Ctrl.

Candidate genes

To determine whether genes with related developmental roles were associated with the selected variants, the topologically associating domains (TADs) derived from mouse embryonic stem cells as defined elsewhere (*25*) were re-mapped onto mm10 co-ordinates. Genes within the TAD

overlapping within 500 kbp of the peak window ("core span") were then cross-referenced against annotated knockout phenotypes (Mouse Genome Informatics, http://www.informatics.jax.org). This broader overlap was chosen to account for genes whose regulatory sequences like enhancers but not their gene bodies fall close to the peak window. We highlight candidate genes showing limb- and bone-related phenotypes, e.g., with altered limb bone lengths or epiphyseal growth plate morphology, as observed in Longshanks (*10*), of the following categories (along with their Mammalian Phenotype Ontology term and the number of genes): "abnormal tibia morphology/MP:0000558" (212 genes), "short limbs/MP:0000547" and "short tibia/MP:0002764" (223 genes), "abnormal cartilage morphology/MP:0000163" (321 genes), "abnormal osteoblast morphology/MP:0004986" (122 genes). Note that we exclude compound mutants or those conditional mutant phenotypes involving transgenes. To determine if the overlap with these genes are significant, we performed 1000 permutations of the core span using `bedtools v2.22.1 shuffle` with the `-noOverlapping` option (*54*) and excluding ChrY, ChrM and the unassembled scaffolds. We then followed the exact procedure as above to determine the number of genes in the overlapping TAD belonging to each category. We reported the quantile rank as the *P*-value, ignoring ties. To determine other genes in the region, we list all genes falling within the entire hitchhiking window (Table S3).

Identification of putative limb enhancers

We downloaded publicly available chromatin profiles, derived from E14.5 limbs, for the histone H3 lysine-4 (K4) or lysine-27 (K27) mono-/tri-methylation or acetylation marks (H3K4me1, H3K4me3 and H3K27ac) generated by the ENCODE Consortium (*29*). We intersected the peak calls for the enhancer-associated marks H3K4me1 and H3K27ac and filtered out those overlapping promoters [H3K4me3 and promoter annotation according to the FANTOM5 Consortium (*55*)].

Enrichment analysis

To calculate enrichment through the whole range of Δz^2, a similar procedure was taken as in *Candidate genes* above. For knockout gene functions, genes contained in TADs within 500 kbp of peak windows were included in the analysis. We use the complete database of annotated knockout phenotypes for genes or spontaneous mutations, after removing phenotypes reported under conditional or polygenic mutants. For gene expression data, we retained all genes which have been reported as being expressed in any of the limb structures, by tracing each anatomy ontological term through its parent terms, up to the top level groupings, e.g., "limb", in the Mouse Genomic Informatics Gene Expression Database (*56*). For E14.5 enhancers, we used a raw 500 kbp overlap with the peak windows, because enhancers, unlike genes, may not have intermediaries and may instead represent direct selection targets.

For coding mutations, we first annotated all SNPs for their putative effects using `snpEff v4.0e` (*57*). To accurately capture the per-site impact of coding mutations, we used per-site Δz^2 instead of the averaged 10 kbp window. For each population, we divided all segregating SNPs into up to 0.02 bands based on per-site Δz^2. We then tracked the impact of coding mutations in genes *known to be expressed in limbs*, as above. We reported the sum of all missense ("moderate" impact), frame-shift, stop codon gain or loss sites ("high impact"). A linear regression was used to evaluate the relationship between Δz^2 and the average impact of coding SNPs (SNPs with high or moderate impact to all coding SNPs).

For regulatory mutations, we used the same bins spanning the range of Δz^2, but focused on the subset of SNPs falling within the ENCODE E14.5 limb enhancers. We then obtained a weighted average conservation score based on an averaged phastCons (*58*) or phyloP (*59*) score in ±250 bp flanking the SNP, calculated from a 60-way alignment between placental mammal genomes [downloaded from the UCSC Genome Browser (*60*)]. We reported the average conservation score of all SNPs within the bin and fitted a linear regression on log-scale. In particular, phastCons scores range from 0 (un-conserved) to 1 (fully conserved), whereas phyloP

is the $|log_{10}|$ of the *P*-value of the phylogenetic tree, expressed as a positive score for conservation and a negative score for lineage-specific accelerated change. We favored using phastCons for its simpler interpretation.

Impact of coding variants
Using the same SNP effect annotations described in the section above, we checked whether any specific SNP with significant site-wise Δz^2 in either LS1 or LS2 cause amino acid changes or protein disruptions and are known to cause limb defects when knocked out. For each position we examined outgroup sequences using the 60-way placental mammal alignment to determine the ancestral amino acid state and whether the selected variant was consistent with purifying vs. diversifying selection. The resulting 12 genes that match these criteria are listed in Table S5.

Association with human height loci
To test if loci known to be associated with human height are clustered with the selected loci in the LS lines, we downloaded the set of previously published 697 SNPs (*61*). In order to facilitate mapping to mouse co-ordinates, each SNP was expanded to 100 kbp centering on the SNP and converted to mm10 positions using the `liftOver` tool with the multiple mapping option disabled (*60*). We were able to assign positions in 655 out of the 697 total SNPs. Then for each of the 810 loci above the $H_{INF,\ no\ LD}$ threshold the minimal distance to any of the mapped human loci was determined using `bedtools closest` with the `-d` option (*54*). Should a region actually overlap, a distance of 0 bp was assigned. To generate the permuted set, the 810 loci were randomly shuffled across the mouse autosomes using the `bedtools shuffle` program with the `-noOverlapping` option. Then the exact same procedure as the actual data was followed to determine the closest interval. The resulting permuted intervals follow an approximately normal distribution, with the actual observed results falling completely below the range of permuted results.

In situ hybridization
Detection of specific gene transcripts were performed as previously described in (*62*). Probes against *Nkx3-2*, *Rab28*, *Bod1l* and *Gli3* were amplified from cDNA from wildtype C57BL/6NJ mouse embryos (Table S6). Amplified fragments were cloned into pJET1.2/blunt plasmid backbones in both sense and anti-sense orientations using the CloneJET PCR Kit (Thermo Fisher Scientific, Schwerte, Germany) and confirmed by Sanger sequencing using the included forward and reverse primers. Probe plasmids have also been deposited with Addgene. *In vitro* transcription from the T7 promoter was performed using the MAXIscript T7 *in vitro* Transcription Kit (Thermo Fisher Scientific) supplemented with Digoxigenin-11-UTP (Sigma-Aldrich) (MPI Tübingen), or with T7 RNA polymerase (Promega) in the presence of DIG RNA labelling mix (Roche) (University of Calgary). Following TURBO DNase (Thermo Fisher Scientific) digestion probes were cleaned using SigmaSpin Sequencing Reaction Clean-Up columns (Sigma-Aldrich) (MPI Tübingen), or using Illustra MicroSpin G-50 columns (GE Healthcare) (University of Calgary). During testing of probe designs, sense controls were used in parallel reactions to establish background non-specific binding.

ATAC-seq library preparation and sequencing pipeline
ATAC-seq was performed on dissected C57BL/6NJ E14.5 forelimb and hindlimb. Nuclei preparation and tagmentation were performed as previously described in (*26*), with the following modifications. To minimize endogenous protease activity, cells were strictly limited to 5 + 5 minutes of collagenase A treatment at 37 °C, with frequent pipetting to aid dissociation into single-cell suspensions. Following wash steps and cell lysis, 50 000 nuclei were tagmented with expressed *Tn5* transposase. Each tagmented sample was then purified by MinElute columns (Qiagen) and amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs) using a uniquely barcoded i7-index primer (N701-N7XX) and the N501 i5-index primer. PCR thermocycler programs were 72°C for 4 min, 98°C for 30 s, 6 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 1 min, and final extension at 72°C for 4 min. PCR-enriched samples were taken

through a double size selection with PEG-based SPRI beads (Beckman Coulter) first with 0.5X ratio of PEG/beads to remove DNA fragments longer than 600 bp, followed by 1.8X PEG/beads ratio in order to select for Fraction A as described in (*63*). Pooled libraries were run on the HiSeq 3000 (Illumina) at the Genome Core Facility at the MPI Tübingen Campus to obtain 150 bp paired end reads, which were aligned to mouse mm10 genome using `bowtie2 v.2.1.0` (*64*). Peaks were called using `MACS14 v.2.1` (*65*).

Multiplexed chromosome conformation capture (4C-Seq)
Chromosome conformation capture (3C) template was prepared from pooled E14.5 liver, forelimb and hindlimb buds (n = 5–6 C57BL/6NJ embryos per replicate), with improvements to the primer extension and library amplification steps following (*66*). The template was amplified with Q5 High-Fidelity Polymerase (New England Biolabs GmbH, Frankfurt am Main, Germany) using a 4C adapter-specific primer and a pool of 6 *Nkx3-2* enhancer viewpoint primers [and, in a separate experiment, a pool of 8 *Gli3* enhancer-specific viewpoint primers; Table S7]. Amplified fragments were prepared for Illumina sequencing by ligation of TruSeq adapters, followed by PCR enrichment. Pooled libraries were sequenced by a HiSeq 3000 (Illumina) at the Genome Core Facility at the MPI Tübingen Campus with single-end, 150 bp reads. Sequence data were processed using a pipeline consisting of data clean-up, mapping, and analysis based upon `cutadapt v1.10` (*67*); `bwa v0.7.10-r789` (*48*) ; `samtools v1.2` (*68*); `bedtools` (*54*) and `R v3.2.0` (*53*). Alignments were filtered for ENCODE blacklisted regions (*69*) and those with `MAPQ` scores below 30 were excluded from analysis. Filtered alignments were binned into genome-wide *BglII* fragments, normalized to Reads Per Kilobase of transcript per Million mapped reads (RPKM), and plotted and visualized in `R`.

Plasmid construction
Putative limb enhancers corresponding to the F0 and F17 alleles of the *Gli3* G2 and *Nkx3-2* N1 and N3 enhancers were amplified from genomic DNA of Longshanks mice from the LS1 F0 (9 mice) and F17 (10 mice) generations and sub-cloned into pJET1.2/blunt plasmid backbone using the CloneJET PCR Kit (Thermo Fisher Scientific) and alleles were confirmed by Sanger sequencing using the included forward and reverse primers (Table S8). Each allele of each enhancer was then cloned as tandem duplicates with junction *SalI* and *XhoI* sites upstream of a β-globin minimal promoter in our reporter vector (see below). Constructs were screened for the enhancer variant using Sanger sequencing. All SNPs were further confirmed against the rest of the population through direct amplicon sequencing.

The base reporter construct pBeta-lacZ-attBx2 consists of a β-globin minimal promoter followed by a *lacZ* reporter gene derived from pRS16, with the entire reporter cassette flanked by double *attB* sites. The pBeta-lacZ-attBx2 plasmid and its full sequence have been deposited and is available at Addgene.

Pronuclear injection of F0 and F17 enhancer-reporter constructs
The reporter constructs containing the appropriate allele of each of the 3 enhancers were linearized with *ScaI* (or *BsaI* in the case of the N3 F0 allele due to the gain of a *ScaI* site) and purified. Microinjection into mouse zygotes was performed essentially as described (*70*). At 12 d after the embryo transfer, the gestation was terminated and embryos were individually dissected, fixed in 4% paraformaldehyde for 45 min and stored in PBS. All manipulations were performed by R.N. or under R.N.'s supervision at the Transgenic Core Facility at the Max Planck Institute of Cell Biology and Genetics, Dresden, Germany. Yolk sacs from embryos were separately collected for genotyping and all embryos were stained for *lacZ* expression as previously described (*71*). Embryos were scored for *lacZ* staining, with positive expression assigned if the pattern was consistently observed in at least two embryos.

Genotyping of time series at the Nkx3-2 *N3* locus

Allele-specific primers terminating on SNPs that discriminate between the F0 from the F17 N3 enhancer alleles were designed (rs33219710 and rs33600994; Table S9). The amplicons were optimized as a qPCR reaction to give allele-specific, present/absent amplifications (typically no amplification for the absent allele, otherwise average ΔCt > 10). Genotyping on the entire breeding pedigree of LS1 (n = 602), LS2 (n = 579) and Ctrl (n = 389) was performed in duplicates for each allele on a Bio-Rad CFX384 Touch instrument (Bio-Rad Laboratories GmbH, Munich, Germany) with SYBR Select Master Mix for CFX (Thermo Fisher Scientific) and the following qPCR program: 50ºC for 2 min, 95ºC for 2 min, 40 cycles of 95ºC for 15 s, 58ºC for 10 s, 72ºC for 10 s. In each qPCR run we included individuals of each genotype (LS F17 selected homozygotes, heterozygotes and F0 major allele homozygotes). For the few samples with discordant results between replicates, DNA was re-extracted and re-genotyped or otherwise excluded.

Supplementary Information

Supplementary Notes

Supplementary Figures

Supplementary Tables

Supplementary Material

Supplementary Methods

Supplementary Notes

Major considerations in constructing the simulations

In the Longshanks experiment, the highest-ranking male and the highest-ranking female from each family were chosen to breed with the highest-ranking mice from other families within a line (i.e., disallowing sibling matings). Thus, if we disregard non-Mendelian segregation, and the fraction of failed litters (15%), selection acts solely within families, on the measured traits. Such selection does not distort the pedigree, and allows us to follow the evolution of each chromosome separately.

Our simulations track the inheritance of continuous genomes by following the junctions between regions with different ancestry. In principle, we should simulate selection under the infinitesimal model by following the contributions to the trait of continuous blocks of chromosomes across the whole genome. However, this is computationally challenging, since the contributions of all the blocks defined by every recombination event have to be tracked. Instead, we follow a large number of discrete biallelic loci checking that the number is sufficiently large to approach the infinitesimal limit (fig. S2C). We made a further slight approximation by only explicitly modelling discrete loci on one chromosome at a time. We divided the breeding value of an individual into two components. The first, V_g, is a contribution from a large number of unlinked loci, due to genes on all but the focal chromosome, as represented by the infinitesimal model. The values of this component amongst offspring are normally distributed around the mean of the parents, with its variance being:

$$V_M = (V_A/2)\,(1 - \beta)\,(1 - F_{ii} - F_{jj})$$

where: $V_{A,0}$ is the initial genetic variance, and
F_{ii}, F_{jj} are the probabilities of identity between distinct genes in each parent, i, j;
F_{ii}, F_{jj} are calculated from the pedigree;
β is the fraction of genome on the focal chromosome.

The second component, V_s, is the sum of contributions from a large number, n, of discrete loci, evenly spaced along the focal chromosome (here we used 1,0000), and contributing a fraction β of the initial additive variance. We choose these to have equal effects and random signs, $\pm\alpha$, such that initial allele frequencies $p_0 = q_0 = \frac{1}{2}$, and equal effects α, such that $\beta V_{A,0} = 2\sum_{i=1}^{n} \alpha^2 p_{i,0} q_{i,0}$. The initial population consists of 28 diploid individuals, matching the experiment, and loci have initial frequencies of 1, 4, 12 and 28 out of the diploid total of 56 alleles, in equal proportions. Inheritance is assumed to be autosomal, with no sex-linkage. This choice of equal effects approaches most closely to the infinitesimal model, for a given number of loci.

