ENHANCing the limb: from micro- to macro-evolution

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

João P. L. Castro

aus Ribeira Grande, Portugal

Tübingen

2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	16 Juni 2020
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Alfred Nordheim
2. Berichterstatter:	Dr. Yingguang Frank Chan

Acknowledgements

I would like to thank:

My advisor, **Frank Chan**, for the opportunity of this PhD adventure, his constant encouragement, the possibility to explore and experiment, his faith in my abilities and extraordinary patience.

Prof. Alfred Nordheim, for agreeing to be my second supervisor, for evaluating my thesis, being member of my PhD and Thesis committee and for his support.

Prof. Detlef Weigel and for being part of my PhD and Thesis committee and for his support and inputs over the years.

Prof. Oliver Betz for accepting to be part of my Thesis committee.

Felicity Jones, for all helpful comments, suggestions and shared knowledge.

Our collaborators, **Campbell Rolian**, **Nick Barton**, **John Cobb**, **Ronald Naumann**, **Joost Gribnau**, **Tino Hochepied**, **Marc Van Montagu**, **Claude Libert** and their lab members.

Sarah Danes, **Sybille Patheiger** and **Herta Soffel** for administrative support.

My lab mates and friends: **Layla**, **Kavita**, **Andreea, Bil, Dingwen, Elena, Enni, Marek, Stan, Ludmila, Muhua, Stefano, Li, Nico, Vrinda, Melanie, JP, Saad, Caroline, Insa, Moritz, Volker, Min, Sebastian K., Luba, Mish, Ivana, Sebastian, Effie, Hannes and Michal** for making these past years an amazingly fun and interesting journey, for all scientific and non-scientific discussions, for all lunch and coffee times.

A special thanks to **Volker**, **Insa** and **Moritz** for translating and prof-reading the abstract, wouldn't be able to do it without you.

Layla, **Kyle**, **Kavita** and **Effie**, for always being there for me, for keeping me sane, for all time spent together and general awesomeness.

My family, **Mário, Paula, Lídia, Joana, Rúben, José F., Sofia, Miguel, José L., Fátima L., Fátima P., Teresa** and all others, for their constant support, for being there and continuous motivation to pursue my dreams. I Love you guys.

Table of Contents

Abstract

Understanding the molecular basis of the diverse morphological forms found within and across species is a longstanding goal in evolutionary biology. One especially relevant class of *cis*-regulatory elements are enhancers. This is because mutations affecting enhancers tend to be tissue- or stage-specific, which allows adaptation to proceed with relatively less harmful side effects in other organs or tissues.

In **Chapter 2** I explore how enhancers help drive morphological selection response within-species. We scanned the genomes of the Longshanks mice, which are mice selectively bred over 20 generations, for a 13% increase in tibiae. Against a backdrop of polygenic response, we found the bone repressor *Nkx3-2*, and specifically its enhancers, to be among the strongest contributor towards increased tibia length. I used transgenics to compare the enhancer activity of the F0 and F17 alleles at 3 candidate enhancers (two near the *Nkx3-2* gene; and one near the limb developmental regulator gene, *Gli3*). We found that both loss-of-function (*Nkx3-2*) and gain-of-function (*Gli3*) alleles contributed to the selection response.

In **Chapter 3**, we explored an approach to study macro-evolutionary variations across species. One of the major barriers to such study is the inability to perform direct genetic crosses due to hybrid sterility. We tackle the species barrier problem by inducing mitotic recombination *in vitro* in hybrid embryonic stem cells (including cross-species hybrids between *Mus musculus* and *Mus spretus*). This was achieved via *Blm* inhibition by the small molecule ML216. We further show, that the resultant mitotic recombinant cells can be used for genetic mapping by connecting tioguanine drug resistance to variations at the *Hprt* locus. Furthermore, *in vitro* recombinant stem cells can be used for rederivation of animals through laser-assisted morula injection, thus allowing the acquisition of morphological data.

Here, through a multidisciplinary approach, we show that enhancer modulation contributes to morphological diversity and selection response within-species and provide a new methodology for enhancer study across-species, thus enabling the study of evolutionary developmental variations in genetic backgrounds that would otherwise be challenging to obtain. Overall, these studies highlight the relevance of enhancers in morphological diversification and provide new tools for their study.

Zusammenfassung

Das Verständnis der molekularen Grundlage der unterschiedlichen inner- und zwischenartlichen Morphologien ist ein langjähriges Ziel der Evolutionsbiologie. Eine zentrale Klasse von *cis*-wirkenden Elementen sind Enhancer. Dies liegt insbesondere daran, dass Mutationen in Enhancern in der Regel spezifisch für ein Gewebe oder Entwicklungsstadium sind und deshalb eine Anpassung ermöglichen, die nur wenige negative Effekte auf andere Organe und Gewebe hat.

In **Kapitel 2** untersuche ich den Beitrag von Enhancern zur morphologische Selektion innerhalb einer Art. Dafür wurden die Genome von „Longshanks"-Mäusen untersucht, deren Tibialänge über 20 Generationen um 13% selektiv erhöht wurde. Im Rahmen dieser polygenen Selektion, trug der Knochen-Repressor *Nkx3-2*, und speziell seine Enhancer, den größten Anteil zu der Verlängerung der Tibiaknochen bei. Experimente mit Transgenen zum Testen der Aktivität von Enhancerkandidaten (zwei nahe *Nkx3-2*; einer nahe des Regulatorgens für Gliedmaßen, *Gli3*) in F0 und F17 Allelen zeigten, dass sowohl der Funktionsverlust (*Nkx3-2)* als auch der Funktionsgewinn (*Gli3*) dieser Allele zu der Selektionsantwort beitrugen.

In **Kapitel 3** untersuchten wir mithilfe eines neuen Ansatzes die makro-evolutionären Unterschiede zwischen Arten, deren Kreuzung aufgrund von Hybridsterilität auf herkömmlichen Wege nicht möglich ist. Durch Induktion von *in vitro* mitotischer Rekombination in embryonalen Hybridstammzellen der beiden Mausspezies *Mus musculus* und *Mus spretus*, konnte das Sterilitätsprobelm der Hybride umgangen werden. Dies konnte durch die Inhibierung von *Blm* mit Hilfe des Moleküls ML216 erreicht werden. Durch Verbindung von Variationen im *Hprt* Lokus mit der Resistenz gegen das Medikament Thioguanine wurde gezeigt, dass mitotisch-rekombinante Zellen für genetische Kartierung geeignet sind. Darüber hinaus können diese Zellen mittels laserbasierter Morula-Injektionen zur Entstehung von Embryonen und somit zum Gewinn morphologischer Daten genutzt werden.

Dieser interdisziplinäre Ansatz zeigt, dass die Modulation von Enhancern zur morphologischen Vielfalt und Selektionsantwort innerhalb von Arten beiträgt und liefert eine neue Methode zur zwischenartlichen Enhancer-Analyse. Diese ermöglicht die Untersuchung evolutionär bedingter Variationen auch vor herausfordernden genetischen Hintergründen. Insgesamt zeigt diese Studie die Relevanz von Enhancern in der morphologischen Diversifizierung und bietet neue Analyseansätze.

Chapter 1: Aims and introduction

Understanding the role of genetic and molecular mechanisms in the evolution of morphological diversity within and between species is a longstanding goal in evolutionary and developmental biology[1-3]. Owing to the importance and pleiotropism of genes required for proper body development and patterning, changes to protein coding sequences themselves can be fatal[3,4]. Furthermore, such genes are deeply conserved and shared across divergent taxa. These observations lead to the so-called "*cis*-regulatory hypothesis of morphological evolution", which proposes that morphology evolves primarily through tweaking the *cis*-regulatory networks that control developmental gene expression, leading to alterations of pattern (heterotopy), timing (heterochrony), or level of expression of genes[1,2,4-9] (although cases where coding changes cause phenotypic alterations have been reported[10-13]).

In my dissertation research, I explored how enhancers, a category of *cis*-regulatory elements, contribute to intra- and inter-species morphological diversity. Here, I tackle these questions in two different ways: by exploring how enhancers contribute to morphological variation at the micro-evolutionary scale in the context of a mouse artificial selection experiment; and by developing a technique that opens the possibility to map how enhancers affect macro-evolutionary changes across species barriers.

In **Chapter 2**, taking advantage of an artificial selection experiment, called Longshanks, where mice were selected for longer tibia, we identified multiple loci across the genome that contribute to tibia length increase. Furthermore, for two genes related to tibia length, I found that both enhancer gain- and loss-of-function mutations can contribute to selection response, with as few as 3 SNPs being enough to alter enhancer activity.

In **Chapter 3,** I contributed to an effort to address one of the major barriers to the study of enhancers across species: hybrid sterility, which prevents crosses between species to be generated. We circumvented the species barrier problem by inducing stem cell mitotic recombination *in vitro* in F1 hybrid embryonic mouse stem cells, such as from a cross between *Mus musculus* and *Mus spretus*. We further show that the mitotic recombinant cells can be used for genetic mapping and,

excitingly, rederivation of animals through laser-assisted morula injection, thus allowing the acquisition of morphological data.

The aforementioned studies provide further support to the *cis*-regulatory hypothesis by showing that enhancer alterations can contribute to intraspecies morphological diversity and opening new research possibilities to investigate the *cis*-regulatory hypothesis at the interspecies level.

1.1 Morphological evolution and enhancers:

The magnitude and nature of genomic changes required for morphological evolution have not yet been clearly defined under a single paradigm[14], and research continues to gauge the importance, impact, and frequency of deletions, *de novo* creation, or modification of existing regulatory elements, epistatic interactions between such elements and associated pleiotropic effects[4,15-18]. To tackle these questions the scientific community employs powerful population genetic tools that allow the association of genotype to phenotype, such as evolve and re-sequence experiments[19], quantitative trait loci mapping (QTL)[20,21], and genome-wide association studies (GWAS)[22]; as well as genome wide regulatory elements screening techniques (such as ChIP-seq[23] and ATAC-seq[24]). When coupled, population genetics and regulatory-element screening offer a powerful approach to address the role of *cis*-regulation in morphological evolution.

Several categories of *cis*-regulatory elements that contribute to phenotypic diversity have been reported, including promoters[25,26], enhancers[27-31], silencers[32] and potentially insulators[33,34]. Enhancers are particularly relevant to the genesis of intra- and inter-species morphological diversity[28,29], but to gauge their importance we need to look into how enhancers orchestrate gene expression. As stated above, developmental genes are often pleiotropic, i.e., mutations in such genes affect many downstream phenotypes because they control patterning of multiple body structures[4,35]. A classic example of pleiotropy in vertebrates comes from *Hox* gene clusters, which are required for the proper patterning and development of the axial skeleton[36], limbs[37], reproductive tract[38], central nervous system[39] and muscle[40]. Other relevant examples, referenced in **Chapter 2**, are: the *Nkx3-2* gene that is

required for the development of the spleen[41,42], pancreas[41], and axial and limb skeleton[42,43]; and the *Gli3* gene, whose function is essential for proper development of the limbs[44,45], eyes[46], brain[47], and craniofacial structures[48]. Given their roles in development and the precise spatiotemporal control of gene expression required, it is no surprise that coding mutations in pleiotropic genes lead to severe morphological defects at birth or even perinatal lethality (as reported for *Nkx3-2*[42,49] and *Gli3*[50,51]).

Much of the control of gene expression during development can be attributed to enhancers, which act in a tissue and time specific manner through recruitment of specific sets of transcriptional factors[9,28,29,52]. Indeed, several pleiotropic developmental genes have been shown to possess elaborate enhancer networks that control the timing and space of expression. These include the *HoxD* cluster whose early and late limb expression is controlled by different sets of enhancers[53], and the complex *Shh* enhancer network that orchestrates specific limb and central nervous system expression patterns[54]. The *HoxD* and *Shh* enhancer networks also demonstrate another crucial feature of enhancers: modularity, i.e., each enhancer contributes independently and cumulatively to overall gene expression[9,54-56]. Furthermore, some enhancers exist as groups of redundant elements, as shown for *Gli3*[57] and *Shox2*[57] limb enhancers, thus providing developmental robustness[57]. Enhancers' tissue/organ specificity, modularity, and redundancy make them great targets for evolution, as these characteristics allow for tissue/organ-specific gene expression changes while simultaneously buffering the risk of lethality and facilitating the accumulation of genetic variation[58]. Several vertebrate and invertebrate studies have highlighted how impactful enhancer modification can be on species morphology[28-31]. Research performed in *Drosophila* showed that both intra- and inter-species morphological diversity can be attributed to gene expression variation caused by enhancer sequence alteration[27,30,31]. In vertebrates, enhancer deletion can significantly alter morphological traits such as the presence/absence of a pelvic spine in natural populations of stickleback fish[29] and the absence of limbs in snakes[28]. To understand how enhancer sequence and activity divergence mediate intra- and inter-species morphological diversity, it is essential to first understand their functional and structural organization.

1.2 Enhancer structure and control of transcriptional output:

Understanding enhancer function and structure is crucial to comprehend how enhancer sequence alteration can affect gene transcription, and lead to the development of novel techniques for enhancer study.

Enhancers are DNA sequences that make contact with gene promoters to initiate and/or amplify gene transcription in a distance- and orientation-independent manner, often spanning and skipping whole genes in between the enhancer and promoter[52,59,60]. These *cis*-regulatory elements range in size from tens[61] to hundreds[29,52,60] of base pairs, and are comprised of protein binding sites (such as for transcriptional factors and RNA polymerase II) and spacer elements[62-65]. An enhancer's "syntax" is the arrangement of these binding sites and spacer elements, which can affect the activity of the enhancer. A report on a tissue-specific enhancer in the *Ciona* genus of sea squirts showed that fully optimizing the enhancer's syntax can lead to a drastic alteration of enhancer activity, even causing the enhancer to be active in ectopic tissues/organs[62]. Surprisingly, robust tissue specificity was achieved with optimal spacing and low affinity biding sites[62]. This shows that there is ample ground for tweaking the transcriptional output and tissue specificity of an enhancer by altering its syntax, i.e., changing the number, type, affinity, spacing, and orientation of its transcription factor binding sites[58,62].

An additional feature of enhancers is the absence of nucleosomes in the enhancer "body", possibly due to displacement by transcription factors. Nucleosomes flanking enhancers also undergo post-translational modifications such as mono- or tri-methylation at the lysine 4 residue of the histone subunit H3 (abbreviated as "H3K4me" or "H3K4me3") [66-68], or acetylation at the lysine 27 residue ("H3K27ac")[69]. Identifying these features associated with chromatin accessibility or remodeling have enabled the development of techniques for genome-wide screening of enhancers, e.g., transposase accessible chromatin followed by sequencing (ATAC-seq)[24], which finds open chromatin regions, and chromatin immunoprecipitation sequencing (ChIP-seq)[23], which locates DNA regions associated with specific proteins or histone modifications. Additional reports have shown that the presence of RNA polymerase-II at enhancers leads to their bidirectional transcription into a class of non-coding RNAs known as eRNAs

(enhancer RNA)[70,71]. This latter signature can be captured most efficiently by polymerase run-on assays that can provide a read-out of active transcription, e.g., global run-on sequencing (GRO-seq)[72].

To increase a gene's transcriptional output, proper and specific contact must be established between the enhancer and the target gene's promoter. The enhancer's syntax-determined assembly of DNA-protein complexes is crucial for this enhancer-promoter contact; as such, tweaking enhancer-promoter communication offers another method for enhancer-mediated change in gene transcription. Enhancer-promoter affinity can even influence how ubiquitous or restricted gene expression is[76,77]. Furthermore, enhancer-promoter communication may require the interaction between specific transcriptional factors (biochemical compatibility)[75,76]. The mechanism of communication between enhancers and promoters is actively researched, and four models have been proposed[73,74]: linking, tracking, looping, and phase separation (**Table 1** provides a brief description of each model). Studies exploring enhancers' impact on morphology often overlook how enhancer sequence alteration affects enhancer-promoter contacts, and thus the significance of enhancer-promoter affinity alteration in the evolution of vertebrate morphological diversity remains poorly understood.

Table 1: Models of enhancer-promoter communication.

Model	Description	Range
Linking[73,74,78,79]	A chain of transcriptional proteins establishes direct enhancer-promoter link.	Very short range. Between promoter and proximal enhancer elements.
Tracking[80-82]	Enhancer-promoter contact is formed by unidirectional migration of the enhancer and its associated protein complexes towards the promoter. Intervening chromatin forms a loop.	15 kbp or less.
Looping[73,74,83]	Enhancer-promoter contact through looping of the intervening DNA sequence. Initial loop is hypothesized to result from random chromatin collision and stabilized by high protein-protein affinity.	Up to 1 Mbp.
Phase separation[84-86]	Specific to "super-enhancers". Forms membraneless liquid-liquid phase separated areas of high concentration of transcriptional factors.	Large range.