The decrease in genetic variance due to random drift is measured by the inbreeding coefficient, defined as the probability of identity by descent, relative to the initial population. We distinguish the identity between two distinct genes within a diploid individual, F_w,from the probability of identity between two genes in different individuals, F_b. The overall mean identity between two genes chosen independently and at random from all $2N$ genes is

$\bar{F} = \frac{2(N-1)F_b + F_w + 1}{2N}$. The proportion of heterozygotes in the population decreases by a factor of $1 - F_w$, the variance in allele frequency increases with $\bar{F}$, and the genetic diversity, $\mathbb{E} = [2pq]$, decreases as $1 - \bar{F}$.

fig. S2A shows that in the absence of selection, the identity F_b increases slower than expected under the Wright–Fisher model with the actual population sizes (compare light shaded lines with black lines). These differences are a consequence of the circular mating scheme, which was designed to slow the loss of variation. The dotted line show the average F, estimated from the loss of heterozygosity in 50 replicate neutral simulations, each with 10^4 loci on a chromosome of length *R*=1 Morgan. These are close to the prediction from the pedigree (light shaded lines), validating the simulations.

The thick colored line in fig. S2A show F, estimated in the same way from simulations that include truncation selection on a trait with within-family variance V_s/V_e = 0.584 (a value we abbreviate as $\theta = 1$), which matches the observed selection response and parent-offspring regression. The rate of drift, as measured by the gradient in F over time, is significantly faster in simulations with selection, by 6.7% in LS1 and 9.8% in LS2 (Student's *t*-test $P \leq 0.008$ in LS1 and $P \leq 0.0005$ in LS2). However, this effect of selection would not be detectable from any one replicate, since the standard deviation of the rate of drift, relative to the mean rate, is ~13% between replicates. On average, the observed loss of heterozygosity fits closely to that expected from the pedigree (large dot with error bars), though there is wide variation among chromosomes (filled dots), which is substantially higher than seen in simulations seeded with SNP at linkage equilibrium (compare filled and open dots).

We then performed 100 simulations, seeding each founding generation with actual genotypes and using actual pedigrees, selection pressure or heritability parameters (within-family heritability h^2 of the fitness dimension: 0.51; fig. 2B). A main conclusion from our modelling is that the overall allele frequencies were hardly perturbed by varying selection from random drift to even doubling the selection intensity. Upon closer examination, it became clear that under the standard "infinitesimal" model, selection could generate a weak but detectable excess of allele frequency sweeps compared to strict neutrality with no selection (fig. S2C, SNP classes 1/56 and 4/56). However, it may take many replicates (assuming no parallelism) for this excess to become statistically significant. Taken at face value, this result echoes many "evolve-and-resequence" (E&R) experiments based on diverse base populations that show only weak evidence of selective sweeps at specific loci (*19*, *72*).

Broader patterns and analyses of parallelism

On a broader scale, we also observed greater extent of parallelism globally than in the simulated results or with the empirical Ctrl line. For example, out of the 2405 and 2991 loci found above the $H_{INF,\ no\ LD}$ cut-off in LS1 and LS2, 398 were found in both lines (13%; χ^2 test, N~150,000 windows; χ^2=2901.4, d.f.=1, $P \leq 1 \times 10^{-10}$); whereas we found only 10 or 7 overlaps in Ctrl–LS1 or Ctrl–LS2 comparisons, respectively. This difference is statistically significant (940 significant Ctrl loci at the $H_{INF,\ no\ LD}$ threshold; N~150,000 windows; Ctrl–LS1: χ^2=0.7; Ctrl–LS2: χ^2=6.0; both *P* = n.s.; see also fig. S8). In fact, there was not a single window out of a total of 8.4 million windows from the 100 replicates where both simulated LS1 and LS2 replicates simultaneously cleared the $H_{INF,\ no\ LD}$ threshold. In contrast to our earlier analysis in single LS replicates, the parallel selected loci in both LS replicates loci may provide the strongest evidence yet to reject the infinitesimal model.

Enrichment for genes with functional impact on limb development

To determine what types of molecular changes may have mediated the selection response, we performed a gene set enrichment analysis. We asked if the outlier loci found in the Longshanks lines were enriched for genes affecting limb development (as indicated by their knockout phenotypes) and found increasingly significant enrichment as the allele frequency shift Δz^2 cut-off became increasingly stringent (fig. S7A). The "limb/digital/tail" category of affected anatomical systems in the Mouse Genomic Informatics Gene Expression Database (*56*) showed the greatest excess of observed-to-expected ratio out of all 28 phenotype categories (the excluded "normal" category also showed no enrichment). In contrast, genes showing knock-out phenotypes in most other categories did not show similar enrichment as Δz^2 became more stringent (fig. S7A). For genes expressed in limb tissue, there was a similar, but weaker increase, with the enrichment only appearing at higher Δz^2 cut-off. We did not observe similar enrichment using data and thresholds derived from Ctrl (fig. S7A, lower panels). To investigate the impact on regulatory sequences, we obtained 21,211 limb enhancers predicted by ENCODE chromatin profile at a stage immediately preceding bone formation (Theiler Stage 23, at approximately embryonic day E14.5) (*29*). We found likewise an enrichment throughout the range of significance cut-offs (fig. S7A). Again, there was no similar enrichment in Ctrl.

Clustering with loci associated with human height

Since tibia lengths directly affect human height, we tested if an association exists between loci controlling human height (*61*) and a set of 810 loci at the $P \leq 0.05$ significance level under $H_{INF,\ no\ LD}$ described here. After remapping the human loci to their orthologous mouse positions (n = 655 out of 697 total height loci; data from GIANT Consortium), we detected significant clustering with the 810 peak loci (mean pairwise distance to remapped height loci: 1.41 Mbp vs. mean 1.69 Mbp from 1000 permutations of shuffled peak loci, range: 1.45–1.93 Mbp; n = 655 height loci and 810 peak loci; $P < 0.001$, permutations). We interpret this clustering to suggest that a shared and conserved genetic program exist between human height and tibia length and/or body mass.

Genome-wide analysis of the role of coding vs. cis*-acting changes in response to selection*

We examined the potential functional impact of coding or regulatory changes as a function of Δz^2 in all three lines. For coding changes, we tracked the functional consequences of coding SNPs of moderate to high impact (missense mutations, gain or loss of stop codons, or frame-shifts). Whereas we found only mixed evidence of increased coding changes as Δz^2 increased in the LS lines, there was a depletion of coding changes in Ctrl line as Δz^2 increased, possibly due to purifying or background selection (fig. S7B; linear regression, LS1: $P \leq 0.015$, slope > 0; LS2: $P = 0.62$, n.s., slope ≈ 0; Ctrl: $P \leq 5.72\times10^{-9}$, slope < 0).

For regulatory changes, we used sequence conservation in limb enhancers overlapping a SNP as a proxy for functional impact. In contrast to the situation for coding changes, where the correlations differed between LS1 and LS2, the potential impact of regulatory changes increased significantly as a function of Δz^2 in both LS lines (fig. S7B): within limb enhancers, SNP-flanking sequences became increasingly conserved at highly differentiated SNPs (phastCons conservation score, ranging from 0 to 1 for unconserved to completely conserved positions; linear regression, log-scale, $P < 1.05\times10^{-9}$ for both, slopes > 0). This relationship also exists for the Ctrl line, albeit principally from lower Δz^2 and conservation values ($P < 0.8\times10^{-3}$,

slope > 0; fig. S7B). Taken together, our enrichment analysis suggests that while both coding and regulatory changes were selected in the Longshanks experiment, the overall selection response may depend more consistently on *cis*-regulatory changes, especially for developmental regulators involved in limb, bone and/or cartilage development (Table 1; Table S3; c.f. Table S4 for coding changes). This is a key prediction of the "*cis*-regulatory hypothesis", especially in its original scope on morphological traits (*1*).

Genes with amino acid changes of potentially major impact

We have further identified 12 candidate genes with likely functional impact on limb development due to specific amino acid changes showing large frequency shifts (albeit only one, *Fbn2*, cleared the stringent $P \leq 0.05$ $H_{INF,\ max\ LD}$ threshold; 6 in LS1, 9 in LS2, of which 3 were shared; Table S4). Consistent with strong selection for tibia development, all 12 genes show limb or tail phenotypes when knocked out, e.g., "short limbs" for the collagen gene *Col27a1* knockout. Most of these genes encode for structural cellular components, e.g., myosin, fibrillin and collagen (*Myo10; Fbn2*; and *Col27a1* respectively), with *Fuz* (fuzzy planar cell polarity protein) being the only classical developmental regulator gene. All but one of these genes have also been shown to have widespread pleiotropic effects with broad expression domains, and their knockouts were often lethal (eight out of 12) and/or exhibit defects in additional organ systems (11 out of 12). Based on this observation, we anticipate that the phenotypic impact of these selected coding missense SNPs (n.b. not knockout) would not be restricted to tibia or bone development.

Estimating the selection coefficient of the top-ranking locus, Nkx3-2*, from changes in allele frequency*

The significant locus on Chr5 containing *Nkx3-2* shows strong changes in SNP frequency in both LS1 and LS2. Here, we estimate the strength of selection on this locus, and the corresponding effect on the selected trait. We approximate by assuming two alternative alleles, and find the selection coefficient implied the observed parallel changes in allele frequency; we then set bounds on this estimate that take account of random drift. Finally, we use simulations that condition on the known pedigree to estimate the effect on the trait required to cause the observed strong frequency changes; these show that linked selection has little effect on the single-locus estimates.

We see strong and parallel changes in allele frequency at multiple steps. There are 14 non-overlapping 10kb windows that have a mean square change in arc-sin transformed allele frequency of $\overline{\Delta z^2} > 2$ in both LS1 and LS2, spanning a 260 kbp region and including 807 SNP. SNP frequencies are tightly clustered, corresponding to two alternative haplotypes (Fig. 5A & fig. S10A). The initial (untransformed) allele frequencies average q_0 = 0.18, 0.17 in LS1, LS2, respectively, and the final frequencies average q_{17} = 0.84, 0.98, respectively (also see fig. S10A, lower panel). These frequencies depend on the arbitrary threshold for which windows to include. However, this makes little difference, relative to the wide bounds on our estimates.

Under constant selection, $log\frac{p}{q}$ changes linearly with time, at a rate equal to the selection coefficient, s. Therefore, a naive estimate of selection is given by $\hat{s} = \frac{1}{T}\log\left[\frac{p_{17}}{q_{17}}\frac{q_0}{p_0}\right]$; thus, $\hat{s}$ =0.19, 0.32 for LS1, LS2, and averages 0.26. Here, males and females with longest legs are chosen to breed; the strength of selection on an additive allele depends on the fraction selected and the within-family trait variance. The former is kept constant, and there is little loss

of variance due to inbreeding (F~0.17), and so assuming constant selection is reasonable (fig. S10B), unless there is strong dominance.

To set bounds on this estimate, we must account for random drift. The predicted loss of diversity over 17 generations, based on the pedigree, is F=0.173, 0.175 for LS1, LS2, which corresponds to an effective size $N_e = 44.9, 44.4$, respectively. Therefore, we calculate the matrix of transition probabilities for a Wright–Fisher population with $2N$ rounded to 90, 89 copies for LS1, LS2, over a range of selection, s. This yields the probability that the number of copies would change from the rounded values of 16/90 to 75/90 in LS1, and from 14/89 to 87/89 in LS2—that is, the likelihood of s, given the observed changes in allele frequency, and the known N_e. There is no significant loss of likelihood by assuming the same selection in both lines; overall, $\hat{s}$ = 0.24 (limits 0.13–0.36; fig. S10C).

Estimating the selection coefficient, accounting for linked loci

The estimates above using the simple approach do not account for selection on linked loci, and do not give the effect on the composite trait. We therefore simulated conditional on the pedigree and on the actual selection regime, as described above, but including an additive allele with effect *A* at the candidate locus on Chr5. The genetic variance associated with the unlinked infinitesimal background, and across Chr5, were reduced in proportion, to keep the overall heritability the same as before $V_a/(V_a + V_e) = 0.539$. The selection coefficient inferred from the simulated changes in allele frequency was approximately proportional to the effect on the trait, with best fit $s = 0.41A/\sqrt{V_e}$ (fig. S10D, left). Assuming this relationship, we can compare the mean and standard deviation of allele frequency from simulations with linked selection, with that predicted by the single locus Wright–Fisher model (points vs. line in fig. S10D, middle & right). These agree well, showing that linked selection does not appreciably change the distribution of allele frequencies at a single locus. This is consistent with fig. S2C, which shows that linked selection only inflates the tail of the allele frequency distribution, an effect that would not be detectable at a single locus.

Combining our estimates of the selection coefficient with the relation $s = 0.41A/\sqrt{V_e}$, we estimate that the locus on Chr5 has effect $\hat{A} = 0.59\sqrt{V_e}$, with 2-unit support limits $0.32\sqrt{V_e}$ to $0.87\sqrt{V_e}$. This single locus is responsible for ~9.4% of the total selection response (limits 3.6–15.5%).

This analysis does not allow for the inflation of effect that might arise from multiple testing. This is hard to estimate, because it depends on the distribution of effects across the genome, and also on the excess variation in estimates due to LD in the founder population. However, we note that if the effect of this locus is large enough that it would certainly be detected in this study, then there is no estimation bias from this source.

We also assume that there are two haplotypes, each with a definite effect. There might in fact be heterogeneity in the effects of each haplotype, for two reasons. First, this region might have had heterogeneous effects in the founder population, with multiple alleles at multiple causal loci. Second, as recombination breaks up the founder genomes, blocks of genome would become associated with different backgrounds. To the extent that genetic variation is spread evenly over an infinitesimal background, this latter effect is accounted for by our simulations, and has little consequence. However, we have not tested whether the data might be explained by more than two alleles, possibly at more than one discrete locus. Testing

such complex models would be challenging, and we do not believe that such test would have much power. However, the estimates of selection made here should be regarded as effective values that may reflect a more complex reality.

Supplementary figures

fig. S1
Artificial selection allowed detailed reconstruction of selection parameters.

fig. S2
Simulating selection on pedigrees.

fig. S3
Broad similarity in molecular diversity in the founder populations for the Longshanks lines and the Control line.

fig. S4
Allele frequencies between the founder populations were very similar.

fig. S5
Selected lines showed more extreme values of Δz^2 than the Control line.

fig. S6
Detailed Δz^2 profiles at the LS significant loci.

fig. S7
Loci associated with selection response in Longshanks lines show enrichment for limb function likely associated with *cis*-acting mechanisms

fig. S8
Changes in Δz^2 across lines

fig. S9
Gene expression patterns at the *Gli3* and *Nkx3-2* candidate intervals.

fig. S10.
Selection at the *Nkx3-2* locus.

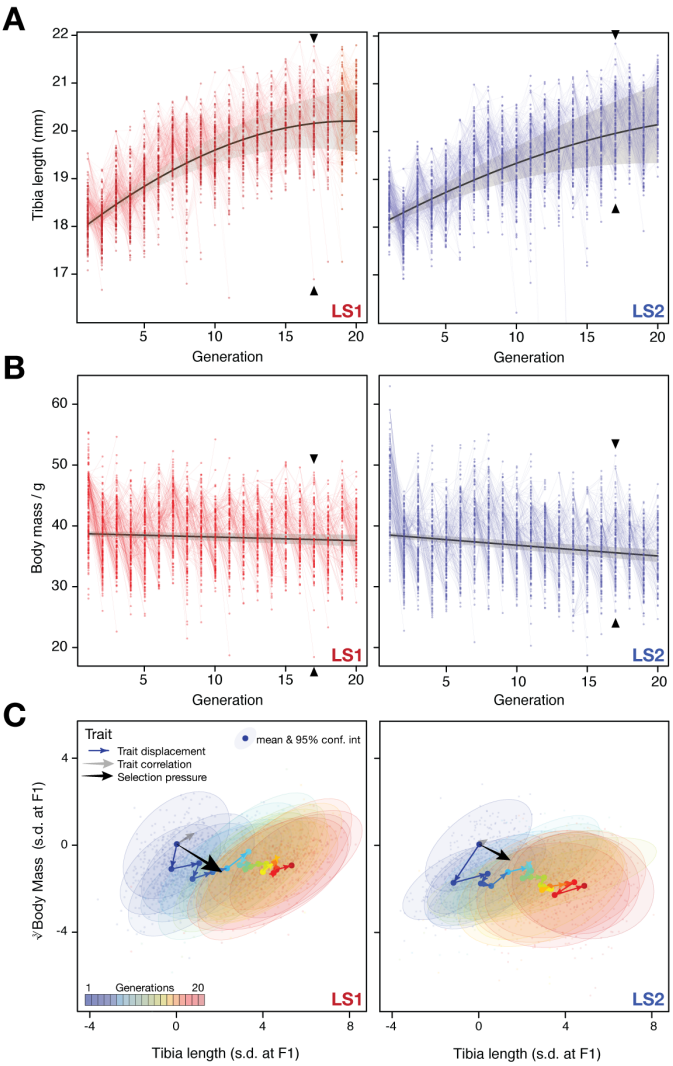


fig. S1. Artificial selection allowed detailed reconstruction of selection parameters. Rapid response to selection produced mice with progressively longer tibiae (**A**) and slightly lower body weight (**B**) within 20 generations. Having complete records throughout the selection experiment makes it possible to reconstruct the selection response for both phenotypes and genotypes in detail. Individuals varied in tibia length in both Longshanks lines (LS1, left; LS2, right). Lines connect parents to their offspring. The actual selection depended on the within-family and within-sex rank order of the tibia length-to-body mass (cube root) ratio (see (*8*) for details). The overall selection response was immediate and rapid for tibia length (**A**), suggesting a selection response that depended on standing variation among the founders (black lines show the best fitting quadratic function, with shading indicating 95% confidence interval; adjusted R^2 = 0.61 for LS1; 0.43 for LS2). Strong selection response led to rapid increase in tibia length. In contrast, there was only minor decrease in body weight over the course of the experiment. (**C**) Trajectory in selection response shows decoupling of correlation between tibia length and body mass. Despite overall correlation between tibia length and body mass (grey arrow and major axes in confidence envelopes), cumulative trait displacement over the 20 generations (expressed in s. d. units at F1; arrows, dots and 95% confidence envelopes, color-coded according to generation) showed persistent increase in tibia length with only minor change in body mass along the general direction of selection pressure (black arrows from F1; vector length and directions based on logistic regression). This shows that the Longshanks selection experiment was successful in specifically selecting for increased tibia length while keeping relatively unchanged body mass.

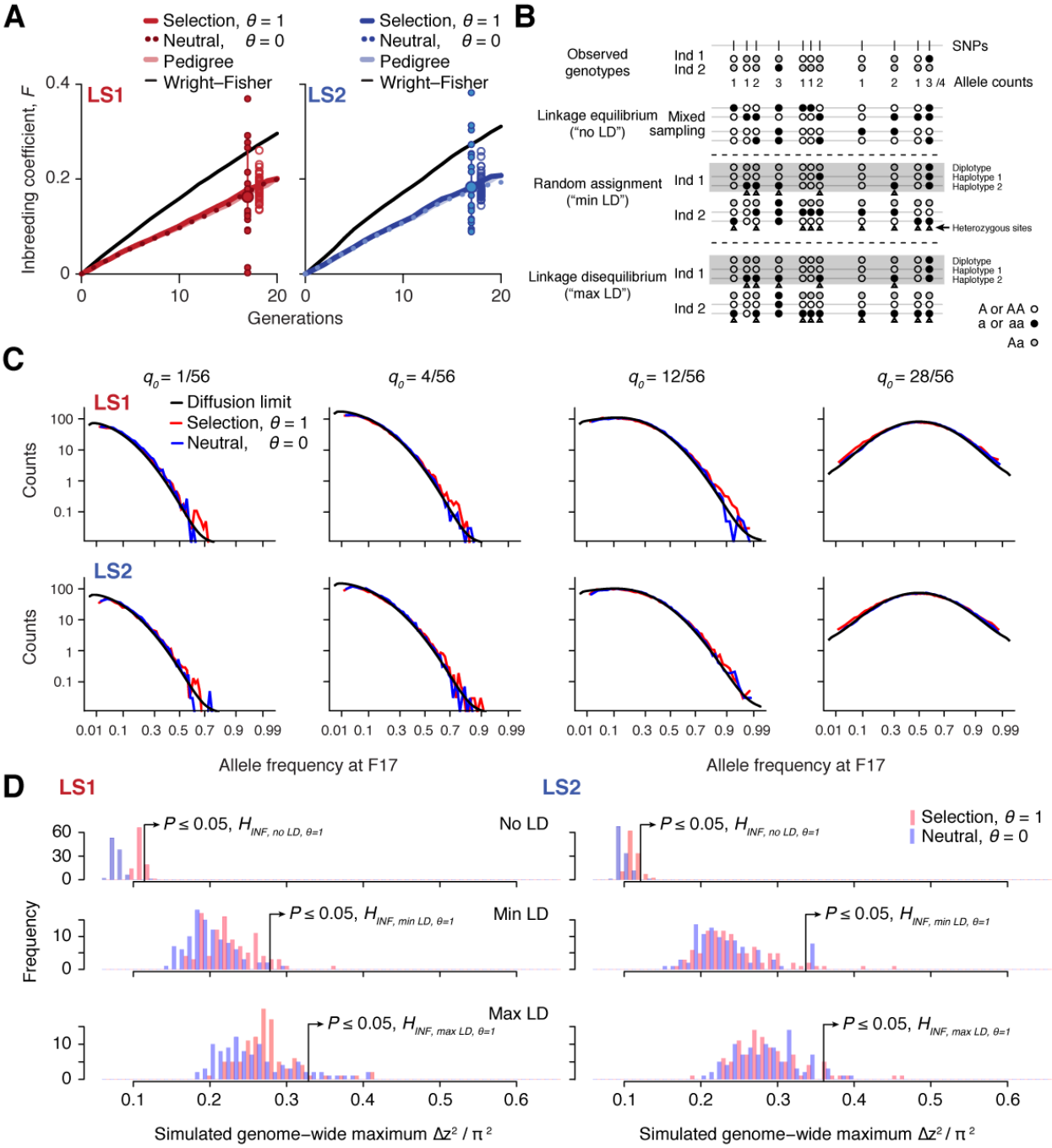


fig. S2. Simulating selection on pedigrees. (**A**) Increase in inbreeding over the course of the Longshanks experiment. The lines show the change in identity between two alleles between diploid individuals, F_b, over 20 generations, as calculated from the pedigree (light shade); an average of 50 neutral simulations without selection (dotted line); or the average of 50 simulation replicates with selection intensity at $V_s/V_e = 0.584$ ($\theta = 1$; thick, dark line). While the F_b trajectories based on pedigree or neutral simulations are indistinguishable, inbreeding increases slightly faster under selection (thick line). The black line shows the increase in identity expected under a Wright–Fisher model with the actual population sizes; under this model, F_w and F_b are close to each other, and to $1-(1-\frac{1}{2N_e})^t$, with N_e equal to the harmonic mean, 24.8. The large dot (with error bar showing the interquartile range among chromosomes) at right show the actual F_b, estimated from the decline in average $2p(1-p)$ over 17 generations. Small filled dots show the estimates from each of the 20 chromosome. Open dots show 40

replicate simulations, made with the same pedigree and the same selection response $\theta = 1$ and sub-sampling from the simulated chromosome according to the actual map length of each of the mouse chromosomes (*45*). The simulation agrees well with the observed genome-wide average. Most of the observed data from chromosomes fall within a range comparable to simulated replicates (compare large dot with open dots), with LD being the likely source of this excess variance. (**B**) Three different schemes to seed founder haplotypes. We simulate founder haplotypes that are consistent with observed genotypes (shown here as black, white and grey dots as the two homozygous and the heterozygous states) by directly sampling from founder individuals in each LS line. Under the linkage equilibrium scheme, we sample from the list of allele counts at all SNPs. This produces founder haplotypes that carry no linkage disequilibrium ("no LD"). Under the random assignment scheme, we sample according to each individual (shown as "diplotypes" within the box for easy comparison). At heterozygote sites in each individual (arrowheads), we randomly assign the alleles to the two haplotypes. This produces founder haplotypes that show minimal LD that is consistent with the observed genotypes ("min LD"). Under the "max LD" assignment scheme, we also sample according to each individual, except that we consistently assign its haplotypes 1 and 2 with reference (white) and alternate (black) alleles, respectively. This maximizes LD in the founder haplotypes ("max LD"). (**C**) Simulated vs. expected allele frequency shifts. The distribution of allele frequencies at generation 17 is compared with the distribution expected with no selection (blue) or with selection (red), given a frequency of 1, 4, 12 or 28 minor alleles out of 56 founding alleles. The black line shows the diffusion limit, calculated for scaled time $\frac{17}{N_e}$, with N_e estimated to be 51.7 and 48 in LS1 and LS2 respectively, from the rate of increase in F, calculated from the pedigree in panel **A** above. (**D**) Significance threshold values under varying LD from 100 simulated replicates (blue: no selection; red: observed selection response in the actual experiment, $\theta = 1$; see panel **B** on LD assignment methods). In order to account for non-independence of adjacent windows due to linkage, a distribution of genome-wide *maximum* Δz^2 was used to determine the significance threshold at each LD level. As seen in previous panels, increasing selection pressure does produce greater shifts in Δz^2 despite using the same pedigree due to a relatively greater proportion of additive genetic variance V_s. However, a far greater impact on Δz^2 is due to changes in LD. This is because weak associations between large numbers of SNP can greatly inflate the variance of Δz^2. Of the three LD levels, "max LD" likely produced overly conservative thresholds, whereas "min LD" may lead to higher false positives. We have opted conservatively to use maximal LD in our analysis.

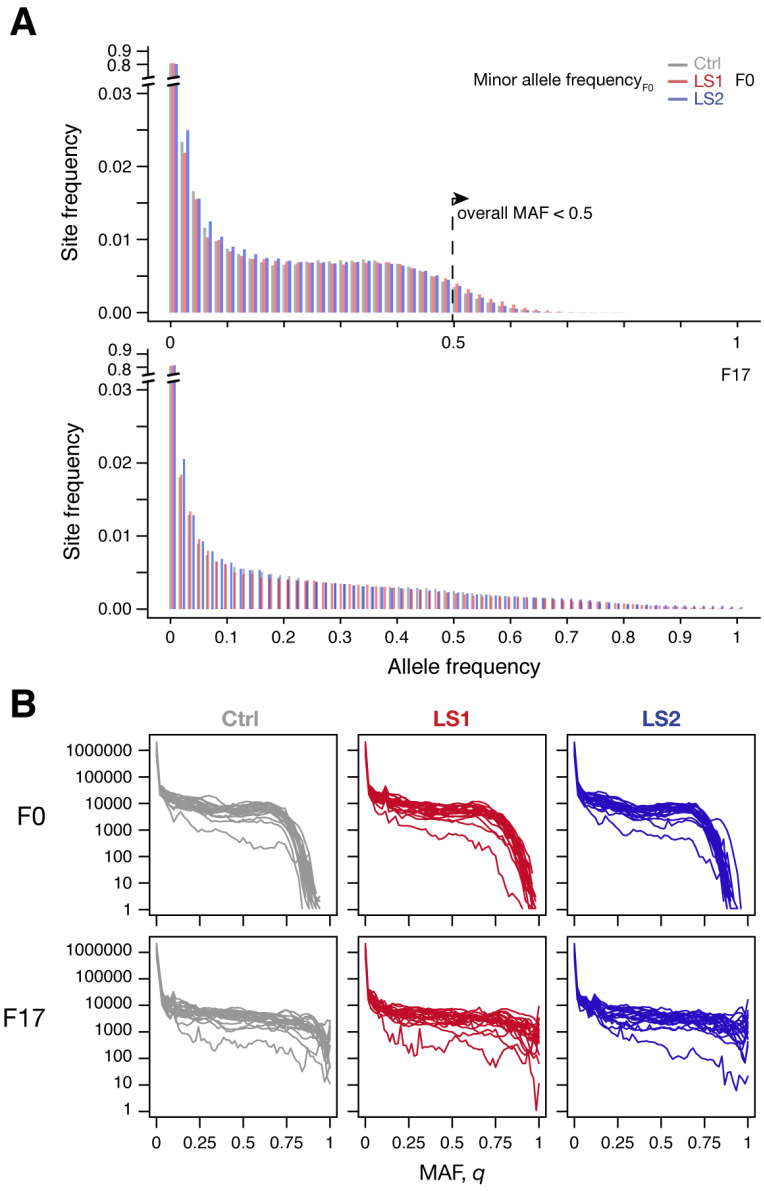


fig. S3. Broad similarity in molecular diversity in the founder populations for the Longshanks lines and the Control line. (**A**) Shown are the site frequency spectra from LS1, LS2 and Control lines at F0 (top, folded, based on a global minor allele frequency or MAF ≤ 0.5) and F17 (bottom, unfolded, but tracking the same minor allele as in F0). Overall the spectra were very similar to each other. The Control population was mostly intermediate in the decay in the rarer alleles. A small number of sites show MAF > 0.5 in each line, even though the overall MAF is ≤ 0.5. After 17 generations, the same alleles were generally more spread out, leading to more broadly distributed spectra. There was again little overall change between the Longshanks and Control lines. Due to differing number of available haplotypes, the F17 bars may appear to change in their grouping. (**B**) Variations between chromosomes shown in each line and generation. The unfolded site frequency spectrum is shown based on the MAF assigned as in **A**. There is substantial variation between chromosomes, which shows increased distortions in F17.