1.3 Techniques for enhancer identification and characterization:

A plethora of methodologies for enhancer identification and characterization is available to the scientific community. These techniques can be categorized into three groups: DNA based, RNA based, and chromatin based.

Methods based on DNA sequencing include enhancer reporter assays[87,88], which determine an enhancer's area of activity, and techniques for enhancer identification such as the enhancer trap assay[87], massively parallel reporter assays (MPRAS)[89], self-transcribing active regulatory region (STARR-seq)[90], and cross-species sequence comparison[91].

RNA based methods include deep transcriptome-wide sequencing (RNA-seq[71]), and RNA run-on assays that provide a snapshot of actively transcribed regions, such as GRO-seq[72], precision run-on sequencing (PRO-seq)[92], and chromatin run-on sequencing (ChRO-seq)[93].

Chromatin based methods include characterization of open or closed chromatin by screening for either: nucleosome free regions, e.g., ATAC-Seq[24], DNAse I hypersensitive sites sequencing (DNAse-seq)[94], and formaldehyde-assisted isolation of regulatory elements (FAIRE-Seq)[93]; or nucleosome bound regions, e.g., micrococcal nuclease digestion and sequencing (MNase-seq[95]). Other chromatin based techniques include 3D chromatin conformation capture (3C[96] and derivative techniques, such as 4C[97] and Capture-C[98]), which identify DNA-DNA contacts, and identification of genomic regions associated with transcription factors, RNA polymerase II, or histone modifications, e.g., ChIP-seq[23,99], chromatin interaction analysis with paired-end tag (Chia-PET)[100] and *in silico* transcriptional factor binding site analysis[101].

To date, no single method can simultaneously identify enhancers, report their strength, and describe their *in vivo* tissue/organ areas of activity; as such, a combination of multiple techniques is required for a thorough characterization of enhancers. **Table 2** summarizes the techniques used in this thesis. In **Chapter 2,** we characterized enhancers with a genome-wide screen for regulatory regions using ChIP-seq and ATAC-seq data, followed by enhancer reporter assays to simultaneously determine if the identified regions were true enhancers and to

measure their respective *in vivo* activity. We also used 4C and *in situ* hybridization to link the identified enhancers with their target genes, thus providing an extensive characterization of enhancer activity.

Table 2: Techniques used for enhancer identification and characterization.

Method	Objective	Advantages	Limitations
ATAC-seq (Assay for transposase-accessible chromatin)	Genome-wide open chromatin mapping.	Low input required, even down to a single cell. Simple and fast.	Does not allow for type-specification of the open chromatin regions. Biased towards small fragments. May lead to false positives due to unintentional closed-chromatin tagmentation.
ChIP-seq (Chromatin Immuno-precipitation)	Determine genome-wide specific DNA-protein association.	Can be used for genome-wide scans of DNA associated with transcriptional factors, histone modifications and other relevant proteins.	Dependent on antibody quality. High GC-content biased. High sequencing coverage required. Can have high background noise caused by cell-type specific variation.
3C (Chromosome conformation capture) and derivative methods (**4C, Capture-C**)	Identification of contacts between different DNA regions.	Links regulatory elements to their targets.	Long and laborious protocol. Tendency for high noise.
STARR-seq (Self-transcribing active regulatory region)	Genome-wide screen of enhancer activity.	High-throughput. No positional integration effects. Simultaneously assess enhancer activity and strength.	Episomal assay. Labor intensive.
Transgenic reporter assay	Visualize regulatory-element's driven expression domains	Can be used *in vivo* and *in vitro*. Allows detailed characterization of regulatory-elements activity.	High-cost. Subject to integration-site effects.
ISH (In situ hybridization) techniques	Visualize physical location of specific targets.	Permits visualization of endogenous gene expression patterns or physical proximity between DNA elements. Probes can be re-used.	Time-consuming and laborious. Low sensitivity and resolution. Dependent on probe quality.

1.4 The vertebrate limb as a model to study enhancers' impact on morphological diversity:

To assess how enhancer variation contributes to morphological diversity, a suitable model system should display both intra- and inter-species variation, be readily available and easy to manipulate. The vertebrate limb fits these requirements as various species-specific morphological adaptations have occurred in the course of evolution in response to various locomotory requirements[28,102]. Such adaptations

also show striking convergence across distant taxa, such as swimming (e.g. fish and whales), flying (e.g., birds and bats), running (e.g. antelopes, wolfs), and hopping (e.g. jerboa and kangaroos). In addition, the major genetic pathways involved in limb development have already been identified in vertebrates[103], and their mechanisms of action and impact on the patterning of all limb axes are well understood[103]. Furthermore, despite sharing the same core developmental program, the fore- and hindlimb can display striking morphological differences (e.g. bats and jerboa), hinting at fore- and hindlimb-specific genetic adaptations, which remain largely unexplored. Lastly, a large array of methodologies for limb studies is available ranging from morphometric analysis[102], to *in vivo* techniques (such as enhancer reporter assays[60]), to genome-wide characterization of regulatory landscapes (such as ATAC-seq[24,104] and ChIP-seq[105]). In **Chapter 2** we use the mouse limb as a model to study the impact of enhancers in intra-species morphological diversification and selection response.

1.5 Summary of findings and conclusions:

A long-standing goal in evolutionary biology is to understand the molecular basis of the diverse morphological forms found within and across species. In **Chapter 2** we explored how *cis*-regulatory elements, such as enhancers, may shape morphological traits at the micro-evolutionary scale. To identify loci that contribute to morphological variation in the mouse hindlimb, we scanned the genomes of the Longshanks mice, which were selectively bred over 20 generations for longer tibiae. Briefly, outbred CD-1 mice were separated into three closed populations with 14 mating pairs each and maintained as two selected lines (Longshanks 1, LS1; and Longshanks 2, LS2) and one control line. The control group was randomly bred, while LS1 and LS2 were selected for having the longest tibia length relative to body mass. After 20 generations, we observed an average tibia length increase of 13.1% and 12.7% respectively in LS1 and LS2, while the control line showed no significant tibia length change (see **Chapter 2, Fig.1, p.31**). We sequenced all breeders from generations 0 (F0) and 17 (F17) of the selection experiment and identified 8 major loci likely contributing to selection response (see **Chapter 2, Fig.2, p.34** and **Table1,**

p.36). Through chromatin profiling (combination of ChIP-seq, ATAC-seq and 4C) of two relevant regions where impactful coding changes were not found (*Nkx3-2* and *Gli3*), we identified possible enhancers contributing to the increase in tibia length in Longshanks mice. We used *in vivo* enhancer reporter assays to compare the activity of the F0 and F17 enhancer alleles at 3 candidate enhancers (designated N1 and N3 near the *Nkx3-2* gene and G2 near the *Gli3* gene) and found that both loss-of-function (N1 and N3) and gain-of-function (G2) alleles contributed to the selection response, with as few as 3 SNPs being sufficient to alter enhancer activity (see **Chapter 2, Fig.4 p.40** and **Fig.4-S1, p.96**). Furthermore, using stickleback enhancer reporter assays, we found that the mouse N1 enhancer is functionally conserved in fishes, hinting at deep conservation of *Nkx3-2* dependent bone maturation mechanisms (see **Chapter 2, Fig.5, p.45**). We demonstrate that combining multidisciplinary techniques is crucial to identify and assay genetic elements that contribute to morphological selection response at high resolution (in this case down to individual SNPs), and thus to understand the impact of enhancers on intra-species morphological diversity.

Overall, in the context of the Longshanks selection experiment, we found that tibia length is a polygenic trait, and that enhancers contributed to morphological selection response. Furthermore, both gain and loss of enhancer function can contribute to morphological evolution. We also found that *cis*-regulatory mechanisms were associated with selected alleles at a genome-wide scale, providing further support to the *cis*-regulatory hypothesis of morphological evolution (see **Chapter 2 Fig.3-S1, p.93**). Future work should determine whether the identified genes and enhancers that control tibia length are also involved in limb morphology variation at the macro-evolutionary scale by employing techniques for genome-wide enhancer screening (such as STARR-seq) or more focused methods (such as cross-species sequence comparison of target enhancers and enhancer reporter assays).

To date, a major barrier to the identification of the genetic basis of morphological evolution arises from the inability to perform mapping experiments across sufficiently diverged taxa due to cross-species hybrid sterility. In **Chapter 3** we show an approach that can overcome this problem by generating *in vitro* mitotic recombinant stem cells (IVR) from mice by suppressing the Bloom syndrome DNA helicase (*Blm*) using a small inhibitory molecule called ML216 (see **Chapter 3 Fig.1 p.120**). We applied this technique to F1 crosses between mouse strains (*Mus*

musculus BL6 x 129) and species (*Mus musculus* x *Mus castaneus* - BL6/CAST; and *Mus musculus* x *Mus spretus* - BL6/SPRET) and successfully obtained IVR stem cell colonies. Whole genome sequencing of recombinant BL6/129 and BL6/CAST clones revealed that one or few recombination events occurred per recombinant clone and that recombination tended to happen near the telomeric ends of chromosomes. However, integration of a selection cassette caused recombination events to occur between the centromere and the cassette integration site (see **Chapter 3, Fig.2, p122**).

Mapping traits in BL6/SPRET hybrids through traditional breeding is challenging due to F1 hybrid male sterility and reduced female fertility[106]; as such, applying classic mapping methodologies, though possible, is difficult. To determine if IVR cells can be used for mapping, we subjected F1 BL6/SPRET stem cells to *in vitro* recombination using ML216 and tested the resulting recombinant cells for tioguanine drug resistance. Following drug exposure we used FACS and whole genome sequencing ("Flow-mapping") on these cells and were able to map drug resistance to the *Hprt* (hypoxanthine-guanine phophoribosyltransferase) locus in 21 days or less (see **Chapter 3, Fig.3, p.126**). Furthermore, through laser-assisted morula injection of BL6/SPRET mitotic recombinant and non-recombinant stem cell clones, we obtained morphological data for individual clones. Using high-resolution X-ray micro computed tomography (microCT), we observed that embryos derived from non-recombinant BL6/SPRET stem cells showed normal development at embryonic day 14.5 (E14.5), while 4 out of 11 embryos derived from IVR stem cell clones showed severe craniofacial malformation and neural tube closure defects (see **Chapter 3, Fig.4, p.128** and **Fig.S8, p.162**). Overall, we highlight the utility of *blm* inhibition to enable genetic mapping across species pairs that suffer from hybrid sterility. Recombinant stem cell clones open the door to explore fundamental evolutionary questions of morphological and developmental diversity that would otherwise be impaired by traditional breeding strategies. Future work should focus on increasing the mitotic recombination rate of the stem cells derived from cross-species F1 crosses, and explore if recombinant clones can be used for mapping cell-type specific traits through *in vitro* differentiation.

In my dissertation, I studied the role of enhancers in intra-species morphological variation, and showed that enhancer activity alteration helps to drive morphological diversity, thus contributing to selection response. Furthermore I

helped to develop a novel method to overcome species barriers and open the door to research enhancer activity in previously hard-to-obtain cross-species genomic combinations. Overall, these studies illuminate the relevance of enhancers in morphological diversification and provide new tools for their study.

1.6 Bibliography:

1 Stern, D. L. Evolutionary developmental biology and the problem of variation. *Evolution* **54**, 1079-1091 (2000).

2 Stern, D. L. & Orgogozo, V. The loci of evolution: how predictable is genetic evolution? *Evolution* **62**, 2155-2177, doi:10.1111/j.1558-5646.2008.00450.x (2008).

3 Prud'homme, B., Gompel, N. & Carroll, S. B. Emerging principles of regulatory evolution. *Proc Natl Acad Sci U S A* **104 Suppl 1**, 8605-8612, doi:10.1073/pnas.0700488104 (2007).

4 Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36, doi:10.1016/j.cell.2008.06.030 (2008).

5 McMahon, A. P., Ingham, P. W. & Tabin, C. J. Developmental roles and clinical significance of hedgehog signaling. *Curr Top Dev Biol* **53**, 1-114 (2003).

6 Kusserow, A. *et al.* Unexpected complexity of the Wnt gene family in a sea anemone. *Nature* **433**, 156-160, doi:10.1038/nature03158 (2005).

7 Halder, G., Callaerts, P. & Gehring, W. J. Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila. *Science* **267**, 1788-1792 (1995).

8 Olson, E. N. Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922-1927, doi:10.1126/science.1132292 (2006).

9 Adachi, Y. *et al.* Conserved cis-regulatory modules mediate complex neural expression patterns of the eyeless gene in the Drosophila brain. *Mech Dev* **120**, 1113-1126, doi:10.1016/j.mod.2003.08.007 (2003).

10 Nackley, A. G. *et al.* Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* **314**, 1930-1933, doi:10.1126/science.1131262 (2006).

11 Stam, L. F. & Laurie, C. C. Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in Drosophila melanogaster. *Genetics* **144**, 1559-1564 (1996).

12 Zwart, M. P. *et al.* Unraveling the causes of adaptive benefits of synonymous mutations in TEM-1 beta-lactamase. *Heredity (Edinb)* **121**, 406-421, doi:10.1038/s41437-018-0104-z (2018).

13 Kamberov, Y. G. *et al.* Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**, 691-702, doi:10.1016/j.cell.2013.01.016 (2013).

14 Nadeau, N. J. & Jiggins, C. D. A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends Genet* **26**, 484-492, doi:10.1016/j.tig.2010.08.004 (2010).

15 Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. & Carroll, S. B. Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. *Nature* **433**, 481-487, doi:10.1038/nature03235 (2005).
16 Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol* **20**, 1377-1419, doi:10.1093/molbev/msg140 (2003).
17 Ludwig, M. Z. *et al.* Functional evolution of a cis-regulatory module. *PLoS Biol* **3**, e93, doi:10.1371/journal.pbio.0030093 (2005).
18 Phillips, P. C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855-867, doi:10.1038/nrg2452 (2008).
19 Schlotterer, C., Kofler, R., Versace, E., Tobler, R. & Franssen, S. U. Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity (Edinb)* **116**, 248, doi:10.1038/hdy.2015.85 (2016).
20 Mackay, T. F., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-577, doi:10.1038/nrg2612 (2009).
21 Allen Orr, H. The genetics of species differences. *Trends Ecol Evol* **16**, 343-350, doi:10.1016/s0169-5347(01)02167-x (2001).
22 Nicod, J. *et al.* Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. *Nat Genet* **48**, 912-918, doi:10.1038/ng.3595 (2016).
23 Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods* **4**, 651-657, doi:10.1038/nmeth1068 (2007).
24 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).
25 Khan, R. *et al.* Genetic variants in the TORC2 gene promoter and their association with body measurement and carcass quality traits in Qinchuan cattle. *PLoS One* **15**, e0227254, doi:10.1371/journal.pone.0227254 (2020).
26 Young, R. S. *et al.* The frequent evolutionary birth and death of functional promoters in mouse and human. *Genome Res* **25**, 1546-1557, doi:10.1101/gr.190546.115 (2015).
27 Rebeiz, M., Jikomes, N., Kassner, V. A. & Carroll, S. B. Evolutionary origin of a novel gene expression pattern through co-option of the latent activities of existing regulatory sequences. *Proc Natl Acad Sci U S A* **108**, 10036-10043, doi:10.1073/pnas.1105937108 (2011).
28 Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642 e611, doi:10.1016/j.cell.2016.09.028 (2016).
29 Chan, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* **327**, 302-305, doi:10.1126/science.1182213 (2010).
30 Rebeiz, M., Pool, J. E., Kassner, V. A., Aquadro, C. F. & Carroll, S. B. Stepwise modification of a modular enhancer underlies adaptation in a Drosophila population. *Science* **326**, 1663-1667, doi:10.1126/science.1178357 (2009).
31 Frankel, N. *et al.* Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598-603, doi:10.1038/nature10200 (2011).