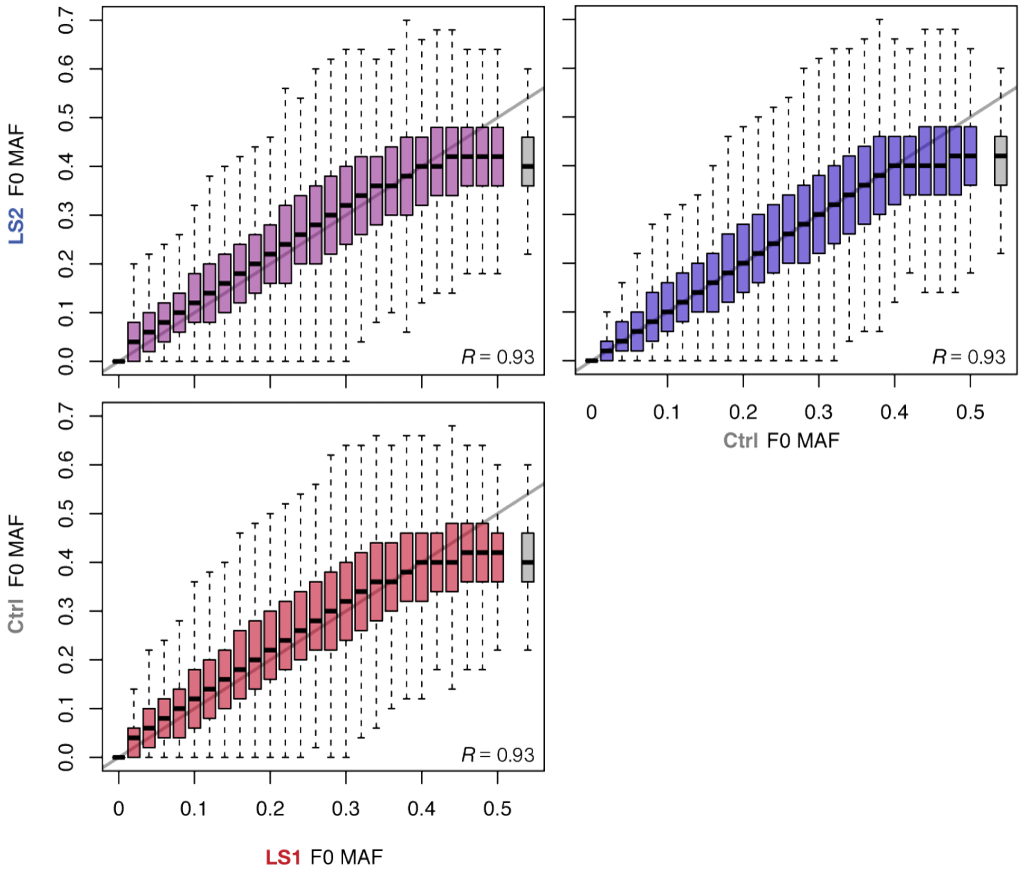


fig. S4. Allele frequencies between the founder populations were very similar. Joint minor allele frequencies shown as box plots in 2% bands between the Control and LS1 (red), LS2 (blue); or the two LS lines (purple). Outliers were omitted for clarity. The overall trends follow closely the parity line (grey line along the diagonal), except at frequencies very close to 0.5. Similar to the site frequency spectra shown in fig. S3A, a small number of sites have a MAF above 0.5 (grey box), because of the use of an overall MAF ≤ 0.5 to determine minor allele status to enable comparisons across lines. Correlations between all pairwise combinations were around 0.93.

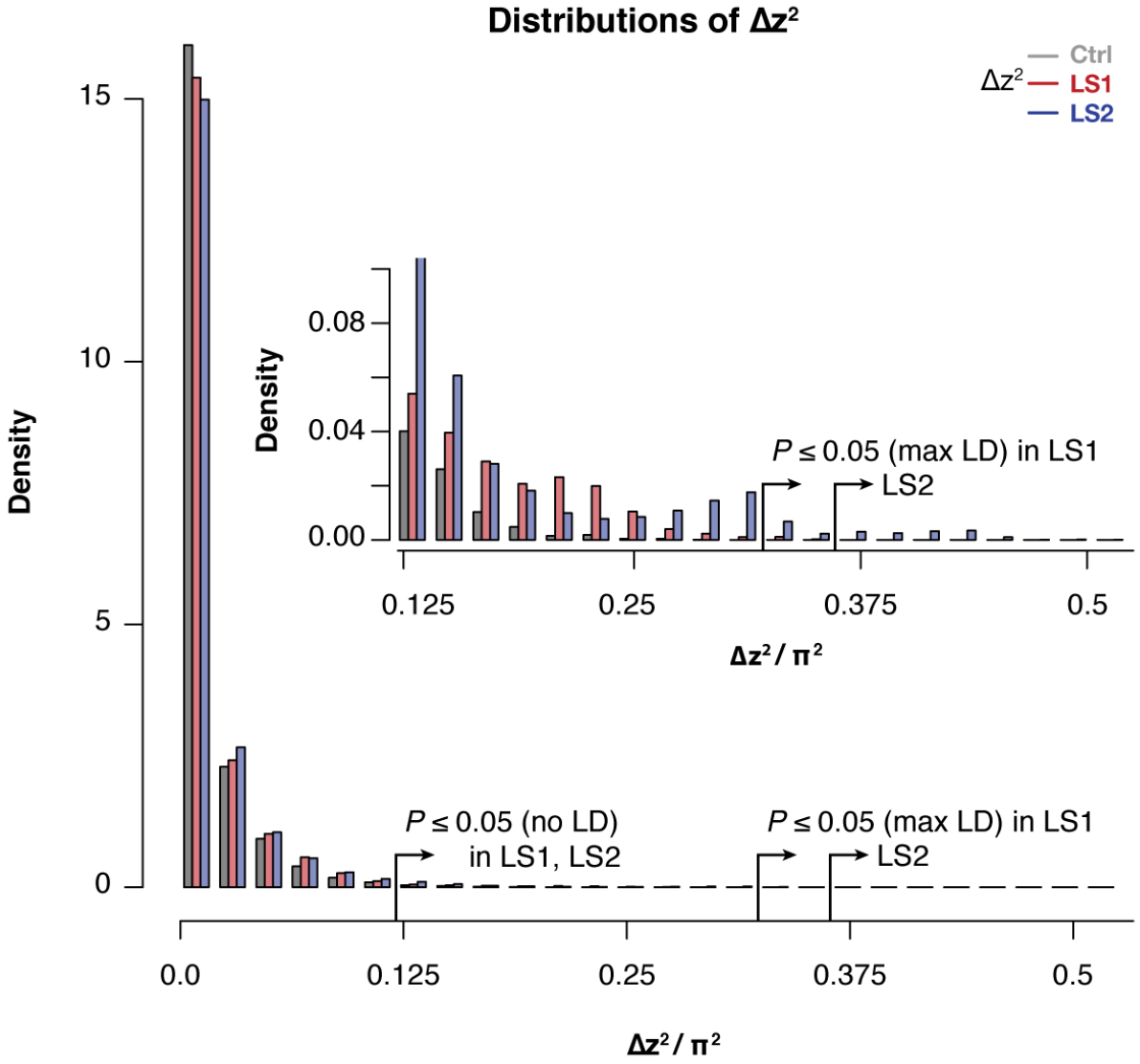


figure S5. Selected lines showed more extreme values of Δz^2 than the Control line. Histogram of within-line Δz^2 values in 10 kbp windows across the genome in the LS1, LS2 and Control. Overall similarity is high across all 3 lines, but there was an excess of large Δz^2 value starting from as low as < 0.1 π^2. This pattern becomes clearly distinct above the threshold value of 0.125 , which corresponds to the lenient significance threshold $P \leq 0.05$ under $H_{INF,\ no\ LD}$ (inset). There were clearly an excess of windows in LS2 above the more stringent $P \leq 0.05$ threshold under $H_{INF,\ max\ LD}$. Such excess supports discrete loci contributing to selection response in LS2 that give rise to greater distortion of Δz^2 spectra.

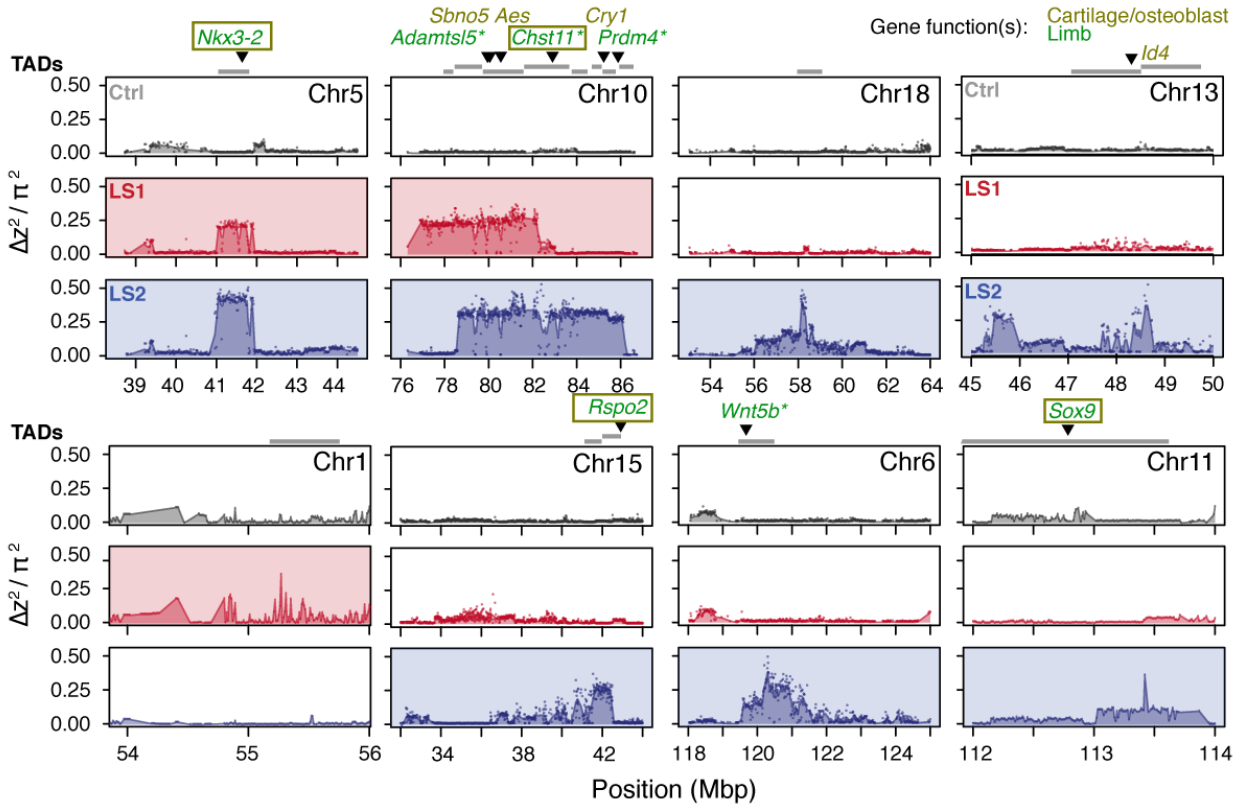


fig. S6. Detailed Δz^2 profiles at the LS significant loci. For each significant locus, Δz^2 profiles are shown for Ctrl (grey), LS1 (red) and LS2 (blue). Plots are shaded if the locus is significant in a given line. TADs within 250 kbp of the significant signals are shown as grey bars above each locus, with gene whose knockout phenotypes of the following categories highlighted: “abnormal tibia morphology”, “short limb”, “short tibia”, “abnormal cartilage morphology”, “abnormal osteoblast morphology”. The gene symbols are colored according to the gene function(s) in limb development (green), in bone development (yellow) or both (boxed). Gene symbols marked by asterisk (*) have reported “short tibia” or “short limb” knockout phenotypes. All of the above categories show significant enrichment at the 8 loci (number of genes per category: 4–7, nominal $P \leq 0.03$, see Supplementary Methods, section *Candidate genes* for details on the permutation), except “abnormal cartilage morphology”, with 4 genes and a nominal *P*-value of 0.083. No overlap was found with any gene in these categories from the three significant loci from the Ctrl line.

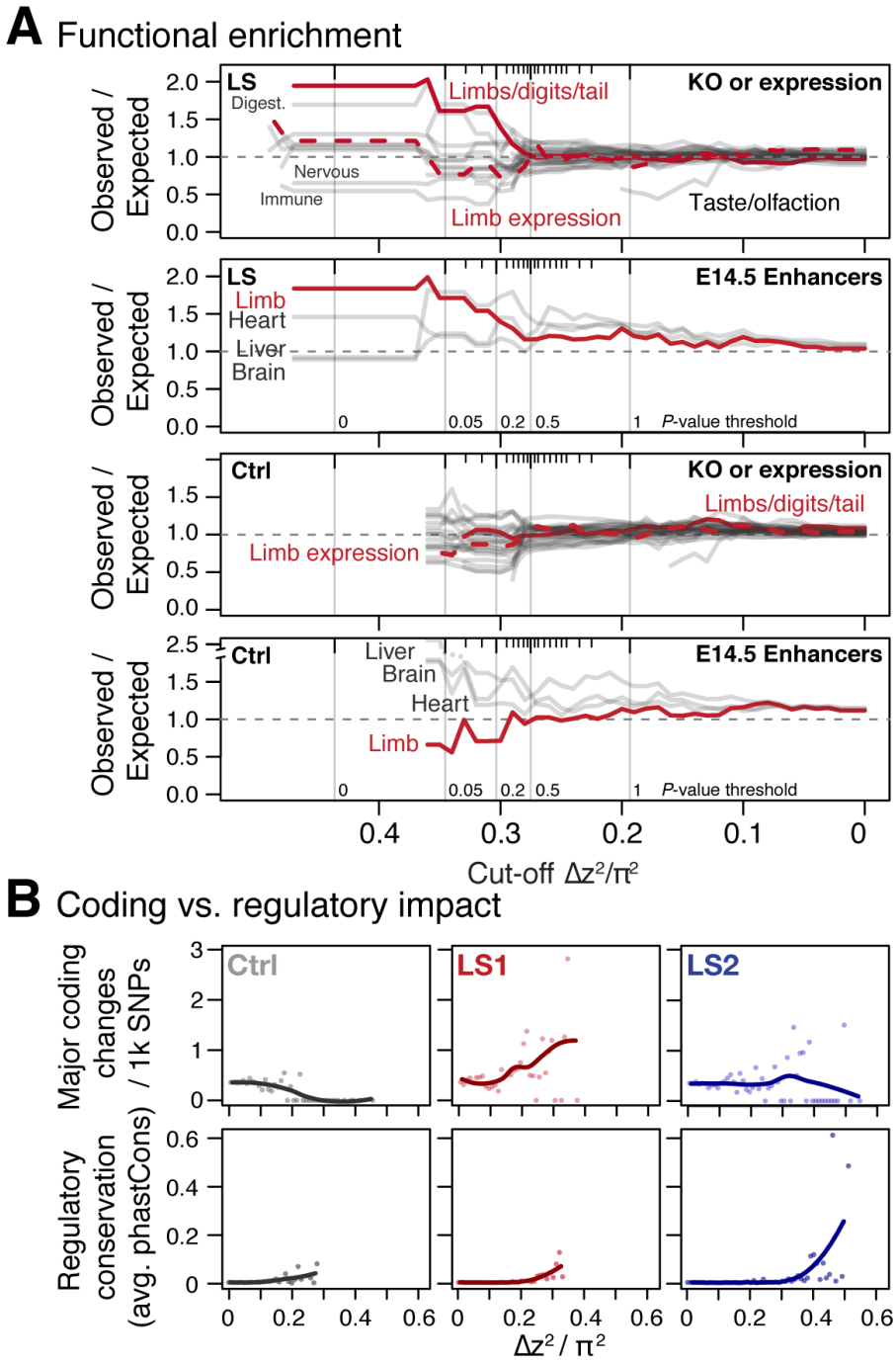


fig. S7. Loci associated with selection response in Longshanks lines show enrichment for limb function likely associated with *cis*-acting mechanisms. (**A**) Gene set enrichment analysis of knock-out phenotypes (KO) showed that selection response (here shown as Δz^2 cut-off values, see Supplementary Methods for details on cut-off values and inclusion criteria) were found among TADs containing limb and tail developmental genes (red solid lines) or genes with limb expression (red dotted lines) in LS lines (top) but not in Ctrl (bottom). Among KO phenotypes, limb defects show the greatest excess out of 28 phenotypic categories (other grey lines, with other extreme categories labeled, the "normal" category is excluded here). Among developmental enhancers for limb, heart, liver and brain tissue, we also observed an association with Δz^2 peaks in LS lines (top) for limb but not in Ctrl lines (bottom). The simulated significance thresholds based on $H_{INF,\ max\ LD}$ are also shown for reference (vertical

grey lines). The data from the LS line suggest that enrichment start to increase around $P \leq 0.5$ threshold and remained largely stable by $P \leq 0.05$, corresponding a cut-off of around 0.33 π^2. (**B**) Coding vs. regulatory impact. Frequency of moderate to major coding changes (top panels, amino acid changes, frame-shifts or stop codons), or average conservation score of regulatory sequences immediately flanking SNPs (based on conservation among 60 eutherian mammals; bottom panels) were used as proxies to estimate the functional impact of coding and regulatory mutations, respectively. In LS1, major coding changes became more common at high Δz^2 ranges; however the rate of SNPs with potentially major phenotype consequences did not increase in LS2 and in fact seems to decrease in Ctrl. In contrast, regulatory changes showed increased conservation associated with greater allele frequency shifts or Δz^2 in all three lines, except that SNPs with large shifts and strong conservation were more abundant in LS1 and LS2. Trend lines are shown with LOESS regression but statistical comparisons were performed using linear regressions.