32 Johnson, W. C. *et al.* Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between Drosophila Species. *PLoS Genet* **11**, e1005279, doi:10.1371/journal.pgen.1005279 (2015).
33 Heger, P. & Wiehe, T. New tools in the box: an evolutionary synopsis of chromatin insulators. *Trends Genet* **30**, 161-171, doi:10.1016/j.tig.2014.03.004 (2014).
34 Heger, P., George, R. & Wiehe, T. Successive gain of insulator proteins in arthropod evolution. *Evolution* **67**, 2945-2956, doi:10.1111/evo.12155 (2013).
35 Gelbart, W. M. The decapentaplegic gene: a TGF-beta homologue controlling pattern formation in Drosophila. *Development* **107 Suppl**, 65-74 (1989).
36 Wellik, D. M. Hox patterning of the vertebrate axial skeleton. *Dev Dyn* **236**, 2454-2463, doi:10.1002/dvdy.21286 (2007).
37 Zakany, J. & Duboule, D. The role of Hox genes during vertebrate limb development. *Curr Opin Genet Dev* **17**, 359-366, doi:10.1016/j.gde.2007.05.011 (2007).
38 Taylor, H. S. The role of HOX genes in the development and function of the female reproductive tract. *Semin Reprod Med* **18**, 81-89, doi:10.1055/s-2000-13478 (2000).
39 Goncalves, C. S., Le Boiteux, E., Arnaud, P. & Costa, B. M. HOX gene cluster (de)regulation in brain: from neurodevelopment to malignant glial tumours. *Cell Mol Life Sci*, doi:10.1007/s00018-020-03508-9 (2020).
40 Pineault, K. M. & Wellik, D. M. Hox genes and limb musculoskeletal development. *Curr Osteoporos Rep* **12**, 420-427, doi:10.1007/s11914-014-0241-0 (2014).
41 Asayesh, A. *et al.* Spleen versus pancreas: strict control of organ interrelationship revealed by analyses of Bapx1-/- mice. *Genes Dev* **20**, 2208-2213, doi:10.1101/gad.381906 (2006).
42 Tribioli, C. & Lufkin, T. The murine Bapx1 homeobox gene plays a critical role in embryonic development of the axial skeleton and spleen. *Development* **126**, 5699-5711 (1999).
43 Rainbow, R. S., Won, H. K. & Zeng, L. The role of Nkx3.2 in chondrogenesis. *Front Biol (Beijing)* **9**, 376-381, doi:10.1007/s11515-014-1321-3 (2014).
44 Cao, T., Wang, C., Yang, M., Wu, C. & Wang, B. Mouse limbs expressing only the Gli3 repressor resemble those of Sonic hedgehog mutants. *Dev Biol* **379**, 221-228, doi:10.1016/j.ydbio.2013.04.025 (2013).
45 Bowers, M. *et al.* Limb anterior-posterior polarity integrates activator and repressor functions of GLI2 as well as GLI3. *Dev Biol* **370**, 110-124, doi:10.1016/j.ydbio.2012.07.017 (2012).
46 Wiegering, A., Petzsch, P., Kohrer, K., Ruther, U. & Gerhardt, C. GLI3 repressor but not GLI3 activator is essential for mouse eye patterning and morphogenesis. *Dev Biol* **450**, 141-154, doi:10.1016/j.ydbio.2019.02.018 (2019).
47 Rash, B. G. & Grove, E. A. Shh and Gli3 regulate formation of the telencephalic-diencephalic junction and suppress an isthmus-like signaling source in the forebrain. *Dev Biol* **359**, 242-250, doi:10.1016/j.ydbio.2011.08.026 (2011).
48 Rachel, R. A., Wellington, S. J., Warburton, D., Mason, C. A. & Beermann, F. A new allele of Gli3 and a new mutation, circletail (Crc), resulting from a single transgenic experiment. *Genesis* **33**, 55-61, doi:10.1002/gene.10088 (2002).
49 Simsek-Kiper, P. O. *et al.* A novel NKX3-2 mutation associated with perinatal lethal phenotype of spondylo-megaepiphyseal-metaphyseal dysplasia in a neonate. *Eur J Med Genet* **62**, 21-26, doi:10.1016/j.ejmg.2018.04.013 (2019).

50 Demurger, F. *et al.* New insights into genotype-phenotype correlation for GLI3 mutations. *Eur J Hum Genet* **23**, 92-102, doi:10.1038/ejhg.2014.62 (2015).
51 Naruse, I., Ueta, E., Sumino, Y., Ogawa, M. & Ishikiriyama, S. Birth defects caused by mutations in human GLI3 and mouse Gli3 genes. *Congenit Anom (Kyoto)* **50**, 1-7, doi:10.1111/j.1741-4520.2009.00266.x (2010).
52 Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* **33**, 717-728, doi:10.1016/0092-8674(83)90014-4 (1983).
53 Rodriguez-Carballo, E. *et al.* The HoxD cluster is a dynamic and resilient TAD boundary controlling the segregation of antagonistic regulatory landscapes. *Genes Dev* **31**, 2264-2281, doi:10.1101/gad.307769.117 (2017).
54 Anderson, E., Devenney, P. S., Hill, R. E. & Lettice, L. A. Mapping the Shh long-range regulatory domain. *Development* **141**, 3934-3943, doi:10.1242/dev.108480 (2014).
55 Geyer, P. K. & Corces, V. G. Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in Drosophila melanogaster. *Genes Dev* **1**, 996-1004, doi:10.1101/gad.1.9.996 (1987).
56 Goto, T., Macdonald, P. & Maniatis, T. Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422, doi:10.1016/0092-8674(89)90916-1 (1989).
57 Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239-243, doi:10.1038/nature25461 (2018).
58 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:10.1016/j.cell.2016.09.018 (2016).
59 Levine, M. Transcriptional enhancers in animal development and evolution. *Curr Biol* **20**, R754-763, doi:10.1016/j.cub.2010.06.070 (2010).
60 Lettice, L. A. *et al.* A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**, 1725-1735 (2003).
61 Bertrand, V., Hudson, C., Caillol, D., Popovici, C. & Lemaire, P. Neural tissue in ascidian embryos is induced by FGF9/16/20, acting via a combination of maternal GATA and Ets transcription factors. *Cell* **115**, 615-627, doi:10.1016/s0092-8674(03)00928-0 (2003).
62 Farley, E. K., Olson, K. M., Zhang, W., Rokhsar, D. S. & Levine, M. S. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proc Natl Acad Sci U S A* **113**, 6508-6513, doi:10.1073/pnas.1605085113 (2016).
63 Murre, C. *et al.* Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell* **58**, 537-544, doi:10.1016/0092-8674(89)90434-0 (1989).
64 Koch, F., Jourquin, F., Ferrier, P. & Andrau, J. C. Genome-wide RNA polymerase II: not genes only! *Trends Biochem Sci* **33**, 265-273, doi:10.1016/j.tibs.2008.04.006 (2008).
65 Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition

between poised and active gene expression. *EMBO J* **26**, 2041-2051, doi:10.1038/sj.emboj.7601654 (2007).
66 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).
67 Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40**, 897-903, doi:10.1038/ng.154 (2008).
68 Andersson, R., Sandelin, A. & Danko, C. G. A unified architecture of transcriptional regulatory elements. *Trends Genet* **31**, 426-433, doi:10.1016/j.tig.2015.05.007 (2015).
69 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
70 De Santa, F. *et al.* A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* **8**, e1000384, doi:10.1371/journal.pbio.1000384 (2010).
71 Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187, doi:10.1038/nature09033 (2010).
72 Lopes, R., Agami, R. & Korkmaz, G. GRO-seq, A Tool for Identification of Transcripts Regulating Gene Expression. *Methods Mol Biol* **1543**, 45-55, doi:10.1007/978-1-4939-6716-2_3 (2017).
73 Vernimmen, D. & Bickmore, W. A. The Hierarchy of Transcriptional Activation: From Enhancer to Promoter. *Trends Genet* **31**, 696-708, doi:10.1016/j.tig.2015.10.004 (2015).
74 Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* **14**, 288-295, doi:10.1038/nrg3458 (2013).
75 Butler, J. E. & Kadonaga, J. T. Enhancer-promoter specificity mediated by DPE or TATA core promoter motifs. *Genes Dev* **15**, 2515-2519, doi:10.1101/gad.924301 (2001).
76 van Arensbergen, J., van Steensel, B. & Bussemaker, H. J. In search of the determinants of enhancer-promoter interaction specificity. *Trends Cell Biol* **24**, 695-702, doi:10.1016/j.tcb.2014.07.004 (2014).
77 Zabidi, M. A. *et al.* Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556-559, doi:10.1038/nature13994 (2015).
78 Meng, H. & Bartholomew, B. Emerging roles of transcriptional enhancers in chromatin looping and promoter-proximal pausing of RNA polymerase II. *J Biol Chem* **293**, 13786-13794, doi:10.1074/jbc.R117.813485 (2018).
79 Szutorisz, H. *et al.* Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage. *Mol Cell Biol* **25**, 1804-1820, doi:10.1128/MCB.25.5.1804-1820.2005 (2005).
80 Ho, Y., Elefant, F., Liebhaber, S. A. & Cooke, N. E. Locus control region transcription plays an active role in long-range gene activation. *Mol Cell* **23**, 365-375, doi:10.1016/j.molcel.2006.05.041 (2006).
81 Routledge, S. J. & Proudfoot, N. J. Definition of transcriptional promoters in the human beta globin locus control region. *J Mol Biol* **323**, 601-611, doi:10.1016/s0022-2836(02)01011-2 (2002).

82 Zhu, X. *et al.* A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Res* **35**, 5532-5544, doi:10.1093/nar/gkm595 (2007).
83 Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr Biol* **12**, 439-445, doi:10.1016/s0960-9822(02)00695-4 (2002).
84 Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23, doi:10.1016/j.cell.2017.02.007 (2017).
85 Nair, S. J. *et al.* Phase separation of ligand-activated enhancers licenses cooperative chromosomal enhancer assembly. *Nat Struct Mol Biol* **26**, 193-203, doi:10.1038/s41594-019-0190-5 (2019).
86 Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842-1855 e1816, doi:10.1016/j.cell.2018.10.042 (2018).
87 Kvon, E. Z. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* **106**, 185-192, doi:10.1016/j.ygeno.2015.06.007 (2015).
88 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308, doi:10.1016/0092-8674(81)90413-x (1981).
89 Melnikov, A. *et al.* Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**, 271-277, doi:10.1038/nbt.2137 (2012).
90 Arnold, C. D. *et al.* Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
91 Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I. & Hardison, R. C. Cross-species sequence comparisons: a review of methods and available resources. *Genome Res* **13**, 1-12, doi:10.1101/gr.222003 (2003).
92 Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-1476, doi:10.1038/nprot.2016.086 (2016).
93 Chu, T. *et al.* Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat Genet* **50**, 1553-1564, doi:10.1038/s41588-018-0244-3 (2018).
94 Galas, D. J. & Schmitz, A. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res* **5**, 3157-3170, doi:10.1093/nar/5.9.3157 (1978).
95 Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* **474**, 516-520, doi:10.1038/nature10002 (2011).
96 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
97 Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17-32, doi:10.1016/j.ymeth.2019.07.014 (2020).
98 Davies, J. O. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat Methods* **13**, 74-80, doi:10.1038/nmeth.3664 (2016).
99 Villar, D. *et al.* Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566, doi:10.1016/j.cell.2015.01.006 (2015).

100 Li, G. *et al.* Chromatin Interaction Analysis with Paired-End Tag (ChIA-PET) sequencing technology and application. *BMC Genomics* **15 Suppl 12**, S11, doi:10.1186/1471-2164-15-S12-S11 (2014).
101 Fang, Y., Wang, Y., Zhu, Q., Wang, J. & Li, G. In silico identification of enhancers on the basis of a combination of transcription factor binding motif occurrences. *Sci Rep* **6**, 32476, doi:10.1038/srep32476 (2016).
102 Moore, T. Y. *et al.* Multiple phylogenetically distinct events shaped the evolution of limb skeletal morphologies associated with bipedalism in the jerboas. *Curr Biol* **25**, 2785-2794, doi:10.1016/j.cub.2015.09.037 (2015).
103 Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat Rev Genet* **18**, 245-258, doi:10.1038/nrg.2016.167 (2017).
104 Roscito, J. G. *et al.* Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. *Nat Commun* **9**, 4737, doi:10.1038/s41467-018-07122-z (2018).
105 VanderMeer, J. E., Smith, R. P., Jones, S. L. & Ahituv, N. Genome-wide identification of signaling center enhancers in the developing limb. *Development* **141**, 4194-4198, doi:10.1242/dev.110965 (2014).
106 Burgio, G. *et al.* Interspecific recombinant congenic strains between C57BL/6 and mice of the Mus spretus species: a powerful tool to dissect genetic control of complex traits. *Genetics* **177**, 2321-2333, doi:10.1534/genetics.107.078006 (2007).

Chapter 2: An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice

2.1 Declaration of Contributions:

Authors: João P. L. Castro*, Michelle N. Yancoskie*, Marta Marchini, Stefanie Belohlavy, Layla Hiramatsu, Marek Kučka, William H. Beluch, Ronald Naumann, Isabella Skuplik, John Cobb, Nicholas H. Barton, Campbell Rolian+, Yingguang Frank Chan+. (*co-first authorship; +co-last authorship)

João P. L. Castro contributions: Identified candidate enhancers using histone ChIP-seq data from ENCODE, designed and prepared mouse and stickleback enhancer reporter assay constructs, performed and analyzed the results of mouse and stickleback enhancer reporter assays, designed and prepared *in situ* hybridization probes, performed *in situ* hybridization on BL6 mouse embryos; aided manuscript writing, revision and editing.
Significance of author contributions to the collective work: The data obtained from mouse enhancer reporter assays and *in situ* hybridization provides a crucial link between enhancer alteration, gene expression and response to selection, thus providing molecular validation of relevant loci. Stickleback enhancer reporter assays demonstrate deep functional conservation of the N1 enhancer.

Co-author contributions: Michelle N. Yancoskie: performed and analyzed 4C and ATAC-seq experiments, manuscript writing, review and editing; **Marta Marchini**: mouse breeding and phenotyping, pedigree data analysis, biological sample collection. **Stefanie Belohlavy**: pedigree and genomic data analysis, performed simulations, and reviewed the manuscript. **Layla Hiramatsu**: pedigree data analysis, and reviewed the manuscript. **Marek Kučka**: prepared genomic libraries for sequencing, performed RT-qPCR, and reviewed the manuscript. **William H. Beluch**: prepared genomic libraries for sequencing. **Ronald Naumann**: transient mouse transgenic reporter assay injections. **Isabella Skuplik** and **John Cobb:** performed

and analyzed *in situ* hybridization in Longshanks lines. **Nicholas H. Barton**: pedigree and genomic data analysis, performed simulations, reviewed and edited the manuscript. **Campbell Rolian**: experimental design, selection, phenotyping, sample collection, analyzed phenotypic and pedigree data, manuscript writing, review and editing. **Yingguang Frank Chan**: experimental design, pedigree and genomic data analysis, performed simulations, manuscript writing, reviewing and editing.

2.2 Full article

An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice

Authors: João P. L. Castro [1,*], Michelle N. Yancoskie [1,*], Marta Marchini [2], Stefanie Belohlavy [3], Layla Hiramatsu [1], 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,†]
(* indicates equal contribution; † indicates equal contribution)

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 authors:
Campbell Rolian (cprolian@ucalgary.ca)
Yingguang Frank Chan (frank.chan@tue.mpg.de)

eLIFE
Received: 14 September 2018
Accepted: 19 May 2019
Published: 06 June 2019
Reviewing editor: Magnus Nordborg, Austrian Academy of Sciences, Austria.

2.2.1 Abstract:

Evolutionary studies are often limited by missing data that are critical to understanding the history of selection. Selection experiments, which reproduce rapid evolution under controlled conditions, are excellent tools to study how genomes evolve under 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 in two replicates. We synthesized evolutionary theory, genome sequences and molecular genetics to understand the selection response and found that it involved both polygenic adaptation and discrete loci of major effect, with the strongest loci tending to be selected in parallel between replicates. We show that selection may favor de-repression of bone growth through inactivating two limb enhancers of an inhibitor, *Nkx3-2*. Our integrative genomic analyses thus show that it is possible to connect individual base-pair changes to the overall selection response.

2.2.2 Introduction:

Understanding how populations adapt to a changing environment is an urgent challenge of global significance. The problem is especially acute for mammal populations, which are often small and fragmented due to widespread habitat loss. Such populations often show increased inbreeding, leading to the loss of genetic diversity (*1*). Because beneficial alleles in mammals typically come from standing genetic variation rather than new mutations like in microbes, this loss of diversity would ultimately impose a limit on the ability of small populations to adapt. Nonetheless, mammals respond readily to selection in many traits, both in nature and in the laboratory (2-*5*). In quantitative genetics, such traits are interpreted as the overall effect from a large set of loci, each with an infinitesimally small (and undetectable) effect ("infinitesimal model"). Broadly speaking, the infinitesimal model has performed remarkably well across a wide range of selection experiments, and the model is the basis for commercial breeding (*6*). However, it remains unclear what type of genomic change is associated with rapid response to selection,

especially in small populations where allele frequency changes can be dominated by random genetic drift.

While a large body of theory exists to describe the birth, rise and eventual fixation of adaptive variants under diverse selection scenarios (*7*-*12*), few empirical datasets capture sufficient detail on the founding conditions and selection regime to allow full reconstruction of the selection response. This is particularly problematic in nature, where historical samples, environmental measurements and replicates are often missing. Selection experiments, which reproduce rapid evolution under controlled conditions, are therefore excellent tools to understand response to selection—and by extension—adaptive evolution in nature (*4*).

Here we describe an integrative, multi-faceted investigation into an artificial selection experiment, called Longshanks, in which mice were selected for increased tibia length relative to body mass (*13*). The mammalian 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 limb development has been studied extensively in mammals, birds and fishes as a genetic and evolutionary paradigm (*14*). The Longshanks selection experiment thus offers the opportunity to study selection response not only from a quantitative and population genetics perspective, but also from a developmental (*15*) 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. Previously, Marchini *et al.* investigated how selection was able to overcome correlation between tibia length and body mass and produced independent changes in tibia length during the first 14 generations of the Longshanks experiment (*13*). Importantly, that study focused on the phenotypes and inferred genetic correlations indirectly using the pedigree. The current genomic analysis was initiated when the on-going experiment reached generation 17 and extends the previous study by integrating both phenotypic *and* genetic aspects of the Longshanks experiment. By sequencing the initial and final genomes, the current analysis benefits from direct and highly resolved genetic information. Here, with essentially complete information, we wish to answer a number of important questions regarding the factors that determine and constrain rapid adaptation: Are the observed changes in gene frequency due to selection or random drift? Does rapid selection response of a complex trait proceed through innumerable loci of

infinitesimally small effect, or through a few loci of large effect? What type of signature of selection may be associated with this process? Finally, when the same trait changes occur independently, do these depend on changes in the same gene(s) or the same pathways (parallelism)?

2.2.3 Results:

Longshanks selection for longer tibiae

At the start of the Longshanks experiment, we established three base populations with 14 pairs each by sampling from a genetically diverse, commercial mouse stock [Hsd:ICR, also known as CD-1; derived from mixed breeding of classical laboratory mice (*16*)]. In two replicate "Longshanks" lines (LS1 and LS2), we bred mice by pairing 16 males and females (and excluding sibling pairs) with the longest tibia relative to the cube root of body mass for each sex. This corresponds to 15–20% of all offspring [only details essential to understanding our analysis are summarized here. See (*13*) for a detailed description of the breeding scheme]. We kept a third Control line (Ctrl) using an identical breeding scheme, except that breeders were selected at random. In LS1 and LS2, we observed a strong and significant response to selection in tibia length [0.29 and 0.26 Haldane or standard deviations (s.d.) per generation, from a selection differential of 0.73 s.d. in LS1 and 0.62 s.d. in LS2]. Over 20 generations, selection for longer relative tibia length produced increases of 5.27 and 4.81 s.d. in LS1 and LS2, respectively (or 12.7% and 13.1% in tibia length), with a modest decrease in body mass [-1.5% in LS1 and -3.7% in LS2; Student's *t*-test, $P < 2\times10^{-4}$ and $P < 1\times10^{-8}$, respectively; Fig. 1B & C; Fig. 1 – figure supplement 1; n.b. this relationship was in part biased by the F1 generation, which were fed a different diet and phenotyped three weeks later than later generations, see (*13*) for details]. By contrast, Ctrl showed no directional change in tibia length or body mass (Fig. 1C; Student's *t*-test, $P > 0.05$). This approximately 5 s.d. change in 20 generations is rapid compared to typical rates observed in nature [(*17*), but see (*18*)] but is in line with responses seen in selection experiments (*3*, *5*, *19*, *20*).

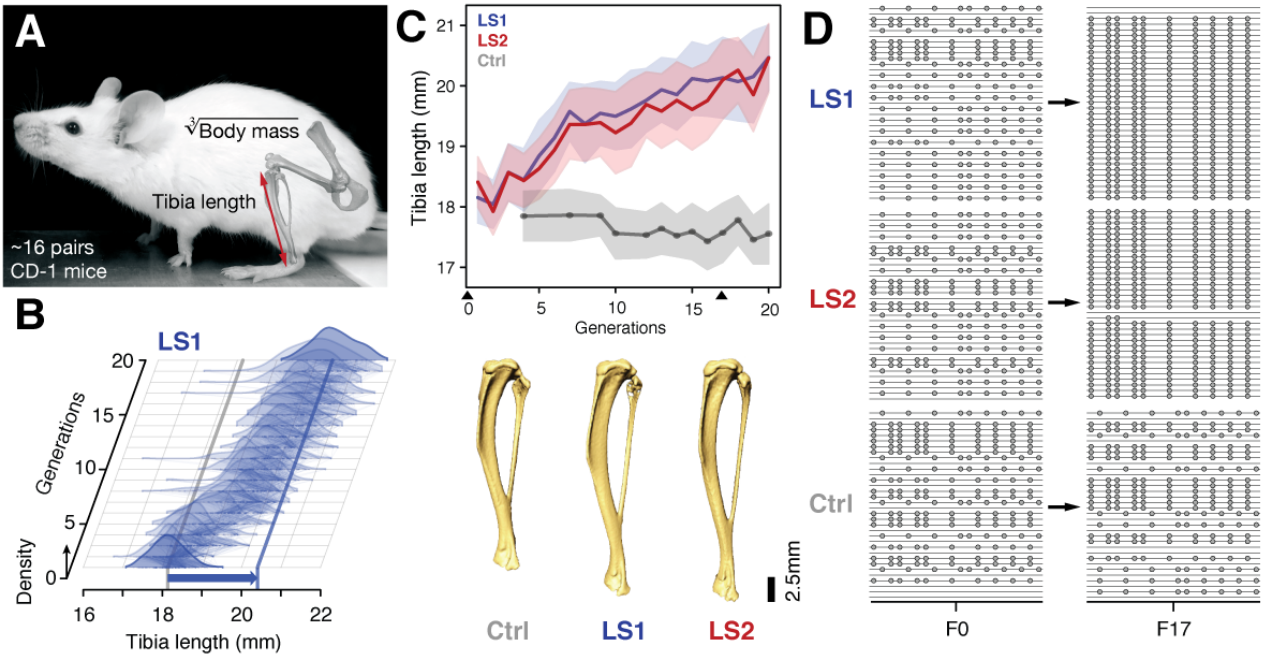


Fig. 1. Selection for Longshanks mice produced rapid increase in tibia length. (**A** and **B**) Tibia length varies as a quantitative trait among outbred mice derived from the Hsd:ICR (also known as CD-1) commercial stock. Selective breeding for mice with the longest tibiae relative to body mass within families has produced a strong selection response in tibia length over 20 generations in Longshanks mice (13%, blue arrow, LS1). (**C**) Both LS1 and LS2 produced replicated rapid increase in tibia length (blue and red; line and shading show mean ± s.d.) compared to random-bred Controls (grey). Arrowheads along the x-axis mark sequenced generations F0 and F17. See Fig. 1 – figure supplement 1 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). Alternatively, selection may elicit a polygenic response, which may involve minor shifts in allele frequency at many loci and therefore may leave a very different selection signature from the one shown here.

Simulating selection response: infinitesimal model with linkage

The rapid but generally smooth increase in tibia length in Longshanks is typically interpreted as evidence for a highly dispersed genetic architecture with no individually important loci contributing to the selection response. This is classically described under quantitative genetics as the infinitesimal model. Crucially, the appropriate null hypothesis for the genomic response here should capture "polygenic adaptation" rather than a neutral model. We therefore 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. Using the actual pedigree and trait measurements, we mapped fitness onto tibia length $\boldsymbol{T}$ and cube-root body mass $\boldsymbol{B}$ as a single composite trait $ln(\boldsymbol{T}\boldsymbol{B}^{\phi})$. We estimated ϕ from actual data as −0.57, such that the ranking of breeders closely matched the actual composite ranking used to select breeders in the selection experiment, based on $\boldsymbol{T}$ and $\boldsymbol{B}$ separately (*13*) (Fig. 1 – figure supplement 2A). We assumed a 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. 1 – figure supplement 2B). We tested models with varying selection intensity and initial linkage disequilibrium (LD), and for each, ran 100 simulated replicates to determine the significance of changes in allele frequency (Fig. 1 – figure supplement 2C–E). This flexible quantitative genetics framework allowed us to explore possible changes in genetic diversity over 17 generations of breeding under strong selection.

In simulations, we followed blocks of genomes 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. 1 – figure supplement 2B) and in the distribution of allele frequencies (Fig. 2 – figure supplement 1B) 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 threshold tests on simulations that were seeded with SNPs drawn with LD consistent with the initial haplotypes (Fig. 1 – figure supplement 2C & E; see Supplementary Notes).

Because our simulations assume infinitesimal effects of loci, allele frequency shifts exceeding this stringent threshold suggest 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; Supplementary File 1). 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; Supplementary File 2; Fig. 2 – figure supplement 1A & Fig. 2 – figure supplement 2). 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 (Supplementary File 2). This is consistent with random sampling from an outbred breeding stock. By F17, the number of segregating SNPs dropped to around 5.8 M (Supplementary File 2). This 13% drop in diversity (0.9M SNPs genome-wide) is predicted by drift. Our simulation confirmed this and moreover, showed that selection contributed negligibly to the drop in diversity (Supplementary Notes, Fig. 1 – figure supplement 2B, D).

We conclude that despite the strong selection on the LS lines, there was little perturbation to genome-wide diversity. Indeed, the changes in diversity in 17 generations were remarkably similar in all three lines, despite Ctrl not having experienced selection on relative tibia length (Fig. 2 – figure supplement 1A). Hence, and consistent with our simulation results (Fig. 1 – figure supplement 2B, D), changes in global genome diversity had little power to distinguish selection from neutral drift despite the strong *phenotypic* selection response.

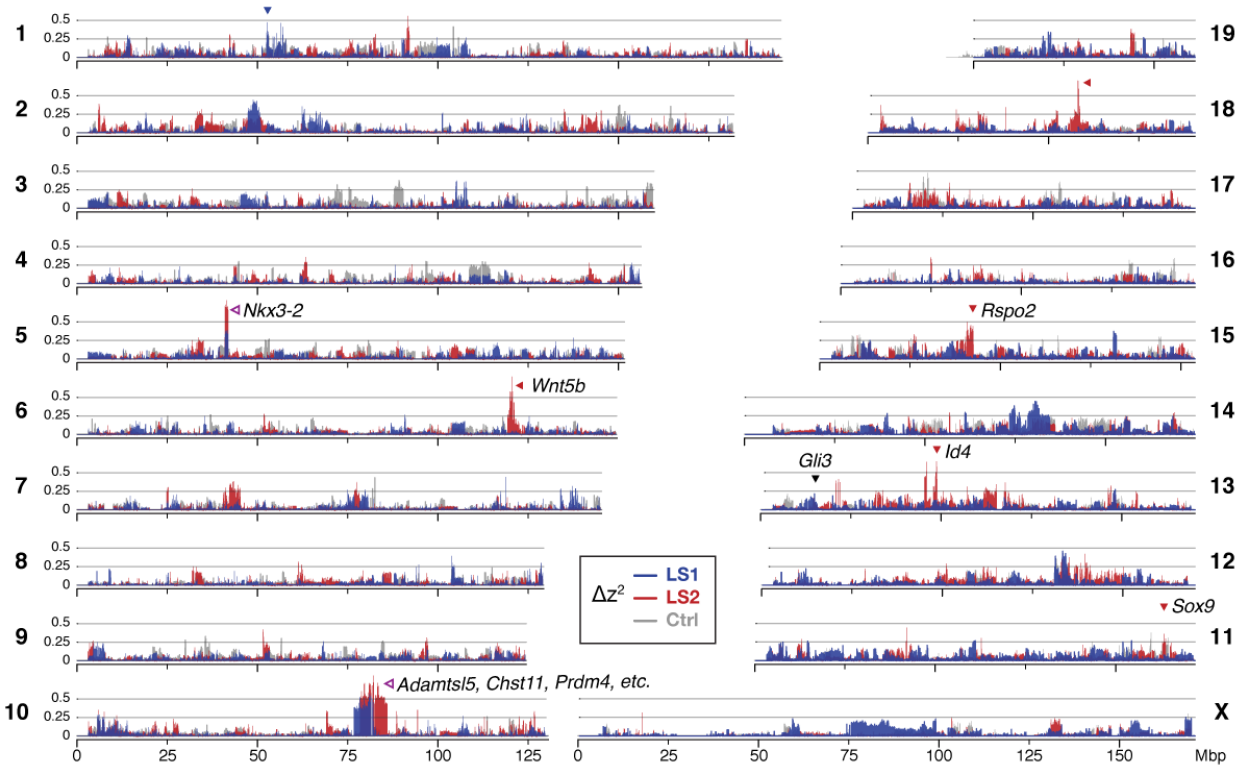


Fig. 2. Widespread genomic response to selection for increased tibia length. Allele frequency shifts between generations F0 and F17 in LS1, LS2 and Ctrl lines are shown as Δz^2 profiles across the genome (plotted here as fraction of its range from 0 to π^2). The Ctrl Δz^2 profile (grey) confirmed our 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 (blue) and LS2 (red) 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 (arrowheads, blue: LS1; red: LS2; purple: parallel; see also Fig. 1 – figure supplement 2E, Fig. 2 – figure supplement 2, Fig. 2 – figure supplement 3) beyond a polygenic background, which may explain a majority of the selection response and yet leave little discernible selection signature. Candidate genes are highlighted (Table 1). An additional *a priori* candidate limb regulator *Gli3* is indicated with a black arrowhead.

We next asked whether specific loci 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 (i.e., loci) that had significant shifts in allele frequency in LS1 and/or LS2 (corresponding to 9 and 164 clustered windows respectively at $P \leq 0.05$ under $H_{INF,\ max\ LD}$; $\Delta z^2 \geq 0.33\ \pi^2$; genome-wide $\Delta z^2 = 0.02 \pm 0.03\ \pi^2$; Fig. 2; Fig. 1 – figure supplement 2D, Fig. 2 – figure supplement 2, Fig. 2 – figure supplement 3; see Methods for details) and 8 windows in 3 clusters in Ctrl (genome-wide $\Delta z^2 = 0.01 \pm 0.02\ \pi^2$). The eight loci each overlapped between 2 to 179 genes and together contained 11 candidate genes with known roles in bone, cartilage and/or limb development (e.g., *Nkx3*-2 and *Sox9*; Table 1; Fig. 2 – figure supplement 3, Fig. 3 – figure supplement 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 SNPs coding for different amino acids, suggesting that for the majority of loci with large shifts in allele frequency, gene regulation was likely important in the selection response (Fig. 3 – figure supplement 1; Supplementary File 3; 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 performed better than moderate LD or no LD (Fig. 1 – figure supplement 2E); 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. 2 – figure supplement 2). Thus, we conclude that although the genetic basis of the selection response in the Longshanks experiment may be largely polygenic, evidence strongly suggests discrete loci with major effect, even when each line is considered separately.

Rk	Chr	Span (Mbp)	Peak	Core (kbp)	TAD (kbp)	Genes	Δq LS1	Δq LS2	Δq Ctrl	Type	Candidate genes
1	5	38.95–45.13	41.77	900	720	*3*	0.69	0.86	−0.14	Parallel	*Nkx3-2*
2	10	77.47–87.69	81.07	5360	6520	*175*	0.79	0.88	−0.04	Parallel	*Sbno5, Aes, Adamtsl5*, Chst11*, Cry1, Prdm4**
3	18	53.63–63.50	58.18	220	520	*4*	0.05	0.78	−0.06	LS2-specific	-
4	13	35.59–55.21	48.65	70	2600	*22*	0.24	0.80	−0.03	LS2-specific	*Id4*
5	1	53.16–57.13	55.27	10	720	*4*	0.65	0.01	−0.23	LS1-specific	-
6	15	31.92–44.43	41.54	10	680	*3*	−0.23	0.66	0.02	LS2-specific	*Rspo2**
7	6	118.65–125.25	120.30	130	1360	*12*	−0.03	0.79	−0.15	LS2-specific	*Wnt5b**
8	11	111.10–115.06	113.42	10	2120	*2*	−0.14	0.66	−0.15	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 ***Dixon et al. (2012)***.

Candidate genes, 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'.