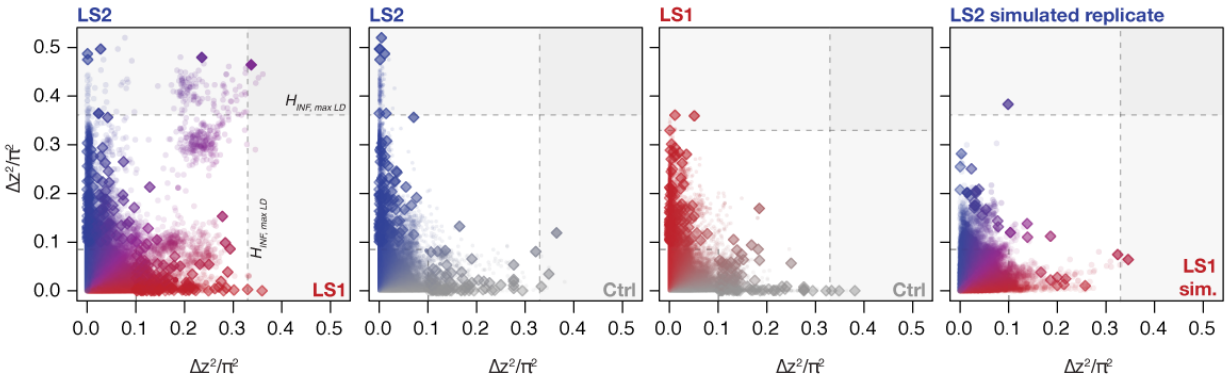


fig. S8. Changes in Δz^2 across lines. Shown are changes in Δz^2 in individual 10 kbp windows (all windows: circles; peak windows: diamonds). Generally there were no clear differences in Δz^2 along the axes except a slight increase in LS2. When taken as a joint LS1–LS2 comparison, however, we observed that many windows show shifts in both LS1 and LS2 (left panel). In contrast, only very few windows show parallelism in Ctrl–LS2 and Ctrl–LS1 comparisons (middle two panels). The right panel shows a single selected simulated replicate (selection pressure $\frac{V_s}{V_e} = 0.58$; maximum LD) found to have among the greatest extent of parallel Δz^2 among the replicates. The excess in parallel loci in observed results is clear both among the significant loci at $P \leq 0.05$ under $H_{INF,\ max\ LD}$ and highly significant at the more relaxed $H_{INF,\ no\ LD}$ threshold.

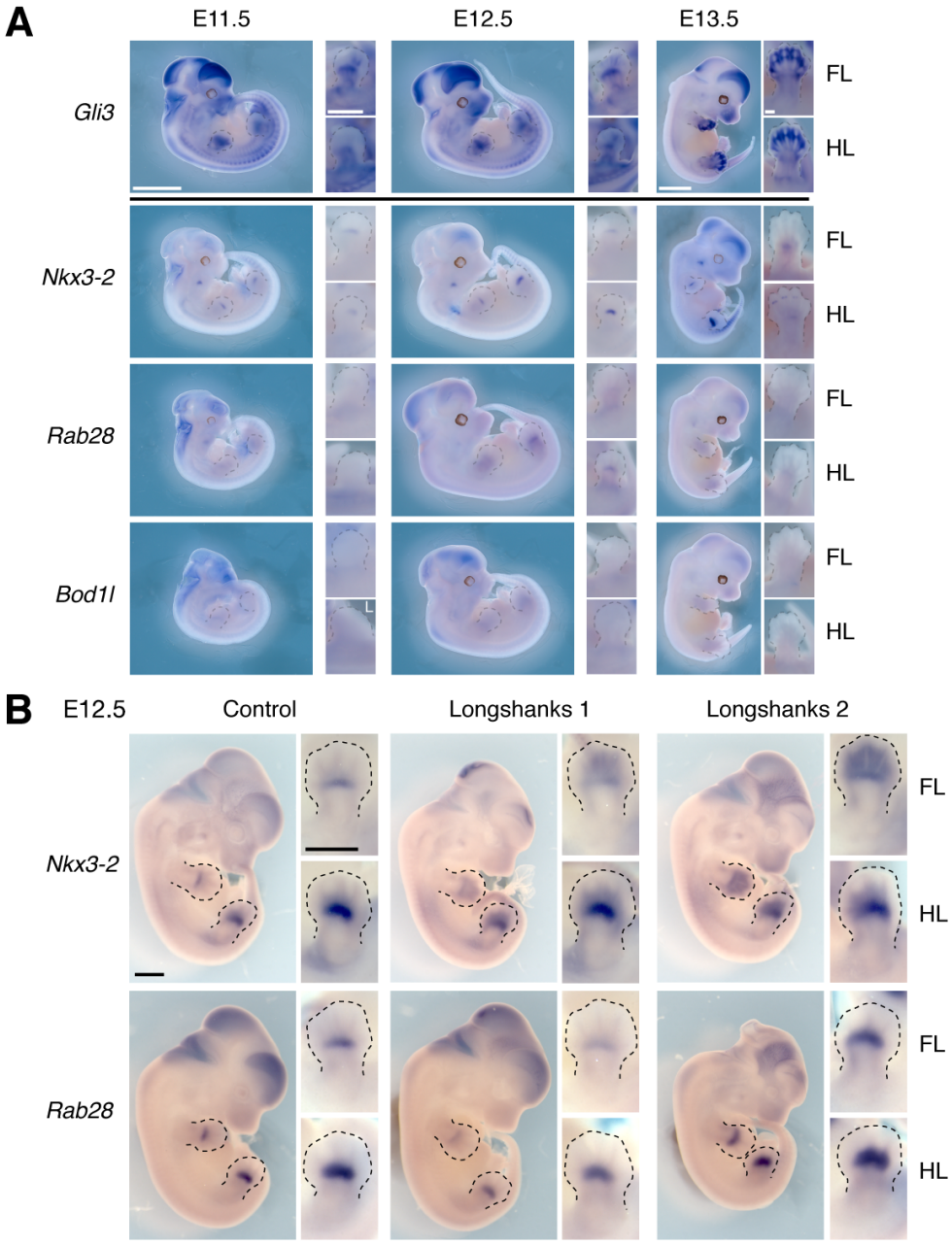


fig. S9. Gene expression patterns at the *Gli3* and *Nkx3-2* candidate intervals. (**A**) *Gli3* expression was determined using *in situ* hybridization. *Gli3* was robustly expressed during limb development in both developing fore- and hindlimb buds, especially in the autopod (hand/foot plate). Lower panel shows expression of *Nkx3-2* and its neighboring genes *Rab28* and *Bod1l*. The stronger expression of *Nkx3-2* in the developing limb buds as well as the known role of *Nkx3-2* in bone maturation (*32*) strongly argues for *Nkx3-2* being the gene underlying the selection response at the Chr5 locus. Scale bars: 1 mm for whole-mounts; 0.5 mm for limb buds. "L" indicates left side. (**B**) We collected E12.5 embryos from each line and performed *in situ* hybridization to determine the sites

and level of expression of *Nkx3-2* and *Rab28* in the Longshanks (right columns) and Control (left column) lines. Both genes are expressed in similar sites overall and specifically in the developing fore- and hindlimb buds in the region of the presumptive zeugopods. These data indicate common sites of expression and rule out qualitative presence/absence differences in expression. (**C**) Although the N1 enhancer pattern appear to differ from endogenous *Nkx3-2* expression, it matches the pattern of *Nkx3-2* expression, as indicated in (*32*). The use of a *Nkx3-2 Cre*-driver line suggested possibly undetected early expression of *Nkx3-2* prior to bone formation in the limb buds (lineage tracing experiment using a *Cre*-driver and revealed through crossing to *Rosa26R*, a *lacZ* reporter line). Image modified from (*32*), reused with permission. Scale bar: 1mm. h, humerus; r, radius; u, ulna; ti, tibia; fi, fibula; sty, stylopod; z, zeugopod; a, autopod; scp, scapula; mt, metacarpals; d, digits.

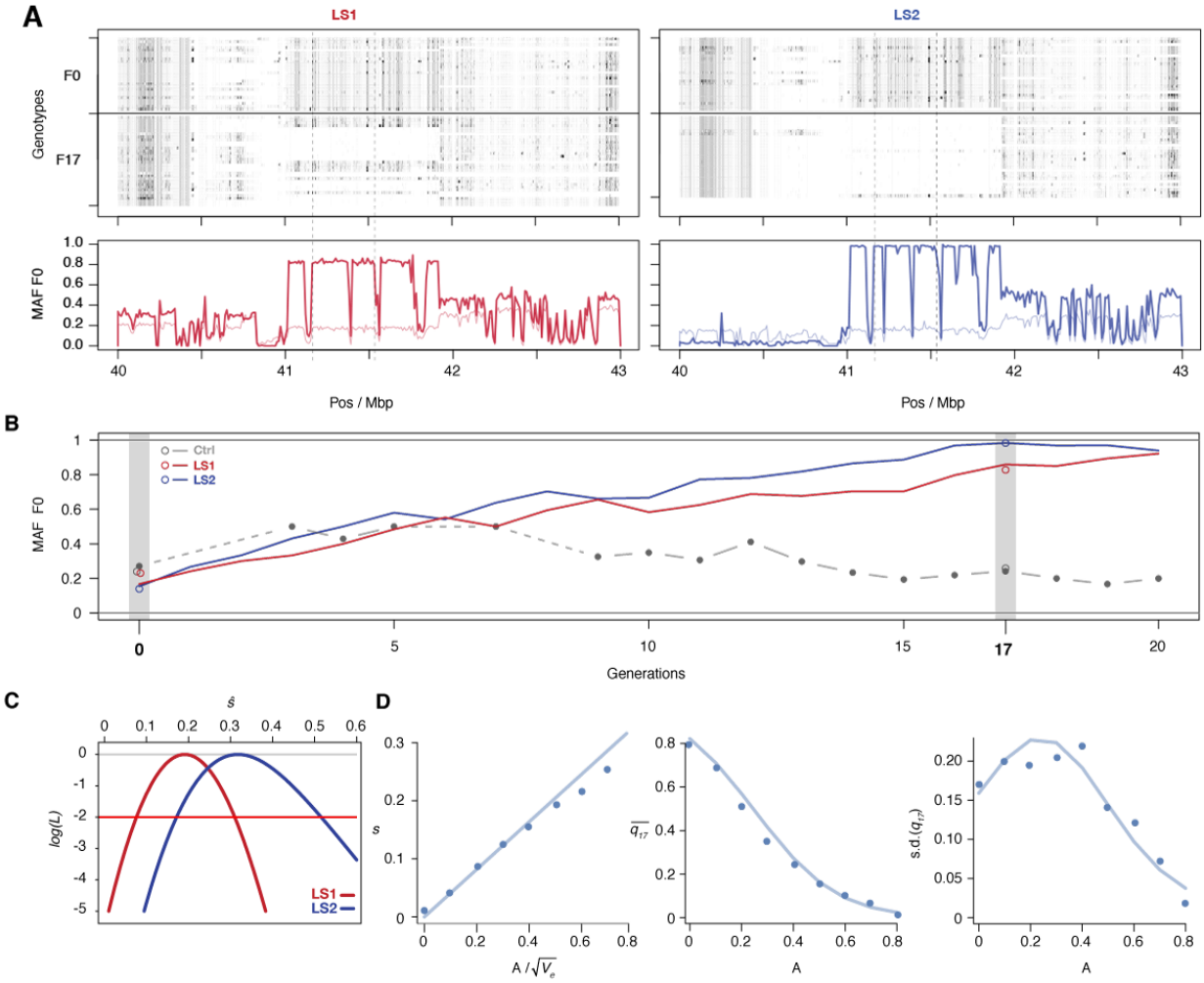


fig. S10. Selection at the *Nkx3-2* locus. (**A**) Raw genotypes from the F0 and F17 generations from LS1 (left) and LS2 (right) are shown, clearly indicating the area under the selective sweep. The genotype classes are shown as C57BL/6J homozygous (BL6, light grey), heterozygous (black) and alternate homozygous (dark grey). Lower Panel: Tracking MAF from both lines show that the originally rare F0 allele (thin line) rose to high frequency at F17 (thick lines). The plateau profile from both lines suggested that an originally rare allele became very common by F17 in both lines (see raw genotypes). Note that in LS2 F17 the region is fixed for the BL6 allele except the bottom-most individual). (**B**) Allele frequency of the selected allele (minor F0 allele) at N3 over 20 generations (red: LS1; blue: LS2; grey broken line: Ctrl; results from N1 were basically identical). Actual frequencies from genotyped samples in the Ctrl line are marked with filled circles. Dashed lines indicate missing Ctrl generations. Open circles at generations 0 and 17 (shaded) indicate allele frequencies from whole genome sequencing. The allele frequency fluctuated in Ctrl without major changes, whereas there was a generally linear increase from around 0.2 to 0.86 (LS1) and 0.98 (LS2) by generation 17. (**C**) The log likelihood of the selection coefficient, *s*, for LS1, LS2 (red, blue), based on transition probabilities for a Wright–Fisher population with the appropriate N_e. The horizontal red line shows a loss of log likelihood of 2 units, which sets conventional 2-unit support limits. (**D**) Simulations of an additive allele with effect *A* on the trait; 40 replicates for each value of *A*. Left: The selection coefficient, estimated from the change in mean allele frequency, as a function of $A/\sqrt{V_e}$; the line shows the least-squares fit $s = 0.41A/\sqrt{V_e}$. Middle: dots show the mean allele frequency at generation 17; the line shows the prediction from the single-locus Wright–Fisher model, given $s = 0.41A/\sqrt{V_e}$. Right: the same, but for the standard deviation of allele frequency.

Supplementary Tables

Table S1.
Sequencing summary.

Table S2.
Pairwise F_{ST} and segregating sites (S) between populations.

Table S3.
Full details on the eight discrete loci.

Table S4.
SNPs found within the *Nkx3-2* enhancers N1–N3 and their expected impact

Table S5.
Detected protein-coding changes with large allele frequency shift in amino acids.

Table S6.
Oligonucleotides for *in situ* hybridization probes

Table S7.
Oligonucleotides for multiplexed 4C-seq of enhancer viewpoints at the *Nkx3-2* locus.

Table S8.
Oligonucleotides for amplifying the enhancers at the *Nkx3-2* and *Gli3* loci.

Table S9.
Oligonucleotide primers for allele-specific genotyping of the N3 enhancer.

Pool	n	Total reads	Mapped Sequence (Gbp)	Fold-coverage (x)	Median cov./Sample (x)
Ctrl, F0	25	1,856,046,931	251.3	92.0	2.42
LS1, F0	26	1,858,953,260	256.1	93.8	2.82
LS2, F0	25	2,011,646,609	283.2	103.7	3.57
Ctrl, F17	32	1,882,838,451	260.2	95.3	2.97
LS1, F17	32	2,071,122,164	292.1	107.0	2.95
LS2, F17	31	1,897,174,855	267.0	97.8	2.93
Total	**169**	**11,577,782,270**	**1609.9**	**589.5**	**2.91**

Table S1. Sequencing Summary. For each line and generation, we individually barcoded all available individuals and pooled for sequencing. We aimed for a sequencing depth of around 100x coverage for 50–64 haplotypes per sample. Since the CD-1 mice were founded by an original import of 7 females and 2 males, we expect a maximum of 18 segregating haplotypes at any given locus. This sequencing design should give sufficient coverage to recover haplotypes genome-wide. Our successful genome-wide imputation results validated this strategy.

S \ F_{ST}	Ctrl F0	LS1 F0	LS2 F0	Ctrl F17	LS1 F17	LS2 F17
Ctrl F0	7,218,921	0.00095	0.00176	0.06067	0.09601	0.10694
LS1 F0	7,213,186	6,612,653	0.00100	0.07731	0.08045	0.10830
LS2 F0	7,306,904	7,282,910	7,239,774	0.07535	0.09505	0.09634
Ctrl F17	7,101,376	7,118,072	7,227,135	5,847,708	0.17469	0.17942
LS1 F17	7,121,979	7,062,183	7,214,508	6,795,924	5,813,162	0.17631
LS2 F17	7,218,921	7,155,580	7,239,774	6,832,220	6,778,532	5,749,742

Table S2. Pairwise F_{ST} and segregating sites (S) between populations. As expected, there is a general trend of decrease in diversity after 17 generations of breeding. Globally, there was a 13% decrease in diversity, but F17 populations still retained on average ~5.8M segregating SNPs (diagonal). There was very little population differentiation, as indicated by low F_{ST} among the three founder populations, however F_{ST} increases by 100-fold among lines by generation F17 (above diagonal, orange boxes). Within-line F_{ST} is intermediate in this respect, increasing by about half that amount.

Rk	Chr	Span (Mbp)	Peak	Genes	Core span	TAD span	Genes	*s*	Type	Candidate genes*	All genes
1	**5**	**38.95–45.13**	**41.77**	**14**	**41.77–41.78**	**41.05–41.81**	***3***	**0.26**	**Parallel**	***Nkx3-2***	Bod1l, Bst1, C1qtnf7, Cc2d2a, Cd38, Cpeb2, Fbxl5, Fgfbp1, Hs3st1, Ldb2, ***Nkx3-2†***, Prom1, Rab28, Tapt1
2	**10**	**77.47–87.69**	**81.07**	**190**	**78.30–85.09**	**77.98–86.58**	***175***	**0.21**	**Parallel**	***Sbno5, Aes, Adamtsl5*, Chst11*, Cry1, Prdm4****	Abca7, Adamtsl5*, Adat3, ***Aes*†***, Agpat3, Aire, Aldh1l2, Amh, Ankrd24, Ap3d1, Apba3, Apc2, Appl2, ***Arid3a***, Ascl1, Ascl4, Atcay, Atp5d, Atp8b3, Bpil2, Bsg, Btbd11, Btbd2, C2cd4c, Casp14, Ccdc105, Cdc34, Celf5, Cfd, ***Chst11*†***, Cirbp, Ckap4, Cnn2, Creb3l3, ***Cry1†***, Csnk1g2, Cstb, D10Jhu81e, D10Wsu102e, D10Wsu52e, Dapk3, Dazap1, Diras1, Dnmt3l, Dohh, Dos, ***Dot1l***, Eef2, Efna2, Eid3, Elane, Fam108a, Fbxo7, Fgf22, Fhl4, Fstl3, Fzr1, Gadd45b, Gamt, Gipc3, Glt8d2, ***Gna11***, Gna15, Gng7, Gpx4, Grin3b, Gzmm, Hcfc2, Hcn2, Hmg20b, Hmha1, Hsp90b1, Icosl, Ilvbl, Itgb1bp3, Itgb2, Izumo4, Jsrp1, Kiss1r, Klf16, Krtap10-10, Krtap10-4, Krtap12-1, Lingo3, Lmnb2, LOC16697, Lrrc3, Lsm7, Madcam1, Map2k2, Matk, Mbd3, Med16, Mex3d, Midn, Mier2, Mknk2, Mobkl2a, Mrpl54, Mterfd3, Mum1, Ncln, Ndufs7, Nfic, Nfyb, Nt5dc3, Nuak1, Oaz1, Odf3l2, Olfr1351, Olfr1352, Olfr1353, Olfr1354, Olfr1355, Olfr1356, Olfr1357, Olfr57, Olfr8, Onecut3, ORF61, Pah, Palm, Pcsk4, Pdxk, Pfkl, Pias4, Pip5k1c, Plekhj1, Plk5, Polr2e, Polr3b, Polrmt, Ppap2c, ***Prdm4****, Prssl1, Prtn3, Ptbp1, Pttg1ip, Pwp1, Pwp2, Reep6, Rexo1, Rfx4, Ric8b, Rnf126, Rps15, Rrp1, S1pr4, ***Sbno2†***, Scamp4, Sf3a2, Sgta, Shc2, Sirt6, Slc1a6, Slc39a3, Slc41a2, Snord37, Stab2, Stk11, Sumo3, Syde1, Syn3, Tbxa2r, Tcf3, Tcp11l2, ***Tdg***, Theg, Thop1, Timm13, Timp3, Tjp3, Tle2, Tle6, Tmprss9, ***Trappc10***, Trpm2, Txnrd1, Ube2g2, Ugcr11,

										Vmn2r80, Vmn2r81, Vmn2r82, Vmn2r83, Wdr18, Zbtb7a, Zfp781, Zfp873, Zfp938, Zfr2
3	18	53.63–63.50	58.18	53	**58.18–58.19**	**57.96–59.08**	*4*	0.18	LS2-specific	-	*Ablim3, Adamts19, Adrb2, Afap1l1, Aldh7a1, Apcdd1, Arhgef37, Arsi, Camk2a, Cd74, Cdx1, Cep120, Chsy3,* ***Csf1r*†****, Csnk1a1, Csnk1g3, Ctxn3, Dctn4, Fam38b,* ***Fbn2****, Fbxo38, Gramd3, Grpel2, Hmgxb3, Htr4, Iigp1, Il17b, Isoc1, Lmnb1, March3, Megf10, Myoz3, Napg, Ndst1, Pcyox1l, Pde6a,* ***Pdgfrb****, Phax, Ppargc1b, Prrc1, Rbm22, Rps14, Sh3tc2, Slc12a2,* ***Slc26a2*†****, Slc27a6, Slc6a7, Spink10, Spink13, Spink7, Synpo, Tcof1†, Zfp608*
4	13	35.59–55.21	48.65	112	**48.65–48.66**	**47.06–49.74**	*22*	0.18	LS2-specific	*Id4*†*	*Arl10,* ***Aspn****, Atxn1, Auh, Barx1, Bicd2, Bmp6, C78339, Cage1,* ***Cap2****, Ccdc90a, Cd83, Cdhr2, Cdyl, Cenpp, Cks2, Cltb, Cplx2, Dek, Diras2, Drd1a, Dsp, Dtnbp1, Ecm2,* ***Edn1†****, Eef1e1, Eif4e1b, Elovl2, F13a1, Faf2, Fam120a, Fars2, Fbxw17, Fgd3, Fgfr4, Gadd45g, Gcm2, Gcnt2, Gfod1, Gmpr, Gprin1, Higd2a, Hivep1, Hk3, Hrh2, Iars,* ***Id4*†****, Ippk, Jarid2, Kdm1b, Kif13a, Ly86, Lyrm4, Mak,* ***Msx2*†****, Muted, Mylip, Nedd9, Nfil3, Nhlrc1, Ninj1, Nol7, Nol8, Nop16, Nrn1, Nsd1, Nup153, Nxnl2, Ofcc1, Ogn, Omd, Pak1ip1, Phactr1, Phf2, Ppp1r3g, Ptpdc1, Ranbp9, Rbm24, Riok1, Rnf144b, Rnf182, Rnf44,* ***Ror2*†****, Rpp40, Rreb1, S1pr3,* ***Secisbp2****, Sema4d, Sfxn1, Shc3, Sirt5, Slc35b3, Sncb, Snrnp48, Spin1, Sptlc1, Ssr1, Susd3, Sykb, Tbc1d7, Tcfap2a, Thoc3, Tmem14c, Tmem170b, Tpmt, Tspan17, Txndc5, Uimc1, Unc5a, Wnk2, Zfp169, Zfp346*
5	1	53.16–57.13	55.27	21	**58.27–58.28**	**57.96–59.08**	*4*	0.14	LS1-specific	-	*Ankrd44,Asnsd1,Boll,Ccdc150,Coq10b,Gtf3c3,Hecw2,Hsfy2,Hspd1,Hspe1,Mars2,Mobkl3,Ormdl1,Osgepl1,Pgap1,Plcl1,Pms1,Rftn2,****Satb2*****,Sf3b1,Stk17b*
6	15	31.92–44.43	41.54	57	**41.54–42.55**	**41.13–42.93**	*3*	0.14	LS2-specific	*Rspo2*†*	*Abra, Angpt1, Ankrd46, Atp6v1c1, Azin1, Baalc, Cox6c, Cthrc1, Dcaf13, Dpys, Eif3e, Eny2, Fbxo43, Fzd6, Grhl2, Hrsp12, Kcns2, Klf10, Laptm4b,*

									Lrp12, Matn2, Mtdh, Ncald, Nipal2, Nudcd1, Odf1, Osr2, Oxr1, Pabpc1, Pgcp, Polr2k, Pop1, Rgs22, Rims2, Rnf19a, Rpl30, Rrm2b, ***Rspo2*†****, Sdc2, Sema5a, Slc25a32, Snord123, Snx31, Spag1,* ***Stk3****, Tas2r119, Tm7sf4, Tmem74, Tnrc18, Trhr, Tspyl5, Ttc35, Ubr5, Vps13b, Ywhaz, Zfp706, Zfpm2*
7	6	118.65–125.25	120.30	108	**120.30–120.31**	**19.45–120.49**	*12*	0.13	LS2-specific	*Wnt5b**	*A2m, Acrbp, Adipor2, Aicda, Apobec1,* ***Atn1****, Atp6v1e1, B4galnt3, Bcl2l13, Bid, C1ra, C1rb,* ***C1rl****, C1s, C3ar1, Cacna1c, Cacna2d4, Ccdc77, Cd163, Cd27, Cd4, Cdca3, Cecr2, Cecr5, Cecr6, Chd4, Clec4a1,* ***Clec4a2†****, Clec4a3, Clec4a4, Clec4b1, Clec4b2, Clec4d, Clec4e, Clec4n, Clstn3, Cops7a, Dcp1b, Dppa3, Emg1, Eno2, Erc1, Fbxl14, Foxj2, Gapdh, Gdf3, Gnb3, Gpr162, Grcc10, Iffo1, Il17ra, Ing4, Iqsec3, Kdm5a, Klrg1, Lag3, Leprel2, Lpar5, Lpcat3, Lrrc23, Lrtm2, M6pr, Mfap5, Mical3, Mlf2, Mrpl51, Mug1, Mug2, Mug-ps1, Nanog, Nanogpd, Ncapd2, Necap1, Ninj2, Nop2, Pex26, Pex5, Phb2,* ***Phc1****, Ptms,* ***Ptpn6****, Rad52, Rimklb, Rnu7, Scarna10, Slc25a18, Slc2a3, Slc6a12, Slc6a13, Spsb2, Tapbpl, Tpi1, Tuba8, Usp18, Usp5, Vamp1, Vmn2r19, Vmn2r20, Vmn2r21, Vmn2r22, Vmn2r23, Vmn2r24, Vmn2r25, Vmn2r26, Vmn2r27, Wnk1,* ***Wnt5b****, Zfp384*
8	11	111.10–115.06	113.42	16	**113.42–113.43**	**11.54–113.62**	*2*	0.12	LS2-specific	*Sox9**	*Btbd17, Cdc42ep4, Cog1, Cpsf4l, D11Wsu47e, D11Wsu99e, Dnaic2, Gpr142, Gprc5c, Kif19a,* ***Rpl38****, Sdk2, Slc39a11,* ***Sox9*†****, Sstr2, Ttyh2*

***** Genes with short limb, short tibia or abnormal tibia knockout phenotypes

† Genes with cartilage or osteoblast morphology knockout phenotypes

Gene Genes within the TAD span

Table S3. Full details on the eight discrete loci. Listed here are the eight loci shown in Table 1, with additional details on the core span and the TAD span used to identify candidate genes, and a full list of genes within the full span.

			LS Alleles		LS selected allele frequency						Impact of selected alleles		
					Ctrl Gen		LS1 Gen		LS2 Gen				
Enh	**dbSNP ID**	**Position**	**Sel**	**Alt**	**F0**	**F17**	**F0**	**F17**	**F0**	**F17**	**TFBS**	**TF role**	**Impact**
	rs29527943	41,166,187	G	A	0.20	0.27	0.19	0.84	0.14	0.97			
N1	rs32745741	41,166,264	C	G	0.26	0.27	0.15	0.81	0.12	0.97	*HoxD12* loss	+	-
	rs32745739	41,166,505	A	G	0.32	0.24	0.13	0.88	0.12	0.98			
N2	rs48075126	41,392,871	A	G	0.26	0.13	0.17	0.88	0.18	0.82	*Zic1/2/3* gain, *Gli3* gain	-/(+)	-/(+)
	rs33320581	41,536,250	G	T	0.28	0.26	0.17	0.83	0.10	0.98	**2x *Nkx3-2* loss**	**++**	**----**
N3	rs33219710	41,536,431	G	A	0.24	0.26	0.23	0.83	0.14	0.98			
	rs33600994	41,536,498	C	A	0.22	0.19	0.23	0.86	0.16	0.98	*Nkx3-1* loss, *Tcf/Lef1* gain	+/~?	-

Enh: Enhancer
Gen: Generation
TFBS: transcription factor binding site
TF: transcription factor

Table S4. SNPs found within the *Nkx3-2* enhancers N1–N3 and their expected impact. The positions, alleles and frequencies of each SNP is indicated along with their allele frequencies at F0 and F17 generations. The selected allele is the F0 minor allele, which increased to near fixation by F17 in both LS1 and LS2. In contrast, there were only slight frequency change in the Control line. The functional impact of each of these SNPs were determined based on transcription factor binding affinity reported from the UNIPROBE database (*35*). For each predicted difference for transcription factors (TF) known to be involved in limb development, the role of each TF is noted and the net impact of the selected allele is indicated. For N2, while the gain of a *Gli3* binding site should produce a positive effect on enhancer activity on *Nkx3-2*, it is unclear in the literature if there is a positive feedback loop from *Gli3* to *Nkx3-2*. For the N3 enhancer, the canonical *Wnt* pathway has opposite effect on bone maturation and differentiation depending on the differentiation state of the cell. We therefore consider its impact ambiguous.