DOI: https://doi.org/10.7554/eLife.42014.011

Table 1. Major loci likely contributing to selection response. These 8 loci show significant allele frequency shifts in Δz^2 and are ordered according to their estimated selection coefficients according to (*21*). 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 LS1 and LS2, 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 phenotype functions are shown, with asterisks (*) marking those with a "short tibia" knockout phenotype (see also Fig. 2 – figure supplement 3 and Supplementary File 3 for full table).

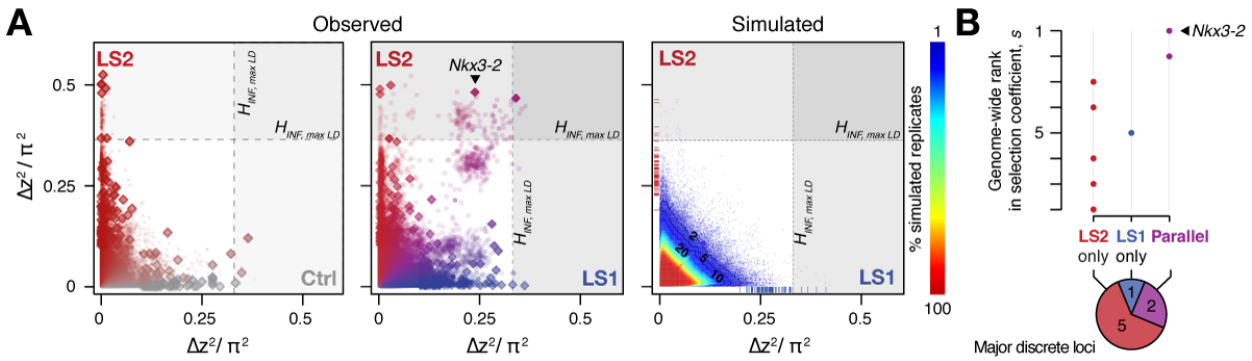


Fig. 3. Selection response in the Longshanks lines was largely line-specific, but the strongest signals occurred in parallel.

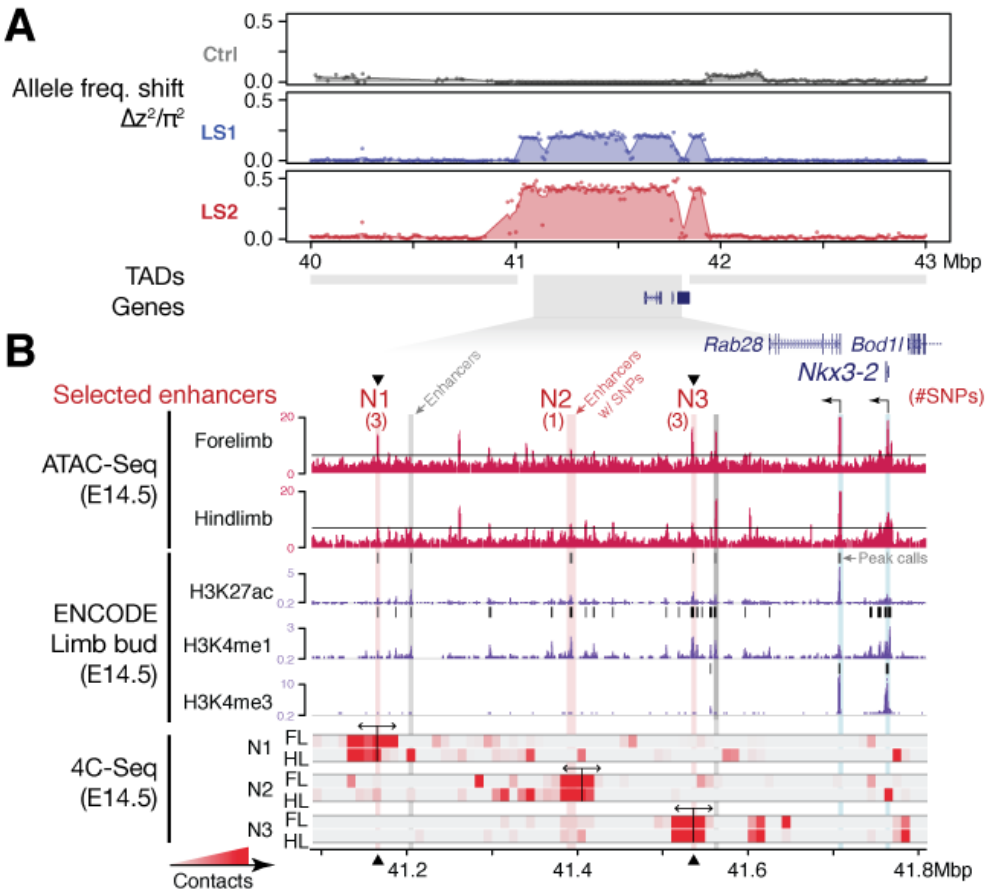
(**A**) Allele frequencies showed greater shifts in LS2 (red) than in Ctrl (grey; left panel; diamonds: peak windows; dots: other 10 kbp windows; see Fig. 3 – figure supplement 2 for Ctrl vs. LS1 and Supplementary Notes for details). Changes in the two lines were not correlated with each other. In contrast, there were many more parallel changes in a comparison between LS1 (blue) vs. LS2 (red; middle panel; adjacent windows appear as clusters due to hitchhiking). 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. 3 – figure supplement 2 for an example replicate). Tick marks along the axes show genome-wide maximum Δz^2 shifts in each of 100 replicate simulations in LS1 (x-axis, blue) and LS2 (y-axis, red), from which we derived line-specific thresholds at the $P \leq 0.05$ significance level. While the frequency shifts from simulations matched the bulk of the observed data well, no simulation recovered the strong parallel shifts observed between LS1 and LS2 (compare middle to right panel, points along the diagonal). (**B**) Genome-wide ranking based on estimated selection coefficients *s* among the candidate discrete loci at $P \leq 0.05$ under $H_{INF,\ max\ LD}$. While six out of eight total loci showed significant shifts in only LS1 or LS2, the two loci with the highest selection coefficients were likely selected in parallel in both LS1 and LS2 (also see middle 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 LS1 and LS2 were more similar to each other than to Ctrl (Fig. 2 & 3A; Fig. 3 – figure supplement 2; Pearson's correlation in Δz^2 from 10 kbp windows: LS1–LS2: 0.21 vs. LS1–Ctrl: 0.06 and LS2–Ctrl: 0.05). Whereas previous genomic studies with multiple natural or artificial selection replicates focused mainly on detecting parallel loci (*23*-*26*), 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 all eight selected alleles were present in the F0 generation in both lines. This prevalence of line-specific loci was consistent under different significance 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 (*27*). Finding just two parallel loci out of 8 discrete loci may appear to be low, given the genetic similarity in the founding generation and the identical selection applied to both Longshanks replicates. However, one should bear in mind the very many genetic paths to increasing tibia length under an infinitesimal model, and that the effect of drift is expected to be very strong in these small populations. In larger populations, the shift in the balance from drift to selection should result in selection being able to favour increasingly subtle variants and thus produce a greater proportion of parallel loci. However, we expect the trend of parallelism being enriched among the top loci to hold. In contrast to the subtle differences within each line in changes in global diversity over 17 generations (Fig. 2 and Fig. 2 – figure supplement 2), we found the signature of parallelism to be significantly enriched in the comparison between the selected replicates (χ^2 test, LS1–LS2: $P \leq 1 \times 10^{-10}$), as opposed to comparisons between each selected lines and Ctrl (LS1–Ctrl: $P > 0.01$ and LS2–Ctrl: $P > 0.2$, both non-significant after correcting for multiple testing), or between simulated replicates (Fig. 3 – figure supplement 2; see Supplementary Notes for details). Because the parallel selected loci between LS1 and LS2 have the highest selection coefficients and parallelism is not generally expected in our

populations, these loci provide the strongest evidence for the role of discrete major loci. As such, the top-ranked parallel locus is the prime candidate for molecular dissection (see Fig. 4 – figure supplement 1 and Supplementary Notes "Molecular dissection of *Gli3*" for an additional *a priori* candidate locus with known limb function).

Molecular dissection of the* Nkx3-2 *locus highlights* cis*-acting changes

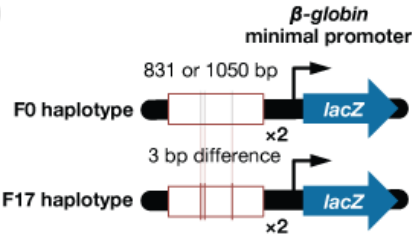
Between the two major parallel loci, we chose the locus on chromosome 5 (Chr5) at 41–42 Mbp for functional validation because it showed the strongest estimated selection coefficient, its clear signature of selection was clear, and crucially for functional characterization, it contains only three genes, including *Nkx3-2* (also known as *Bapx1*), a known regulator of bone maturation (Fig. 2 & 4A) (*29*). At this locus, the pattern of variation resembles a selective sweep spanning 1 Mbp (Fig. 4A). Comparison between F0 and F17 individuals revealed no recombinant in this entire region (Fig. 5 – figure supplement 1A, top panel), precluding fine-mapping using recombinants. We then analyzed the genes in this region to identify the likely target(s) of selection. First, we determined that no coding changes existed for either *Rab28* or *Nkx3-2,* the two genes located within the topologically associating domain (TADs, which mark chromosome segments with shared gene regulatory logic) (*22*). 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. 4 – figure supplement 2B). A third gene, *Bod1l*, straddled the TAD boundary with its promoter located in the neighbouring TAD, making its regulation by sequences in the selected locus unlikely. Consistent with this, *Bod1l* showed only weak or undetectable expression in the developing limb bud (Fig. 4 – figure supplement 2A). We next combined ENCODE chromatin profiles and our own 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 and LS2 (Fig. 4B & C; Fig. 5 – figure supplement 1A).



C

Enh	Position	Alleles Sel/Alt		Allele frequency LS1 Gen F0	LS1 Gen F17	LS2 Gen F0	LS2 Gen F17	Predicted transcription factor (TF) binding sites	TF role	Impact
	41,166,187	G	A	0.19	0.84	0.14	0.97			
N1	41,166,264	C	G	0.15	0.81	0.12	0.97	Hox12	+	-
	41,166,505	A	G	0.13	0.88	0.12	0.98			
N2	41,392,871	A	G	0.17	0.88	0.18	0.82	Zic1/2/3 & Gli3	-/(+)	-/(+)
	41,536,250	G	T	0.17	0.83	0.10	0.98	2x Nkx3-2	++	----
N3	41,536,431	G	A	0.23	0.83	0.14	0.98			
	41,536,498	C	A	0.23	0.86	0.16	0.98	Nkx3-1,Tcf/Lef1	+/~?	-

D



E

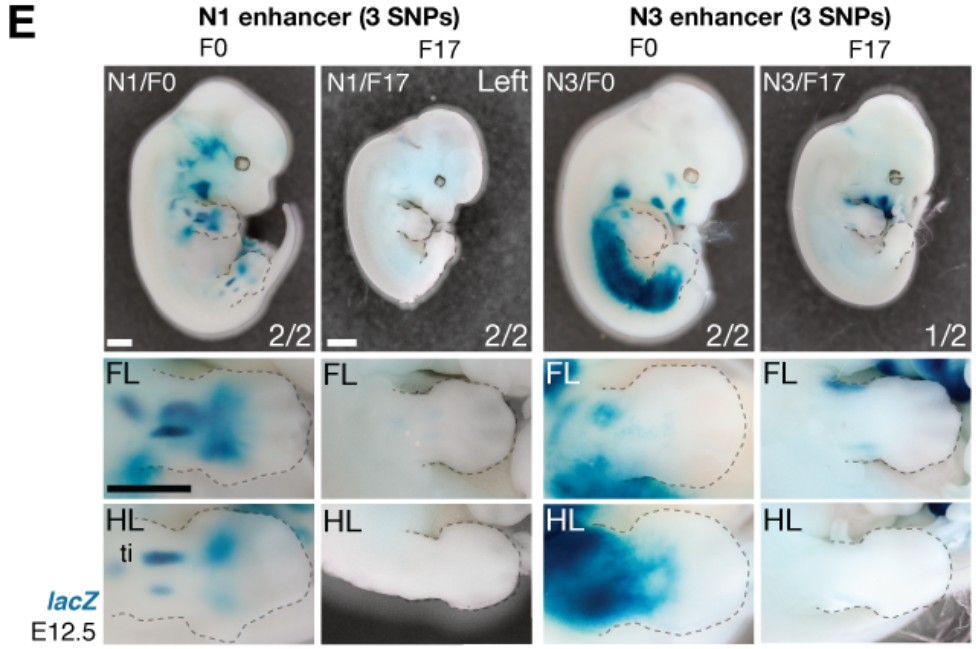


Fig. 4. Strong parallel selection response at the bone maturation repressor *Nkx3-2* locus was associated with decreased activity of two enhancers.

(**A**) Δz^2 in this region of chromosome 5 showed strong parallel differentiation spanning 1 Mbp in both Longshanks but not in the Control line. This 1 Mbp region contains three genes *Nkx3-2*, *Rab28* and *Bod1l* (whose promoter lies outside the TAD boundary, shown as grey boxes). Although an originally rare allele in all lines, this region had swept almost to fixation by generation 17 in LS2 (Fig. 5 – figure supplement 1A). (**B**) Chromatin profiles [ATAC-Seq, red, (*28*); 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 from grey to red showing increasing contacts; Promoters are shown with black arrows and blue vertical shading). (**C**) Selected alleles at 7 SNPs found within the N1–3 enhancers increased ~0.75 in frequency in both LS1 and LS2. Selected alleles at 3 of these sites are predicted to lead to loss (red inhibition circles) of transcription factor binding sites in the *Nkx3-2* pathway (including a SNP in N3 causing loss of 2 adjoining *Nkx3-2* binding sites) and thus reduce enhancer activity in N1 and N3. (**D, E**) 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 hind limb buds (FL, HL; ti: tibia). Fractions indicate number of embryos showing similar *lacZ* staining out of all transgenic embryos. Substituting the F17 enhancer allele (replacing 3 positions each in N1 and N3) led to little observable limb bud expression in both the N1/F17 and N3/F17 embryos, 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.

Chromosome conformation capture assays showed that the N1–3 sequences formed long-range looping contacts with the *Nkx3-2* promoter—a hallmark of enhancers—despite as much as 600 kbp of intervening sequence (Fig. 4B). We next used transgenic reporter assays to determine whether these sequences could drive expression in the limbs. Here, we were not only interested in whether the sequence encoded enhancer activity, but specifically whether the SNPs would affect the activity (Fig. 4C & D). An examination of the predicted transcription factor binding sites showed that both the N1 and N3 enhancers contain multiple SNPs with consistent directional impact on the putative enhancer activity (Fig. 4C). In contrast, the N2 enhancer contains only a single SNP and is predicted to have inconsistent effect on its activity. We therefore excluded the N2 enhancer from further testing. We found that the F0 alleles of the N1 and N3 enhancers (3 SNPs each in about 1 kbp) drove robust and consistent *lacZ* expression in the developing limb buds (N1 and N3) as well as in expanded trunk domains (N3) at E12.5 (Fig. 4E). In contrast, transgenic reporters carrying the selected F17 alleles of the N1 and N3 enhancers drove consistently weak, nearly undetectable *lacZ* expression (Fig. 4E). Thus, switching from the F0 to the F17 enhancer alleles led to a nearly complete loss in activity ("loss-of-function"). This is consistent with the role of *Nkx3-2* as a repressor in long bone maturation (*29*). It should be noted that even though our selective regime favored an increase in the target phenotype (tibia length), at the molecular level we expect advantageous loss- and gain-of-function variants to be equally likely favored by selection. In fact, in an additional functional validation example at the *Gli3* locus, we found a gain-of-function enhancer variant that may have been favored at that locus (see Fig. 4 – figure supplement 1 and Supplementary Notes "Molecular dissection of *Gli3*").