Gene symbol	Chrom	dbSNP ID	SNP position	SNP Freq. change		Amino acid change		Phast Cons	KO phenotype (limbs/digits/ tail)	KO lethal ?	KO overall assessment
				LS1	LS2	LS1	Dir.				
Traf3ip1	1	rs47229691	91,500,846	0.28	0.53	M143T	PUR	0.014	polydactyly, curly tail	**Y**	abnormal nervous system development, abnormal neural tube morphology, absent embryonic cilia, cardiac edema, **embryonic lethality** (complete penetrance), microphthalmia
		rs48526274	91,505,448	0.12	0.55	R313K	PUR	0.706			
		rs37605154	58,063,554	0.04	0.61	Q1426R	PUR	0.997			
		rs250429553	58,209,792	0.02	0.78	R5Q	DIV	1			
Cutal	2	rs226367576	34,882,365	0.36	0.51	I47S	PUR	0.996	polydactyly	N	None
Galnt3	2	rs27987860	66,084,173	0.36	0.00	E633D	PUR	1	abnormal femur morphology	N	Mice homozygous for a knock-out allele exhibit decreased circulating alkaline phosphatase, hypercalcemia, hyperphosphatemia, decreased circulating parathyroid hormone, and male specific postnatal growth retardation, infertility, and increase in bone density.
Col27a1	4	rs33038048	63,225,145	0.20	0.52	T357A	PUR	0.001	short limbs	**Y**	Mice homozygous for an in-frame deletion display **neonatal lethality**, respiratory failure, and severe chondrodysplasia.
		rs32148105	63,225,427	0.30	0.64	N451D	DIV	0			
		rs32529591	63,225,479	0.30	0.60	P468L	DIV	0.001			
		rs27905388	63,272,545	0.23	0.57	V816I	DIV	0.137			
		rs257793193	63,306,703	0.25	0.68	E1176D	DIV	1			
		rs31967053	63,317,333	0.29	0.63	D1316E	PUR	1			
Fuz	7	rs46684086	44,898,988	0.38	-0.42	L264M	DIV	0.637	abnormal apical ectodermal	**Y**	Mice homozygous for a gene trapped allele exhibit neural tube closure defects, abnormal craniofacial
		rs253122287	44,900,158	0.38	-0.42	D353E	PUR	0.669			
		rs243865195	44,900,281	0.38	-0.42	Q394H	DIV	0.987			

		rs31477222	44,900,404	0.38	-0.42	H385R	DIV	1	ridge morphology, polydactyly		morphology, abnormal skeletal morphology, polydactyly, anopthalmia, pulmonary hypoplasia, and cardiac outflow tract defects.
Acan	7	rs262365125	79,096,707	0.40	-0.11	S698R	DIV	0.002	abnormal limb morphology, brachydactyly, brachypodia, short humerus, short femur, abnormal limb development, short limbs, absent caudal vertebrae	**Y**	Spontaneous mutations in this gene lead to dwarfism, cartilage, skeletal and limb anomalies, craniofacial defects, hearing loss and **neonatal death** due to respiratory failure. Homozygotes for an ENU-induced allele show cardiomyopathy as well as cleft palate, disproportionate dwarfism and brachypodia.
		rs32400885	79,099,204	0.41	0.09	V1241A	PUR	0.402			
		rs31327907	79,099,207	0.41	0.09	N1242S	PUR	0.931			
		rs32439390	79,099,513	0.45	-0.13	S1344I	DIV	0.023			
		rs32085027	79,100,023	0.35	-0.02	E1514G	PUR	1			
		rs49012578	79,100,506	0.36	0.05	N1675I	PUR	0.361			
		rs48571987	79,100,667	0.35	0.1	I1729V	PUR	0			
		rs36376581	79,101,084	0.40	-0.09	Y1868H	DIV	0.306			
Kif7	7	rs211772880	79,699,167	0.46	0.04	N1220T	DIV	0.002	abnormal limb bud morphology, polydactyly, postaxial polydactyly, preaxial	**Y**	Mice homozygous for a knock-out allele exhibit **neonatal lethality**, exencephaly, polydactyly, abnormal sternum, edema, abnormal ribs, and abnormal neurogenesis. Mice homozygous for an ENU-induced allele exhibit **prenatal lethality**,
		rs31859391	79,699,507	0.42	0.06	W1137R	PUR	0.992			
		rs222106773	79,708,493	0.58	0.01	S570G	PUR	0.152			

									polydactyly, syndactyly		polydactyly, and abnormal neural tube development and neurogenesis.
Dot1l	10	rs51353227 rs234091350 rs46367861	80,783,290 80,786,131 80,789,444	0.78 0.78 0.63	0.89 0.89 0.69	V371A L756P P1175L	PUR PUR DIV	0.075 0.198 0.367	short tail	**Y**	Mice homozygous for a gene trap allele show **late embryonic lethality**. Mice homozygous for a null allele die by E10.5 displaying a growth arrest, abnormal yolk sac angiogenesis and heart dilation while mutant ES cells show elevated apoptosis, G2 cell cycle arrest, telomere elongation and aneuploidy.
Cap2	13	rs220246885	46,530,979	0.22	0.47	V42M	DIV	1	oligodactyly	**Y**	Mice homozygous for a knock-out allele are smaller, prone to eye infections and show microphthalmia, cardiac conduction defects and dilated cardiomyopathy, predominantly in males. Males are underrepresented at weaning and **~70% die suddenly by 12 weeks of age**, whereas females survive at nearly expected levels.
Syk	13	rs6363905	52,640,675	-0.01	0.51	C482S	PUR	0.986	abnormal foot pad morphology, decreased bone resorption	**Y**	Homozygous null mice have high rates of **postnatal lethality**, exhibit developmental defects of B cells, T cells and osteoclasts, and have defective dendritic cell cross-presentation of antigens from necrotic cells.
Myo10	15	rs13482486	25,732,087	-0.17	0.42	S350P	PUR	0.688	abnormal tail pigmentation, decreased tail pigmentation, abnormal digit morphology,	N	Mice homozygous for a spontaneous null allele display decreased body size, white spotting, decreased tail pigmentation, and syndactyly as well as various eye defects including iris coloboma, corneal opacity, and

									fused phalanges, syndactyly		persistence of hyaloid vascular system.
Fbn2	18	rs13464230	58,023,239	0.10	0.48	K2503E	PUR	0.769	abnormal interdigital cell death, abnormal autopod morphology, fused phalanges, syndactyly	N	Homozygotes for spontaneous, chemically-induced, and targeted null mutations show bilateral syndactyly with fusion of both soft and hard tissues. Deafness found in an X-ray induced allelic mutant is apparently due to the joint disruption of a linked gene.

Table S5. Detected protein-coding changes with large allele frequency shift in amino acids. Listed are genes carrying large frequency changing SNPs affecting amino acid residues. Highlighted cells indicate the line with greater frequency changes ≥ 0.34 (red text with shading). Other suggestive changes are also shown with red numbers in unshaded cells. The changed amino acids are marked using standard notations, with the directionality indicated as "purifying" or "diversifying" with respect to a 60-way protein sequence alignment with other placental mammals. The conservation score based on phastCons was calculated at the SNP position itself, ranging from 0 (no conservation) to 1 (complete conservation) among the 60 placental mammals. For each gene, reported knockout phenotypes of the "limbs/digits/tail" category is reported, along with whether lethality was reported in any of the alleles, excluding compound genotypes. A summary of the mutant phenotype as reported by the Mouse Genome Informatics database is also included to highlight any systemic defects beyond the "limbs/digits/tail" category (lethal phenotypes reported in **bold**).

Gene	Chr	Start	End	Size	Type	Primer (5' - 3')
Rab28	chr5	41,698,405	41,625,451	785	For	AGGTGGCAAGATGTTGGATAAATAC
					Rev	GATCATCAAAGCTTGGAGCAGC
Nkx3-2	chr5	41,763,877	41,762,039	579	For	GCGATCCTCAACAAGAAAGAGGA
					Rev	GCGCTTCTTTCGCGGTTTAG
Bod1l	Chr5	41,832,764	41,828,797	873	For	GATGCCATGTCAATCTTGGAAACC
					Rev	CACTGTGAGTTCGTCATCAGAATC
Gli3	chr5	15,720,149	15,724,574	640	For	CTCATGTTACCAAGAAGCAGCGT
					Rev	TGTCTCTCCTGTTGAGTGTGTTC

Table S6. Oligonucleotides for *in situ* hybridization probes.

VP	Chr	Start	End	Type	Primer (5' - 3')
N1	chr5	41,165,684	41,165,705	Biotin	/5Biosg/GAGTTATCTCTATGGGAGAAGT
	chr5	41,165,733	41,165,752	Nested	CTTGAGTTTGCCACCCAAAC
N2-DS	chr5	41,403,983	41,404,002	Biotin	/5Biosg/TGGCGATCTGAAGAACTAAG
	chr5	41,403,985	41,404,010	Nested	GCGATCTGAAGAACTAAGAAGCTTAG
N3	chr5	41,535,787	41,535,806	Biotin	/5Biosg/GTGGTTGTAAGTAGCAGACA
	chr5	41,535,790	41,535,813	Nested	GTTGTAAGTAGCAGACACAGAGAT

VP: Viewpoint

Table S7. Oligonucleotide primers for multiplexed 4C-seq of enhancer viewpoints at the *Nkx3-2* locus. The 4C-seq adapter and adapter-specific primer sequences are given in (*70*). N2-DS denotes its location as 18 kbp downstream of the actual N2 enhancer. All viewpoints are pointed towards *Nkx3-2* gene body ("+" strand).

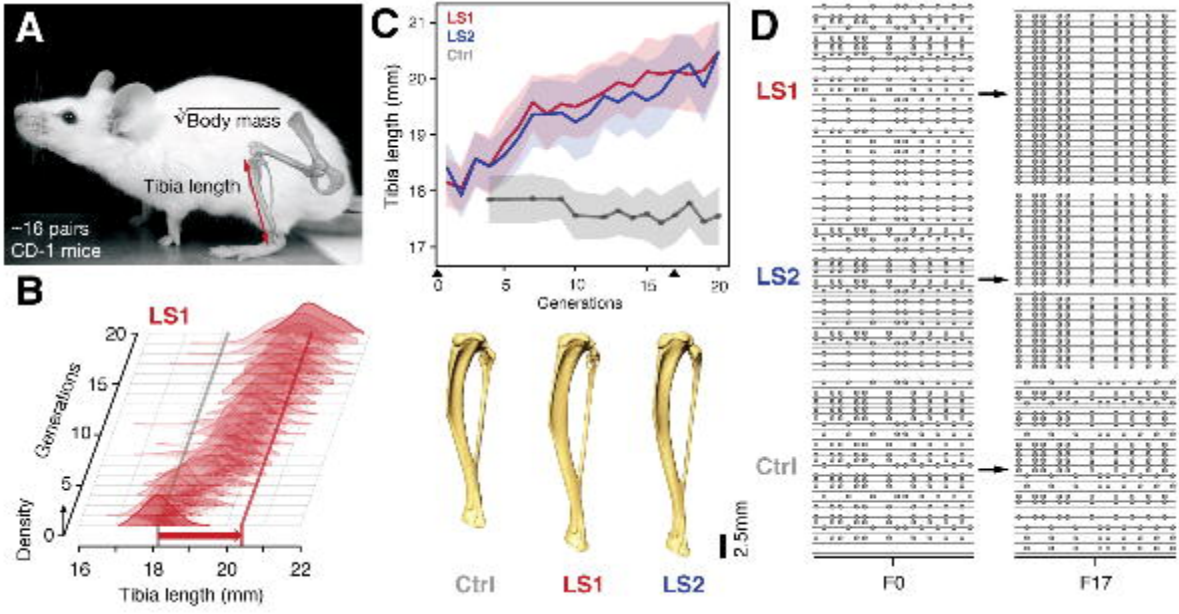
Enh	Chr	Start	End	Size	Type	Primer (5' - 3')
N1	chr5	41,165,777	41,166,605	829	For	TCAGCT**GATATC**TTCAGCTGTCGACGCTGAAC ATTAAAGCCATTACAGGAGCTTTCC
					Rev	TCAGCT**GATATC**TTCAGCTCTCGAGCTGCCAA CAGGCTCCTTCCTATAAACC
N3	chr5	41,536,216	41,537,263	1048	For	TCAGCT**GATATC**TTCAGCTGTCGACCACGAGA CTTAGCCAAACCTGAGGTG
					Rev	TCAGCT**GATATC**TTCAGCTCTCGAGCTTCCTC CCAGTGCATATCTCCCAAC
G2	chr13	15,145,904	15,146,863	960	For	TCAGCT**GATATC**TTCAGCTGTCGACCCTGGGT AAACTGCTTTCTGTAGCTCAAGG
					Rev	TCAGCT**GATATC**TTCAGCTCTCGAGGCGGTGG TTCTGTGGCATGC