At the *Nkx3-2* locus, we hypothesize that the F17 allele causes *de-repression* of bone formation by reducing enhancer activity and *Nkx3-2* expression. Crucially, the F0 N1 enhancer showed activity that presages future long bone cartilage condensation in the limb (Fig. 4E). That is, the observed expression pattern recalls previous results that suggest that undetected early expression of *Nkx3-2* may mark the boundaries and size of limb bone precursors, including the tibia (*30*) (Fig. 4 – figure supplement 2C). Conversely, over-expression of *Nkx3-2* has been shown to cause shortened tibia (even loss) in mice (*31*, *32*). In humans, homozygous frameshift mutations in *NKX3-2* cause the rare disorder spondylo-megaepiphyseal-

metaphyseal dysplasia (SMMD; OMIM: 613330), which is characterized by short-trunk, long-limbed dwarfism and bow-leggedness (*33*). The affected bones in SMMD patients broadly correspond to the expression domains of the two novel N1 (limbs) and N3 (limbs and trunk) enhancers. Instead of wholesale loss of *Nkx3-2* expression, which would have been lethal in mice (*34*) or likely cause major defects similar to SMMD patients (*33*), our *in situ* hybridization data did not reveal qualitative differences in *Nkx3-2* expression domains between Ctrl or LS embryos (Fig. 4 – figure supplement 2B). Taken together, our results recapitulate the key features of a *cis*-acting mode of adaptation: *Nkx3-2* is a broadly expressed pleiotropic transcription factor that causes lethality when knocked out (*34*). We found no amino acid changes between the F0 and F17 alleles that could impact protein function. Rather, selection favoured changes in tissue-specific expression by modular enhancers. 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 mice, this represents a rare example of genetic dissection of a trait to the base-pair level.

Linking molecular mechanisms to evolutionary consequence

We next aimed to determine the evolutionary relevance of the *Nkx3-2* enhancer variants at the molecular and the population levels. At the strongly expressed N3/F0 "trunk and limb" enhancer, we note that the SNPs in the F17 selected allele lead to disrupted *Nkx3-1* and *Nkx3-2* binding sites [Fig. 4C & 5A; UNIPROBE database (*35*)]. This suggests that the selected SNPs may disrupt an auto-feedback loop to decrease *Nkx3-2* activity in the limb bud and trunk domain (Fig. 5A). Using a *GFP* transgenic reporter assay in stickleback fish embryos, we found that the mouse N1/F0 enhancer allele was capable of driving expression in the distal cells but not in the fin rays of the developing fins (Fig. 5A). This pattern recapitulates fin expression of *nkx3.2* in fish, which gives rise to endochondral radials (homologous to ulna/tibia in mice) (*36*). Our results suggest that strong selection may have favored the weaker N1/F17 and N3/F17 enhancer alleles in the context of the Longshanks selection regime despite the deep functional conservation of the F0 variants.

Using theory and simulations, we went beyond qualitative molecular dissection to quantitatively estimate the selection coefficient at the *Nkx3-2* locus and its

contribution to the total selection response in the Longshanks mice. We retraced the selective sweep of the *Nkx3-2* N1 and N3 alleles through targeted genotyping in 1569 mice across all 20 generations. The selected allele steadily increased from around 0.17 to 0.85 in LS1 and 0.98 in LS2 but fluctuated around 0.25 in Ctrl (Fig. 5B). We estimated that such a change of around 0.7 in allele frequency would correspond to a selection coefficient *s* of ~0.24 ± 0.12 at this locus (Fig. 5 – figure supplement 1B; see Supplementary Notes section on *"Estimating selection coefficient of the top-ranking locus, Nkx3-2, from changes in allele frequency"*). By extending our simulation framework to allow for a major locus against an infinitesimal background, we find that the *Nkx3-2* locus would contribute 9.4% of the total selection response (limits 3.6 – 15.5%; see Supplementary Notes section "*Estimating selection coefficient*" for details) in order to produce a shift of 0.7 in allele frequency over 17 generations. To avoid inflation stemming from estimating from outliers, we also independently estimated the contribution of the *Nkx3-2* locus using a linear mixed animal model based on the full genotyped series mentioned above (see Supplementary Notes section "*Estimating selection coefficient, animal model*" for details). Using this alternative approach, we estimated that each selected allele increases tibia length by 0.36% (*N*=1569, 95% conf. int.: .07% – 0.64%, *P* = 0.0171). Multiplying the effect with the increase in the allele frequency suggests that the *Nkx3-2* locus alone would account for approximately 4% of the overall 12.9% increase in tibia length. This lower estimate of around 4% is nonetheless within the bounds of the estimate from simulations. Together, both approaches agree that the *Nkx3-2* locus contribute substantially to the selection response.

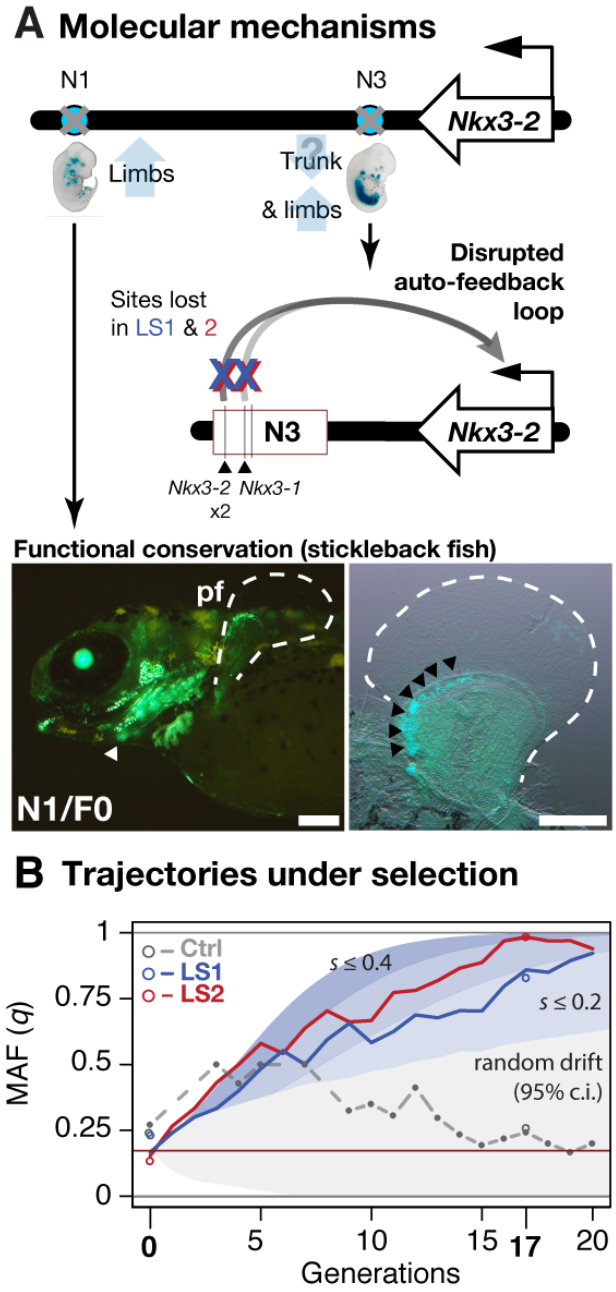


Fig. 5. Linking base-pair changes to rapid morphological evolution.

(**A**) At the *Nkx3-2* locus, we identified two long-range enhancers, N1 and N3 (circles), located 600 and 230 kbp away, respectively. During development, they drive partially overlapping expression domains in limbs (N1 and N3) and trunk (N3), which are body regions that may correlate positively (tibia length) and possibly negatively (trunk with body mass) with the Longshanks selection regime. For both enhancers, the selected F17 alleles carry loss-of-function variants (grey crosses). Two out of three SNPs in the N3 F17 enhancer are predicted to disrupt an auto-feedback loop, likely reducing *Nkx3-2* expression in the trunk and limb regions.

Conversely, the enhancer function of the strong N1 F0 allele is evolutionarily conserved in fishes, demonstrated by its ability to drive consistent GFP expression (green) in the pectoral fins (pf, outlined) and branchial arches (white arrowhead, left) in transgenic stickleback embryos at 11 days post-fertilization. The N1 enhancer can recapitulate *nkx3.2* expression in distal cells specifically in the endochondral radial domain in developing fins (black arrowheads, right). Scale bar: 250 μm for both magnifications. (**B**) Allele frequency of the selected allele (minor F0 allele, *q*) at N3 over 20 generations (blue: LS1; red: LS2; grey broken line: Ctrl; results from N1 were nearly identical due to tight linkage). Observed frequencies from genotyped generations in the Ctrl line are marked with filled circles. Dashed lines indicate missing Ctrl generations. Open circles at generations 0 and 17 indicate allele frequencies from whole genome sequencing. The allele frequency fluctuated in Ctrl due to random drift but followed a generally linear increase in the selected lines from around 0.17 to 0.85 (LS1) and 0.98 (LS2) by generation 17. Shaded contours mark expected allelic trajectories under varying selection coefficients starting from 0.17 (red horizontal line; the average starting allele frequency between LS1 and LS2 founders). The grey shaded region marks the 95% confidence interval under random drift.

2.2.4 Discussion:

A defining task of our time is to understand the factors that determine and constrain how small populations respond to sudden environmental changes. Here, we analyze the replicated and controlled Longshanks experiment to characterize the genomic changes that occur as small experimental populations respond to selection.

An important conclusion from the Longshanks experiment is that selection response can be steady and robust even in extremely bottlenecked populations. That is, we found that tibia length increased readily and repeatedly in response to selection even with as few as 14-16 breeding pairs per generation. The sustained response was possible because the lines were founded with enough standing variation, and generation 17 was still only a fraction of the way to the expected limit for the selection response at ~$2N_e$ generations (*37*), estimated here to be around 90

(see legend for Fig. 1 – figure supplement 2B; Supplementary Notes on "Estimating the selection coefficient"). Although other selective breeding studies using a similar base population of mice encountered selection limits at around generation 20-25 [possibly due to countervailing selection rather than loss of genetic variance for high voluntary wheel running behavior (*38*) and for nest-building behavior (*39*), here all evidence suggest that the Longshanks mice should continue to show increase in tibia length for many more generations.

The estimated N_e of 46 in the Longshanks experiment, while small, is comparable to those in natural populations like the Soay sheep (*40*), Darwin's finches (*41*) or Tasmanian Devils (*5*) (this last study documents a rapid and parallel evolutionary response to transmissible tumor). These populations span a wide range of time in sustained bottlenecking, from the most recent in Tasmanian devils, to likely many millions of years in Darwin's finches. Accordingly, we also expect very different dynamics during short- vs. long-term selection response: for a short bout of selection, such as the 20 generations analyzed in this study, selection response depends overwhelmingly on standing genetic variation, with little to no contribution from *de novo* mutations (*43*, *44*). Over the long term, however, *de novo* mutations would contribute increasingly. In the Longshanks experiment, we observe a robust early response to selection (Fig. 1B & Fig. 1 – figure supplement 1), and a gradual decrease in sequence diversity, consistent with the effect of drift (Fig. 1 – figure supplement 2B & Fig. 2 – figure supplement 1A, Supplementary File 2). There has long been broad empirical support for adaptation from standing genetic variation in nature (*24*, *42*, *45*) and breeding (*46*). At least in the short-run, our result demonstrating robust selection response in the Longshanks experiment provides grounds for some optimism regarding the ability of populations to respond rapidly to changes in their environment.

By combining pedigree records with sequencing of founder individuals, our data had sufficient detail to allow precise modeling of trait response, with predicted shifts in allele frequency distribution that closely matched our results (e.g. Fig. 1 – figure supplement 2D). Furthermore, we functionally validated loci that showed allele frequency shifts outside the model's predictions and found key enhancers of major effect. Connecting trait changes to allele frequency changes at specific loci has been a longstanding objective in selection experiments, with a number of remarkable early attempts (*47*). To date, we know of only a few studies that attempt to explicitly

link traits with changes in allele frequencies (*48*-*51*) and none have systematically tested the underlying architecture against an infinitesimal background. Here, our results imply a mixed genetic architecture with a few discrete loci of large effect amid an infinitesimal background. It remains to be seen whether other evolve-and-resequence (E&R) studies, with different selection pressures and population parameters may reveal similar results.

To put our finding of a mixed genetic architecture into perspective, it is worth noting that the infinitesimal model is still the most predictive model by far in practical quantitative genetics, for diverse domesticated species from cattle to crops, despite its intrinsically unrealistic assumptions (*52*-*54*). In general, current genomic data for many traits is consistent with a very large number of loci, each with a small effect. From a practical point of view, however, the use of an infinitesimal model does not preclude the presence or indeed the importance of few major effect loci. Rather, it simply assumes that they are rare enough to allow reasonable model fit [(*55*), page 878]. Here, we note that it is actually not clear how one might parameterize a generally applicable predictive oligogenic model with more than a single major effect locus. In this study, while we consider the most likely genetic architecture to be a small number of major effect loci together with a polygenic background, we cannot reject other alternative models that could also account for the observed response, such as an effectively infinitesimal model with linkage, as well as models with a few major trait loci.

Among other classical examples of complex traits, such as height or body weight, that may have been subjected to selection, we observe a range of genetic architectures in ways often tightly connected to their population size and/or selection history. Height in humans is often cited as the classical complex trait under possible selection of unknown (and much debated) intensity [see (*56*-*59*)]. It shows high heritability and a highly dispersed genetic architecture (with the top-ranked locus accounting for only 0.8% of the variation explained in cosmopolitan European populations) (*60*, *61*). In contrast, as few as 4 to 6 loci account for 83% and 50% of the variation in height in horses and dogs, respectively (*62*, *63*). In both horses and dogs, selection has been strong and sustained, and breed populations tend to be small. Interestingly, and in line with our experiment, the major allele at the *IGF1* locus stems from a standing genetic variant, despite many factors that may theoretically favor large-effect *de novo* mutations (*64*). In chickens, modern breeding

practice and selection from large populations yielded a highly polygenic genetic architecture for body weight, with some of the best empirical evidence for epistasis (*65*-*68*). Similarly, results from many selection experiments in *Drosophila* suggest that the genetic architecture underlying selection response may involve many genes (*69*-*72*). By contrast, the extreme tail of the effect size distribution (as inferred from Δz^2) from the Longshanks experiment appears to account for a substantial part of the selection response, presumably due to the combined effects of relatively low diversity in commercial mouse stock and the small founding populations. But unlike these previous QTL studies or selection experiments, in which either the genetic architecture of a trait or the selection value were estimated separately, sometimes from only few parental individuals or lines, E&R studies sample a much broader pool of alleles and continually compete them against each other. Thus, our approach allowed simultaneous inference of genetic architecture and distribution of effect sizes, is more likely to be representative of the population at large, and is more akin to genome-wide association studies (GWAS), except that here we can also directly connect a trait to its selective value and capture the trajectory of any given allele.

Parallel evolution is often seen as a hallmark for detecting selection (*25*, *73*-*75*). We investigated the factors that contribute to parallelism in allele frequency shifts over 17 generations by contrasting the two Longshanks replicates against the Control line. However, we observed little parallelism between selected lines and Ctrl, or between simulated replicates under selection, even though the simulated haplotypes were sampled directly from actual founders. This underscores that parallelism depends on both shared selection pressure (absent in Ctrl) *and* the availability of large-effect loci that confer a substantial selection advantage (absent under the infinitesimal model; Fig. 3; Fig. 3 – figure supplement 2). With increasing population size, selection would be better able to detect variants with more subtle effects. This would in turn lower the threshold beyond which the selection advantage of an allele would become deterministic, i.e., exhibit parallelism.

Through in-depth dissection of the *Nkx3-2* locus, our data show in fine detail how the selective value of standing variants depends strongly on the selection regime: the originally common F0 variant of the N1 enhancer show deep functional conservation and can evidently recapitulate fin *nkx3.2* expression in fishes (Fig. 5A). Yet, in the Longshanks experiment selection strongly favored the weaker allele (Fig. 5B). In fact, our molecular dissection of two loci show that both gain-of-function

(*Gli3*) and loss-of-function (*Nkx3-2*) variants could be favored by selection (Fig. 4E & 5A; Fig. 4 – figure supplement 1D). Through synthesis of multiple lines of evidence, our work uncovered the key role of *Nkx3-2*, which was not an obvious candidate gene like *Gli3* due to the lack of abnormal limb phenotype in the *Nkx3-2* knockout mice. To our surprise, the same loss of *NKX3-2* function in human SMMD patients manifests in opposite ways in different bone types as short trunk and long limbs (*33*). This matches the expression domains of our N1 (limb) and N3 (limb and trunk) enhancers (Fig. 5A). Evidently, in the absence of lethal coding mutations, the F17 haplotype was doubly beneficial at both enhancers for the limb and potentially also trunk target tissues under the novel selection regime in the Longshanks selection experiment. We estimate that these enhancer variants, along with any other tightly linked beneficial SNPs, segregate as a single locus, which in turn contributes ~10% of the overall selection response.