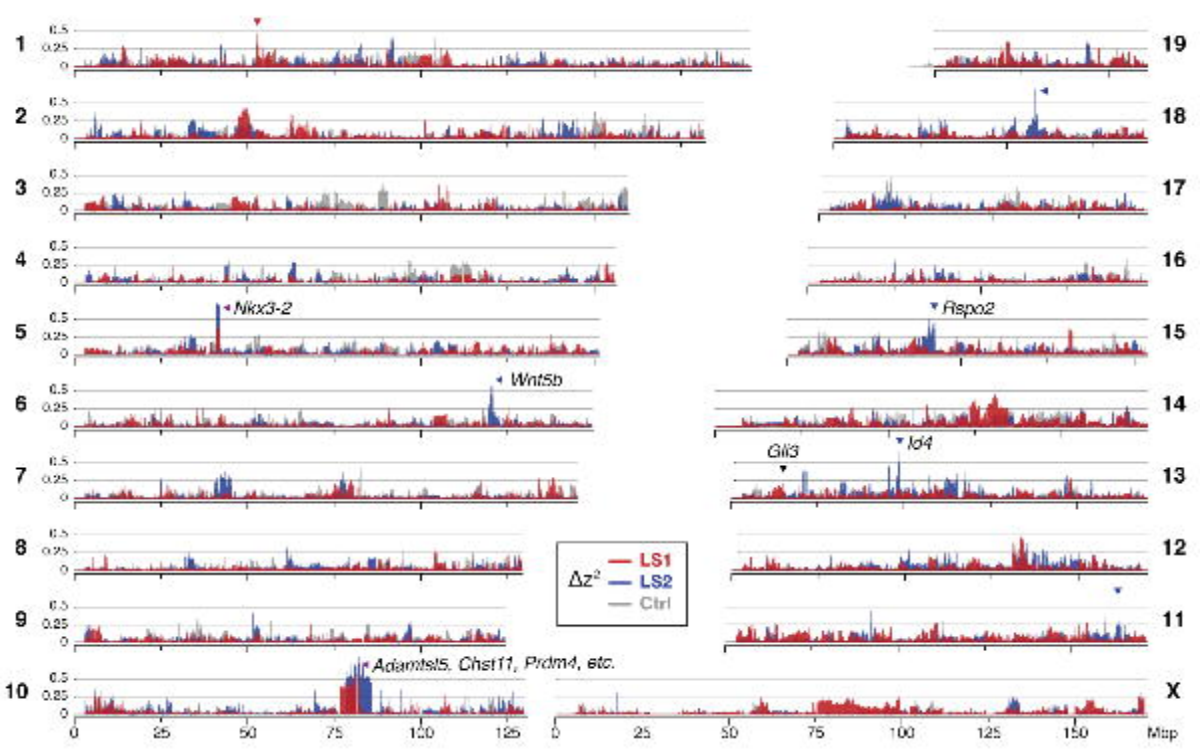
Enh: Enhancer

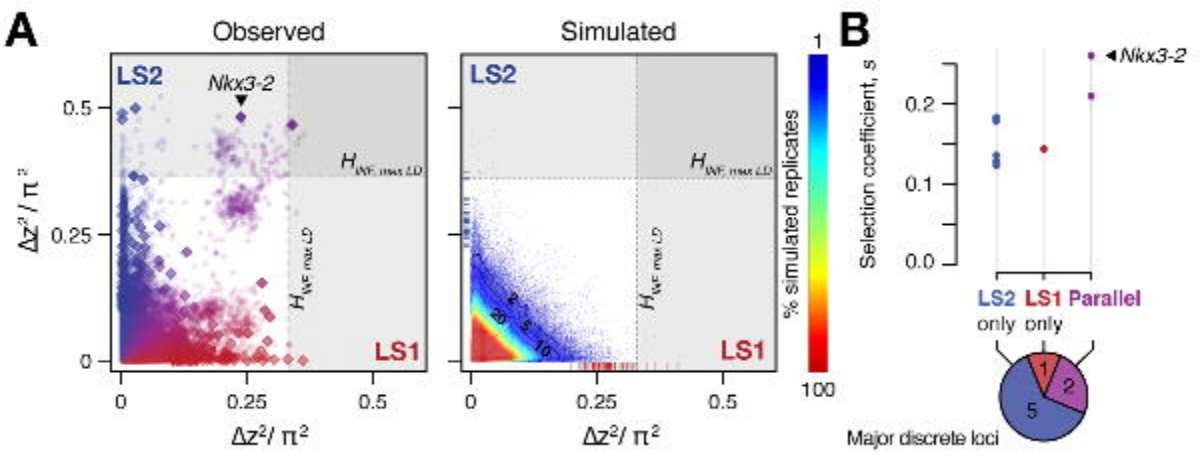
Table S8. Oligonucleotide primers for amplifying the enhancers at the *Nkx3-2* and *Gli3* loci. Each of the amplicons are tagged with *SalI* (forward) or *XhoI* (reverse) sites (underlined) for concatenation and flanked by *EcoRV* sites (underlined and bold) for insertion into the pBeta-lacZ-attBx2 reporter vector upstream of the β-globin minimal promoter.

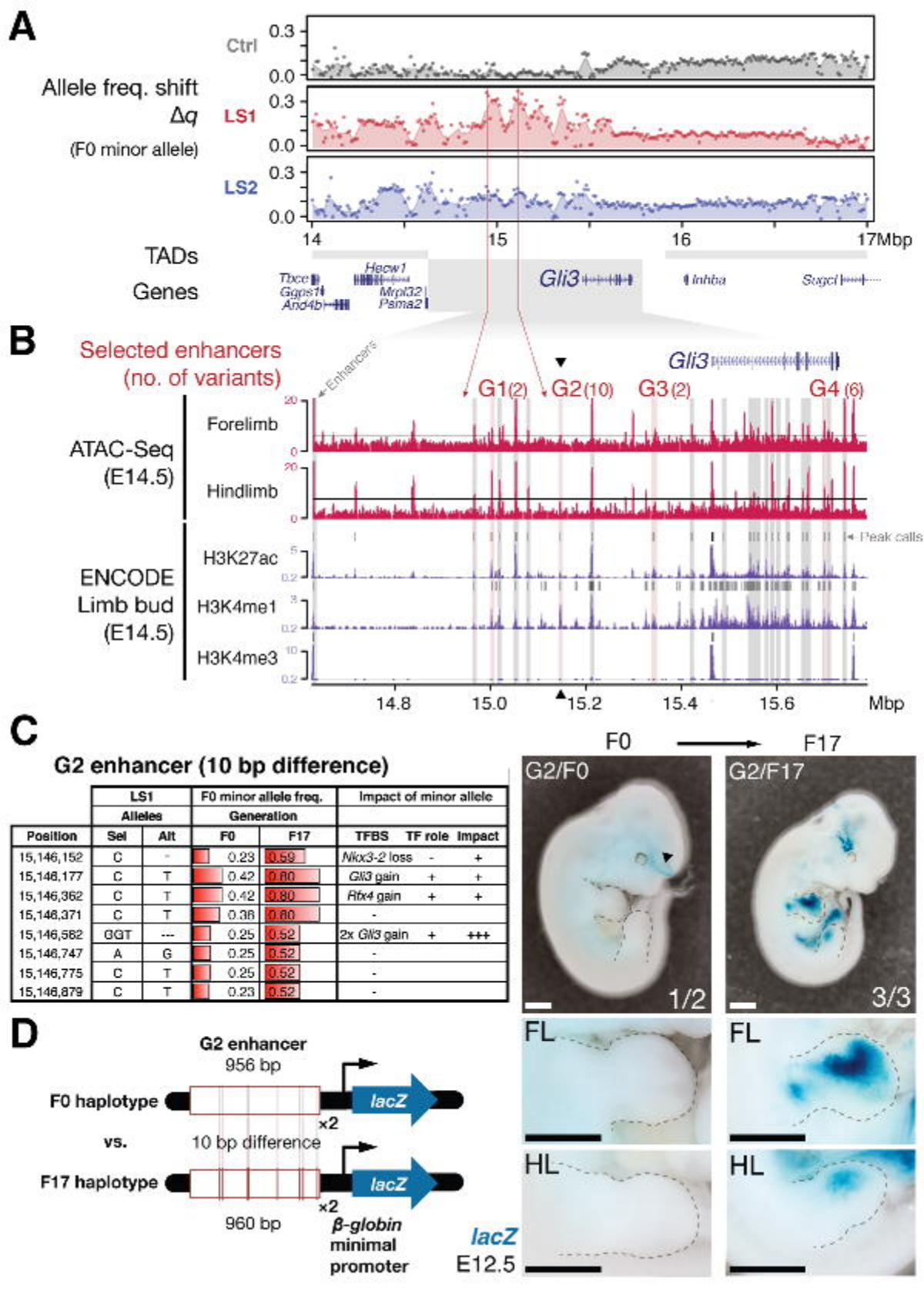
Allele	Chr	Start	End	Size	Type	Primer (5' - 3')
N3-F0 Maj	chr5	41,536,402	41,536,520	119	For	GCAAATCTTGAAAAGTCAGGGAATAAAAT**A**
					Rev	GCTAAATTTTGCCTCCTGCTTT**T**
N3-F17 selected	chr5	41,536,402	41,536,520	119	For	GCAAATCTTGAAAAGTCAGGGAATAAAAT**G**
					Rev	GCTAAATTTTGCCTCCTGCTTT**G**

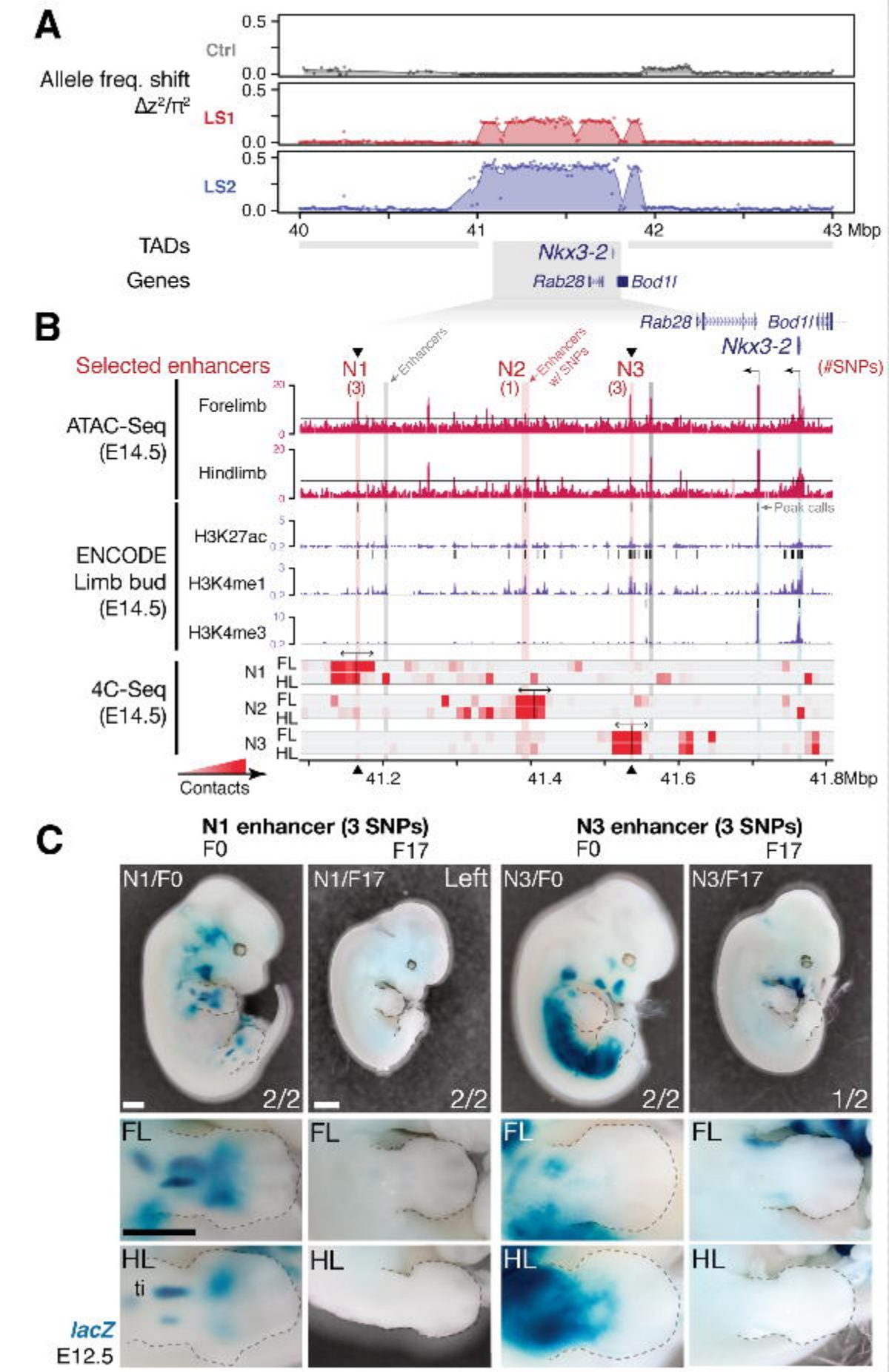
Table S9. Oligonucleotide primers for allele-specific genotyping of the N3 enhancer. The primers were designed to target two SNPs (bold) in the N3 enhancer, rs33219710 and rs33600994.


A
√Body mass
Tibia length
~16 pairs
CD-1 mice
B
LS1
Generations
Density
Tibia length (mm)
C
LS1
LS2
Ctrl
Tibia length (mm)
Generations
Ctrl
LS1
LS2
2.5mm
D
LS1
LS2
Ctrl
F0
F17


1
2
3
4
5
6
7
8
9
10
19
18
17
16
15
14
13
12
11
X
0.5
0.25
0
Nkx3-2
Wnt5b
Rspo2
Gli3
Id4
Adamts5, Chst11, Prdm4, etc.
Δz²
LS1
LS2
Ctrl
0
25
50
75
100
125
150
Mbp


A
Observed
Simulated
LS2
LS1
Nkx3-2
$H_{WF\ max\ LD}$
$\Delta z^2 / \pi^2$
0
0.25
0.5
% simulated replicates
1
100
2
5
10
20
B
Selection coefficient, s
0.0
0.1
0.2
Nkx3-2
LS2 only
LS1 only
Parallel
Major discrete loci
1
2
5


A
Allele freq. shift
Δq
(F0 minor allele)
Ctrl
LS1
LS2
0.3
0.0
14
15
16
17Mbp
TADs
Genes
Hoxw1
Tbce
Ggps1
And4b
Mrpl32
Psma2
Gli3
Inhba
Sugct
B
Selected enhancers
(no. of variants)
Enhancers
G1(2)
G2(10)
G3(2)
G4(6)
ATAC-Seq
(E14.5)
Forelimb
Hindlimb
ENCODE
Limb bud
(E14.5)
H3K27ac
H3K4me1
H3K4me3
Peak calls
14.8
15.0
15.2
15.4
15.6
Mbp
C
G2 enhancer (10 bp difference)
LS1 Alleles
F0 minor allele freq. Generation
Impact of minor allele
Position
Sel
Alt
F0
F17
TFBS
TF role
Impact
15,146,152 C - 0.23 0.59 Nkx3-2 loss - +
15,146,177 C T 0.42 0.80 Gli3 gain + +
15,146,362 C T 0.42 0.80 Rfx4 gain + +
15,146,371 C T 0.38 0.80 -
15,146,582 GGT --- 0.25 0.52 2x Gli3 gain + +++
15,146,747 A G 0.25 0.52 -
15,146,775 C T 0.25 0.52 -
15,146,879 C T 0.23 0.52 -
F0
F17
G2/F0
G2/F17
1/2
3/3
D
G2 enhancer
956 bp
F0 haplotype
lacZ
×2
vs.
10 bp difference
F17 haplotype
lacZ
×2
960 bp
β-globin
minimal
promoter
lacZ
E12.5
FL
HL
FL
HL


A
Allele freq. shift
Δz²/π²
Ctrl
LS1
LS2
0.5
0.0
40
41
42
43 Mbp
TADs
Genes
Nkx3-2
Rab28
Bod1l
B
Selected enhancers
N1
(3)
N2
(1)
N3
(3)
Enhancers
Enhancers w/ SNPs
(#SNPs)
ATAC-Seq (E14.5)
Forelimb
Hindlimb
Peak calls
ENCODE Limb bud (E14.5)
H3K27ac
H3K4me1
H3K4me3
4C-Seq (E14.5)
FL
HL
Contacts
41.2
41.4
41.6
41.8Mbp
C
N1 enhancer (3 SNPs)
F0
F17
N3 enhancer (3 SNPs)
N1/F0
N1/F17
Left
N3/F0
N3/F17
2/2
2/2
2/2
1/2
FL
HL
ti
lacZ
E12.5