Despite our efforts to uncover the mechanism underlying the selective advantage of the *Nkx3-2* locus, much remains unknown. For example, it remains unclear how such a major allele could segregate in the general mouse stock (and as the reference C57BL/6J allele, no less). It could be that this allele has the same effect in the general mouse population but is conditionally neutral under non-selective breeding and simply escaped notice. However, our preliminary exploration in a panel of C57BL/6-by-DBA/2 ("BXD") mice suggested otherwise: mapping of tibia length or mineral density did not reveal this locus as a major QTL determining tibia length (unpublished data kindly provided by Weikuan Gu), suggesting that this allele's effect on tibia length may depend on the genetic background. Alternatively, the broader C57BL/6 allele could be linked to a compensatory mutation that became uncoupled among the founders of the Longshanks lines. Finally, although we do observe the specific N1 and N3 SNP positions as variable across the rodent and indeed the broader mammalian lineages, further work is needed to determine their effect, if any, on limb development.

2.2.5 Conclusion:

Using the Longshanks selection experiment and synthesizing theory, empirical data and molecular genetics, we show that it is possible to identify some of the individual SNPs that have contributed to the response to selection on morphology. In particular, discrete, large-effect loci are revealed by their parallel response. Further work should focus on dissecting the mechanisms behind the dynamics of selective sweeps and/or polygenic adaptation by re-sequencing the entire selection pedigree, testing how the selection response depends on the genetic architecture, and the extent to which linkage places a fundamental limit on our inference of selection. Improved understanding in these areas may have broad implications for conservation, rapid adaptation to climate change and quantitative genetics in medicine, agriculture and in nature.

2.2.6 Material 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 SRA database under accession number SRP165718 and GEO under GSE121564, GSE121565 and GSE121566. Non-sequence data have been deposited at Dryad. 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 (*13*). 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 (*76*). 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 “Major considerations”). 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 the two alleles at heterozygous sites in each individual to its two haplotypes at random, which minimizes linkage disequilibrium but is consistent with observed diploid genotypes (“min LD”); and assigning alleles at heterozygous sites in

each individual to the “reference” and “alternate” haplotype consistently within an interval, which maximises linkage disequilibrium (“max LD”) (Fig. 1 – figure supplement 2C). For details, see legend in Fig. 1 – figure supplement 2.

Significance thresholds

To obtain significance thresholds, we summarized 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. 1 – figure supplement 2E). This procedure controls for the effect of linkage and 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 Kits (Illumina, Inc., San Diego, USA) according to manufacturer’s recommendations or using equivalent *Tn5* transposase expressed in-house as previously described (*77*). 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 from software fastQC v0.10.1 (*78*); trimmomatic v0.33 (*79*); bwa v0.7.10-r789 (*80*); GATK v3.4-0-gf196186 modules BQSR, MarkDuplicates, IndelRealignment (*81*, *82*). 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 (*83*)], 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 (*16*). 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 (*83*) based on genotype posterior likelihoods using a genetic map interpolated from the mouse reference map (*76*) 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. 2 – figure supplement 1A).

Population genetics summary statistics

Summary statistics of the F0 and F17 samples were calculated genome-wide (Weir–Cockerham F_{ST}, π, heterozygosity, allele frequencies *p* and *q*) in adjacent 10 kbp windows or on a per site basis using VCFtools v0.1.14 (*84*). The summary statistic 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 (*85*) 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 signal 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 and LS2. 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 (*22*) 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 include genes whose regulatory sequences (e.g., enhancers), but not necessarily 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 mice (*15*), 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 excluded compound mutants or those conditional mutant phenotypes involving transgenes. To determine if the overlap with these genes were significant, we performed 1000 permutations of the core span using bedtools v2.22.1 shuffle with the -noOverlapping option (*86*) and excluding ChrY, ChrM and 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 (Supplementary File 3).

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 (*87*). We intersected the peak calls for the enhancer-associated marks H3K4me1 and H3K27ac and filtered out peaks overlapping promoters [H3K4me3 and promoter annotation according to the FANTOM5 Consortium (*88*)].

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 used 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 (*89*). 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 (*90*). 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 (*91*) or phyloP (*92*) score in ±250 bp flanking the SNP, calculated from a 60-way alignment between placental mammal genomes [downloaded from the UCSC Genome Browser (*93*)]. 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 matched these criteria are listed in

Supplementary File 4.

Association with human height loci

We downloaded the set of 697 SNPs associated with loci for human height (*61*) to test if these loci cluster with the selected loci in the Longshanks lines. In order to facilitate mapping to mouse co-ordinates, each human 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 (*93*). 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 in the selected Longshanks lines, the minimal distance to any of the mapped human loci was determined using bedtools closest with the -d option (*86*). When a region actually overlapped, a distance of 0 bp was assigned. To generate a 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 followed an approximately normal distribution, with observed results falling completely below the range of permuted results, i.e., closer to height-associated human SNPs.

***In situ* hybridization**

Detection of specific gene transcripts were performed as previously described in (*94*). Probes against *Nkx3-2*, *Rab28*, *Bod1l* and *Gli3* were amplified from cDNA from wildtype C57BL/6NJ mouse embryos (Supplementary File 5). 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 (*28*), with the following modifications. To minimize endogenous protease activity, cells were strictly limited to 5 + 5 minutes of collagenase A treatment at 37 ºC, 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 (*95*). 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 (*96*). Peaks were called using MACS14 v.2.1 (*97*)**.**

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 (*98*). 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; Supplementary File 6]. 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 (*99*); bwa v0.7.10-r789 (*80*) ; samtools v1.2 (*100*); bedtools (*86*) and R v3.2.0 (*85*). Alignments were filtered for ENCODE blacklisted regions (*101*) 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 (Supplementary File 7). 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 in mice

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 (*102*).

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 (*103*). 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; Supplementary File 8). 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.

Transgenic reporter assays in stickleback fish

In sticklebacks, transgenic reporter assays were carried out using the reporter construct pBHR (*74*). The reporter consists of a zebrafish *heat shock protein 70* (*Hsp70*) promoter followed by an *eGFP* reporter gene, with the entire reporter cassette flanked by *tol2* transposon sequences for transposase-directed genomic integration. The *Nkx3-2* N1/F0 enhancer allele was cloned as tandem duplicates using the *NheI* and *EcoRV* restriction sites upstream of the *Hsp70* promoter. Enhancer orientation and sequence was confirmed by Sanger sequencing. Transient transgenic stickleback embryos were generated by co-microinjecting the plasmid (final concentration: 10 ng/µl) and *tol2* transposase mRNA (40 ng/µl) into freshly fertilized eggs at the one-cell stage as described in (*74*).

2.2.7 References:

1. A. A. Hoffmann, C. M. Sgrò, Climate change and evolutionary adaptation. *Nature* **470**, 479-85 (2011), doi: 10.1038/nature09670.
2. C. Darwin, *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
3. P. D. Gingerich, Rates of evolution on the time scale of the evolutionary process. *Genetica* **112-113**, 127-44 (2001).
4. T. Garland, M. R. Rose, *Experimental evolution : concepts, methods, and applications of selection experiments* (University of California Press, Berkeley, 2009).
5. P. D. Keightley. L. Bünger, U. Renne, R. C. Buis, in *Encyclopedia of Genetics*. E. C. R. Reeve, Eds. (Fitzroy Dearborn Publishers, Chicago, 2001), pp. 337-360.
6. B. Walsh, M. Lynch, The infinitesimal model and its extensions in *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018), pp. 1504.
7. J. M. Smith, J. Haigh, The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23-35 (1974), doi: 4407212.
8. N. H. Barton, Linkage and the limits to natural selection. *Genetics* **140**, 821-41 (1995).
9. S. P. Otto, N. H. Barton, Selection for recombination in small population. *Evolution* **55**, 1921-31 (2001), doi: 10.1111/j.0014-3820.2001.tb01310.x.
10. D. B. Weissman, N. H. Barton, Limits to the rate of adaptive substitution in sexual populations. *PLoS Genet.* **8**, e1002740. (2012), doi: 10.1371/journal.pgen.1002740.
11. J. F. Crow, M. Kimura, Evolution in sexual and asexual populations. *Am. Nat.* **99**, 439-50 (1965).
12. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269-94 (1966).
13. M. Marchini, L. M. Sparrow, M. N. Cosman, A. Dowhanik, C. B. Krueger, B. Hallgrimsson, C. Rolian, Impacts of genetic correlation on the independent evolution of body mass and skeletal size in mammals. *BMC Evol. Biol.* **14**, 258. (2014), doi: 10.1186/s12862-014-0258-0.

14. F. Petit, K. E. Sears, N. Ahituv, Limb development: a paradigm of gene regulation. *Nat. Rev. Genet.* **18**, 245-58 (2017), doi: 10.1038/nrg.2016.167.
15. M. Marchini, C. Rolian, Artificial selection sheds light on developmental mechanisms of limb elongation. *Evolution* **72**, 825-37 (2018), doi: 10.1111/evo.13447.
16. B. Yalcin, J. Nicod, A. Bhomra, S. Davidson, J. Cleak, L. Farinelli, M. Østerås, A. Whitley, W. Yuan, X. Gan, M. Goodson, P. Klenerman, A. Satpathy, D. Mathis, C. Benoist, D. J. Adams, R. Mott, J. Flint, Commercially available outbred mice for genome-wide association studies. *PLoS Genet.* **6**, e1001085. (2010).
17. A. P. Hendry, M. T. Kinnison, Perspective: the pace of modern life: measuring rates of contemporary microevolution. *Evolution* **53**, 1637-53 (1999), doi: 10.1111/j.1558-5646.1999.tb04550.x.
18. P. R. Grant, B. R. Grant, Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707-11 (2002), doi: 10.1126/science.1070315.
19. D. S. Falconer, T. F. Mackay, *Introduction to quantitative genetics (4th Ed.)* (Pearson, London, 1996).
20. W. Pitchers, J. B. Wolf, T. Tregenza, J. Hunt, I. Dworkin, Evolutionary rates for multivariate traits: the role of selection and genetic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130252 (2014), doi: 10.1098/rstb.2013.0252.
21. J. B. S. Haldane, *The causes of evolution,* (Longmans, Green and Co., London; New York, 1932).
22. 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.
23. 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.
24. 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.
25. 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.

26. J. K. Kelly, K. Hughes, An examination of the evolve-and-resequence method using *Drosophila simulans*. *bioRxiv* (2018), doi: 10.1101/337188.
27. H. A. Orr, The probability of parallel evolution. *Evolution* **59**, 216-20 (2005).
28. 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.
29. 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.
30. 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.
31. 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.
32. 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.
33. J. Hellemans, M. Simon, A. Dheedene, Y. Alanay, E. Mihci, L. Rifai, A. Sefiani, Y. van Bever, M. Meradji, A. Superti-Furga, G. Mortier, Homozygous inactivating mutations in the *NKX3-2* gene result in spondylo-megaepiphyseal-metaphyseal dysplasia. *Am. J. Hum. Genet.* **85**, 916-22 (2009), doi: 10.1016/j.ajhg.2009.11.005.
34. 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).
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. P. L. Crotwell, P. M. Mabee, Gene expression patterns underlying proximal-distal skeletal segmentation in late-stage zebrafish, *Danio rerio*. *Dev. Dyn.* **236**, 3111-28 (2007), doi: 10.1002/dvdy.21352.
37. A. Robertson, A theory of limits in artificial selection. *Proc. Roy. Soc. Lond. B Biol. Sci.* **153**, 234-49 (1960).
38. V. Careau, M. E. Wolak, P. A. Carter, T. Garland. Limits to behavioral evolution: the quantitative genetics of a complex trait under directional selection. *Evolution* **67**, 3102-19 (2013) doi: 10.1111/evo.12200.
39. A. Bult, C. B. Lynch. Breaking through artificial selection limits of an adaptive behavior in mice and the consequences for correlated responses. *Behav Genet* **30**, 193-206 (2000).
40. A. F. McRae, J. M. Pemberton, P. M. Visscher, Modeling linkage disequilibrium in natural populations: the example of the Soay sheep population of St. Kilda, Scotland. *Genetics* **171**, 251-8 (2005), doi: 10.1534/genetics.105.040972.
41. P. R. Grant, B. R. Grant, Demography and the genetically effective sizes of two populations of Darwin's Finches. *Ecology* **73**, 766-84 (1992), doi: 10.2307/1940156.
42. B. Epstein, M. Jones, R. Hamede, S. Hendricks, H. McCallum, E. P. Murchison, B. Schönfeld, C. Wiench, P. Hohenlohe, A. Storfer, Rapid evolutionary response to a transmissible cancer in Tasmanian devils. *Nat. Commun.* **7**, 12684. (2016), doi: 10.1038/ncomms12684.
43. W. G. Hill. Predictions of response to artificial selection from new mutations. *Genet. Res.* **40**, 255-78 (1982).
44. K. E. Weber, L. T. Diggins. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. *Genetics* **125**, 585-97 (1990).
45. 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.
46. 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.

47. P. D. Keightley, T. Hardge, L. May, G. Bulfield. A genetic map of quantitative trait loci for body weight in the mouse. *Genetics* **142**, 227-35 (1996).
48. D. Kessner, J. Novembre. Power analysis of artificial selection experiments using efficient whole genome simulation of quantitative traits. *Genetics* **199**, 991-1005 (2015), doi: 10.1534/genetics.115.175075.
49. D. P. Rice, J. P. Townsend. A test for selection employing quantitative trait locus and mutation accumulation data. *Genetics* **190**, 1533-45 (2012), doi: 10.1534/genetics.111.137075.
50. N. Chen, I. Juric, E. J. Cosgrove, R. Bowman, J. W. Fitzpatrick, S. J. Schoech, A. G. Clark, G. Coop. Allele frequency dynamics in a pedigreed natural population. *Proc. Nat. Acad. Sci. of U. S. A.* **116**, 2158-64 (2019), doi: 10.1073/pnas.1813852116.
51. S. V. Nuzhdin, C. L. Dilda, T. F. Mackay. The genetic architecture of selection response. Inferences from fine-scale mapping of bristle number quantitative trait loci in *Drosophila melanogaster*. *Genetics* **153**, 1317-31 (1999).
52. W. G. Hill, M. E. Goddard, P. M. Visscher. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* **4**, e1000008 (2008), doi: 10.1371/journal.pgen.1000008.
53. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Traits* (Sinauer, 1998).
54. W. G. Hill, X. S. Zhang. On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. *Genetics* 190, 1131-7 (2012) doi: 10.1534/genetics.111.135681.
55. B. Walsh, M. Lynch, The infinitesimal model and its extensions in *Evolution and Selection of Quantitative Traits* (Oxford University Press, 2018), pp. 1504.
56. M. C. Turchin, C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, J. N. Hirschhorn. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat. Genet.* **44**, 1015-9 (2012) doi: 10.1038/ng.2368.
57. J. J. Berg, G. Coop. A population genetic signal of polygenic adaptation. *PLoS Genet.* **10**, e1004412. (2014), doi: 10.1371/journal.pgen.1004412.

58. N. Barton, J. Hermisson, M. Nordborg. Why structure matters. *eLife* **8** (2019), doi: 10.7554/eLife.45380.

59. M. Sohail, R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, M. C. Turchin, C. W. Chiang, J. Hirschhorn, M. J. Daly, N. Patterson, B. Neale, I. Mathieson, D. Reich, S. R. Sunyaev. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8** (2019), doi: 10.7554/eLife.39702.

60. M. N. Weedon, G. Lettre, R. M. Freathy, C. M. Lindgren, *et al.* A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat. Genet.* **39**, 1245-50 (2007), doi: 10.1038/ng2121.

61. A. R. Wood, T. Esko, J. Yang, S. Vedantam, *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173-86 (2014), doi: 10.1038/ng.3097.

62. S. Makvandi-Nejad, G. E. Hoffman, J. J. Allen, E. Chu, E. Gu, A. M. Chandler, A. I. Loredo, R. R. Bellone, J. G. Mezey, S. A. Brooks, N. B. Sutter. Four loci explain 83% of size variation in the horse. *PLoS ONE* **7**, e39929. (2012), doi: 10.1371/journal.pone.0039929.

63. M. Rimbault, H. C. Beale, J. J. Schoenebeck, B. C. Hoopes, J. J. Allen, P. Kilroy-Glynn, R. K. Wayne, N. B. Sutter, E. A. Ostrander. Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res.* **23**, 1985-95 (2013), doi: 10.1101/gr.157339.113.

64. N. B. Sutter, C. D. Bustamante, K. Chase, M. M. Gray, *et al.* A single *IGF1* allele is a major determinant of small size in dogs. *Science* **316**, 112-5 (2007), doi: 10.1126/science.1137045.

65. Ö. Carlborg, L. Jacobsson, P. Ahgren, P. Siegel, L. Andersson. Epistasis and the release of genetic variation during long-term selection. *Nat. Genet.* **38**, 418-20 (2006), doi: 10.1038/ng1761.

66. P. Wahlberg, O. Carlborg, M. Foglio, X. Tordoir, A. C. Syvänen, M. Lathrop, I. G. Gut, P. B. Siegel, L. Andersson. Genetic analysis of an F(2) intercross between two chicken lines divergently selected for body-weight. *BMC Genomics* **10**, 248. (2009), doi: 10.1186/1471-2164-10-248.

67. C. J. Rubin, M. C. Zody, J. Eriksson, J. R. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallböök, F. Besnier, Ö. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-

Toh, L. Andersson. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587-91 (2010), doi: 10.1038/nature08832.

68. M. E. Pettersson, A. M. Johansson, P. B. Siegel, Ö. Carlborg. Dynamics of adaptive alleles in divergently selected body weight lines of chickens. *G3 (Bethesda)* **3**, 2305-12 (2013), doi: 10.1534/g3.113.008375.
69. A. R. Jha, C. M. Miles, N. R. Lippert, C. D. Brown, K. P. White, M. Kreitman. Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in *Drosophila melanogaster*. *Mol. Biol. Evol.* **32**, 2616-32 (2015), doi: 10.1093/molbev/msv136.
70. R. G. Reeves, D. Tautz. Automated phenotyping indicates pupal size in *Drosophila* is a highly heritable trait with an apparent polygenic basis. *G3 (Bethesda)* **7**, 1277-86 (2017), doi: 10.1534/g3.117.039883.
71. P. Orozco-terWengel, M. Kapun, V. Nolte, R. Kofler, T. Flatt, C. Schlötterer. Adaptation of *Drosophila* to a novel laboratory environment reveals temporally heterogeneous trajectories of selected alleles. *Mol. Ecol.* **21**, 4931-41 (2012), doi: 10.1111/j.1365-294X.2012.05673.x.
72. T. L. Turner, A. D. Stewart, A. T. Fields, W. R. Rice, A. M. Tarone. Population-based resequencing of experimentally evolved populations reveals the genetic basis of body size variation in Drosophila melanogaster. *PLoS Genet*. **7**, e1001336. (2011), doi: 10.1371/journal.pgen.1001336.
73. 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.
74. 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.
75. 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.

The following references appeared in Material and Methods only

76. 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).
77. 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.
78. FastQC A quality control tool for high throughput sequence data. Accessed: 5 Aug. 2016, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
79. 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.
80. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-95 (2010), doi: 10.1093/bioinformatics/btp698.
81. 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.
82. 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.
83. 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.
83. B. L. Browning, S. R. Browning, Genotype imputation with millions of reference samples. *Am. J. Hum. Genet.* **98**, 116-26 (2016).

84. 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 Analysis Group. The variant call format and VCFtools. *Bioinformatics* **27**, 2156-8 (2011), doi: 10.1093/bioinformatics/btr330.

85. R. C. Team, R: A language and environment for statistical computing. *R: A language and environment for statistical computing* (2015).

86. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

87. 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.

88. FANTOM5 Consortium, A promoter-level mammalian expression atlas, *Nature* **507**, 462. (2014).

89. J. H. Finger, C. M. Smith, T. F. Hayamizu, I. J. McCright, *et al.*, The mouse Gene Expression Database (GXD): 2017 update. *Nucleic Acids Res.* **45**, D730-6 (2017).

90. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).

91. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110-21 (2010).

92. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034-50 (2005).

93. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, D. Haussler. The human genome browser at UCSC. *Genome Res*. **12**, 996-1006 (2002).

94. S. D. M. Brown, P. Chambon, M. H. de Angelis, EM PReSS: standardized phenotype screens for functional annotation of the mouse genome. *Nat. Genet.* 37, 1155-1155 (2005).

95. P. Milani, R. Escalante-Chong, B. C. Shelley, N. L. Patel-Murray, X. Xin, M. Adam, B. Mandefro, D. Sareen, C. N. Svendsen, E. Fraenkel, Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Sci. Rep.* **6**, 25474. (2016).

96. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-9 (2012).

97. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. (2008).

98. T. Sexton, S. Kurukuti, J. A. Mitchell, D. Umlauf, T. Nagano, P. Fraser, Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. *Nat Protoc.* **7**, 1335-50 (2012), doi: 10.1038/nprot.2012.071.

99. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, pp-10 (2011).

100. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools, *Bioinformatics* **25**, 2078-9 (2009), doi: 10.1093/bioinformatics/btp352.

101. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012), doi: 10.1038/nature11247.

102. R. J. DiLeone, G. A. Marcus, M. D. Johnson, D. M. Kingsley, Efficient studies of long-distance Bmp5 gene regulation using bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1612-7 (2000).

103. D. P. Mortlock, C. Guenther, D. M. Kingsley, A general approach for identifying distant regulatory elements applied to the *Gdf6* gene. *Genome Res.* **13**, 2069-81 (2003).

104. J. Hadfield. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J. Stat. Software*, *Articles* **33**, 1-22 (2010), doi: 10.18637/jss.v033.i02.

105. P. de Villemereuil, *Estimation of a biological trait heritability using the animal model: How to use the R package MCMCglmm.* (Accessed: 21 Mar. 2019, http://devillemereuil.legtux.org/wp-content/uploads/2012/12/tuto_en.pdf).

106. S. B. Carroll, Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36 (2008), doi: 10.1016/j.cell.2008.06.030.
107. R. Mo, A. M. Freer, D. L. Zinyk, M. A. Crackower, J. Michaud, H. H. Heng, K. W. Chik, X. M. Shi, L. C. Tsui, S. H. Cheng, A. L. Joyner, C. Hui, Specific and redundant functions of *Gli2* and *Gli3* zinc finger genes in skeletal patterning and development. *Development* **124**, 113-23 (1997).
108. T. Nakamura, J. Klomp, J. Pieretti, I. Schneider, A. R. Gehrke, N. H. Shubin, Molecular mechanisms underlying the exceptional adaptations of batoid fins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15940-5 (2015), doi: 10.1073/pnas.1521818112.
109. 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).
110. 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.
111. 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.

2.2.8 Acknowledgements:

We thank Felicity Jones for input into experimental design, helpful discussion and improving the manuscript. We thank the Rolian, Barton, Chan and Jones Labs members for support, insightful scientific discussion and improving the manuscript. We thank the Rolian lab members, the Animal Resource Centre staff at the University of Calgary, and MPI Dresden Animal Facility staff for animal husbandry. We thank Derek Lundberg for help with library preparation automation. We thank Christa Lanz, Rebecca Schwab and Ilja Bezrukov for assistance with high-throughput sequencing and associated data processing; Andre Noll for high-

performance computing support; the MPI Tübingen IT team for computational support. We thank Felicity Jones and the Jones Lab for help with stickleback microinjections. pRS16 was a gift from François Spitz. We thank Mirna Marinič for creating an earlier version of the transgenic reporter plasmid. We are indebted to Gemma Puixeu Sala, William G. Hill, Peter Keightley for input and discussion on data analysis and simulation. We are also indebted to Stefan Mundlos, Przemko Tylzanowki, Weikuan Gu for suggested experiments and sharing unpublished data. We thank Sean B. Carroll, Andrew Clark, John Kelly (reviewer), David Kingsley, Jonathan Pritchard, Matthew Rockman (reviewer), Gregory Wray, and Magnus Nordborg (reviewer) for thoughtful input that has greatly improved our manuscript. J.P.L.C. is supported by the International Max Planck Research School “From Molecules to Organisms”. S.B. and N.B. are supported by IST Austria. C.R. is supported by Discovery Grant #4181932 from the Natural Sciences and Engineering Research Council of Canada and by the Faculty of Veterinary Medicine at the University of Calgary. Y.F.C. is supported by the Max Planck Society.

2.2.9 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 provided funding for the research but no other competing interests.

2.2.10 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 mattings). 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. 1 – figure supplement 2D). 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 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 10,000), 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. 1 – figure supplement 2B 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. 1 – figure supplement 2B shows 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). 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. 1 – figure supplement 2D, SNP classes 1/56 and 4/56). However, it would 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 (*23*, *71*).

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. 3 – figure supplement 2). 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.

Heritability estimate by an animal model

We estimated heritability using linear mixed effect "animal models" with maximum likelihood (Fig. 1 – figure supplement 1D) in the R package MCMCglmm v2.5 [*104*; following guide by (*105*)]. Because the animal model makes inference of the parameter estimates to the base population, to compare heritability as it changed over time we estimated heritability in blocks of 5 generations F0-4, 5-9, 10-14, and 15-19, separately for each selected line. In testing each block, we used the full pedigree to build the relationship matrix but only phenotypes from the individuals in those generations. As an alternative, we tested each block with a truncated pedigree, in which the first generation of each block is treated as unrelated (i.e., the base population). The two methods produced similar results. In all analyses, we standardized the composite trait $ln(\boldsymbol{TB}^{-0.57})$ ($\boldsymbol{T}$ = tibia length in mm; $\boldsymbol{B}$ = cube-root body mass in $\sqrt[3]{}$g; see *Simulating selection response: infinitesimal model with linkage* in main text) within each generation and line to account for fluctuations in mean and variance (*38*). The phenotypic variance was partitioned as V_P = fixed effects + V_A + V_R, where fixed effects were sex, age, and litter size, V_A was additive genetic variance, and V_R was residual variance. Heritability was estimated as $h^2 = V_A / (V_A + V_R)$.

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. 3 – figure supplement 1A). The "limb/digital/tail" category of affected anatomical systems in the Mouse Genomic Informatics Gene Expression Database (*59*) 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. 3 – figure supplement 1A). 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. 3 – figure supplement 1A, 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) (*87*). We found likewise an enrichment throughout the range of significance cut-offs (Fig. 3 – figure supplement 1A). 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 the 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. 3 – figure supplement 1B; 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. 3 – figure supplement 1B): 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. 3 – figure supplement 1B). 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; Supplementary File 3; c.f. Supplementary File 4 for coding changes). This is a key prediction of the "*cis*-regulatory hypothesis", especially in its original scope on morphological traits (*106*).

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; Supplementary File 4). 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.

Molecular dissection of Gli3, *a candidate limb regulator, reveal gain-of-function* cis*-acting changes*

We anticipated that genes related to major limb patterning, like *Gli3*, may contribute to the selection response (*107*, *108*). We thus performed an in-depth molecular dissection of *Gli3*, an important early limb developmental regulator on chromosome 13 (Chr13; Fig. 4 – figure supplement 1A). 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 (*109*) and considering that *Gli3* could be among the many minor loci in the polygenic background contributing to the selection response in LS1.

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) (*22*) and identified putative enhancers using chromatin modification marks from the ENCODE project and our own ATAC-Seq data (Fig. 4 – figure supplement 1B) (*28*, *90*). 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. 4 – figure supplement 1C). 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). We found that only the F17 LS1 allele was able to drive consistent *lacZ* expression in the developing limb buds (Fig. 4 – figure supplement 1D). 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. 4 – figure supplement 1A). Furthermore, substitution of the enhancer sequence with the F0 allele (10 differences out of 956 or 960 bp) abolished *lacZ* expression (Fig. 4 – figure supplement 1D). 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 favoured because it drove stronger expression of *Gli3*, a gene essential for tibia development (*110*) [but see (*111*)].

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. 5 – figure supplement 1A). 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. 5 – figure supplement 1A, 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, $log\frac{q}{p}$ changes linearly with time, at a rate equal to the selection coefficient, s. Therefore, a naive estimate of selection is given by $\hat{s} = \frac{1}{T}\log\left[\frac{q_{17}}{p_{17}}\frac{p_0}{q_0}\right]$ (*21*) thus, $\hat{s}$ =0.19, 0.32 for *q* in LS1, LS2, and averages 0.26. Here, males and females with longest tibia 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 drift (F~0.17). Thus, assuming constant selection is reasonable (Fig. 5B), 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 (note that due to differences in estimation methodology, this N_e differs slightly from that mentioned in Fig. 1 – figure supplement 2 but is largely consistent). 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. 5 – figure supplement 1B).

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. 5 – figure supplement 1C, 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. 5 – figure supplement 1C, 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. 1 – figure supplement 2D, 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.

Estimating the contribution of the* Nkx3-2 *locus using an animal model

We used a linear mixed "animal model" to estimate the effect of the enhancer N3 (of the major locus in Nkx3-2) on the composite selected trait $ln(\boldsymbol{TB}^{-0.57})$, see L. 129 and Fig. 1 – figure supplement 2A. The model was:

V_P = fixed effects + V_A + V_R

where:

fixed effects = sex, generation, litter size (i.e., number of siblings in family), genotype at N3 (0, 1, or 2 copies of F17 allele), and replicate line

V_A = additive genetic variance

V_R = residual variance

We found a small but significant effect of the genotype at enhancer N3 on the composite trait (mean effect = 0.0036; 95% CI: 0.00069–0.0064; *P*=0.017). Given the same body mass $\boldsymbol{B}$, the mean effect corresponds to 0.36% increase in tibia length per copy of the F17 allele, or ~1% of the variance in tibia length at generation F01. The observed increase of this allele from ~0.18 to 0.91, averaged over the two lines, implies that it accounts for ~4% of the total selection response. This is within the confidence limits in the main text, based on the change in SNP frequency (3.6–15.5%) and note that the latter may be biased upwards by ascertainment. However, the exact effect of the allele is difficult to pinpoint in any given generation or population due the nature of the composite trait and change in variance in the composite trait over generations.

2.2.11 Supplementary Figures:

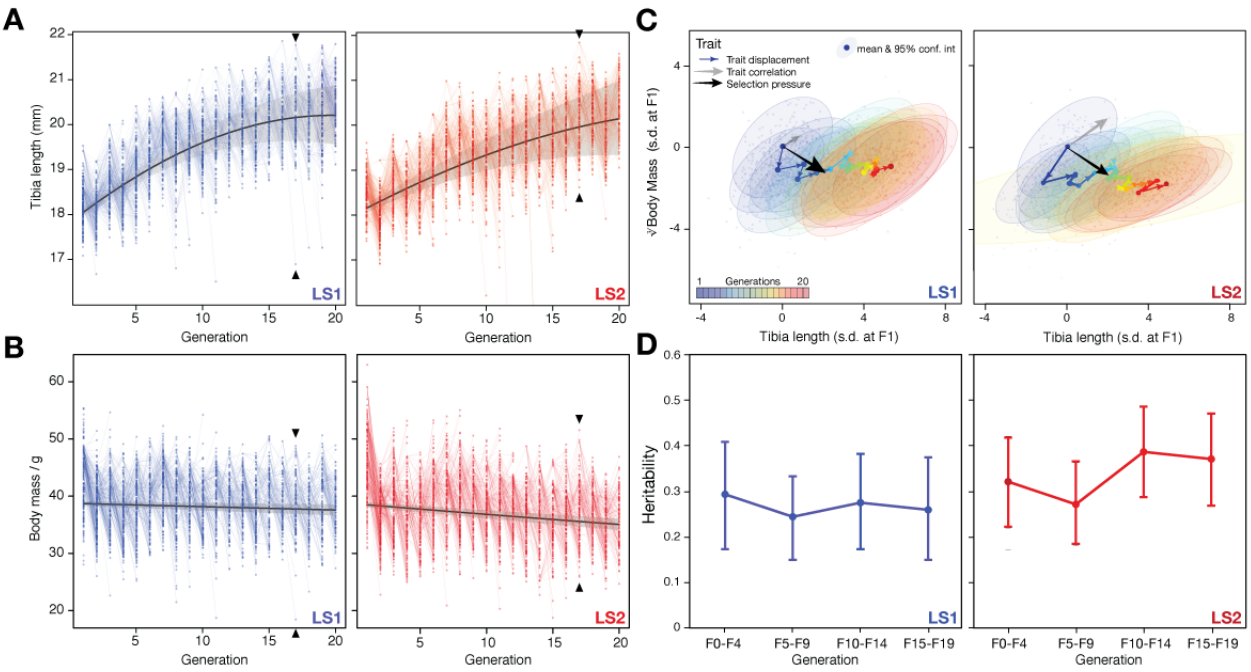


Fig. 1 – figure supplement 1. Artificial selection allowed detailed reconstruction of selection parameters. Rapid response to selection produced mice with progressively longer tibiae (**A**) and slightly lower body weight (**B**) within 20 generations. Having complete records throughout the selection 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 (*13*) 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. (**D**) Despite persistent and strong selection, heritability for the composite trait $ln(\boldsymbol{TB}^{-0.57})$ ($\boldsymbol{T}$ = tibia length in mm; $\boldsymbol{B}$ = cube-root body mass in $\sqrt[3]{}$g) (see *Simulating selection response: infinitesimal model with linkage* in main text) was maintained over 20 generations. Heritability was estimated by a linear mixed "animal model" in which the phenotypic variance was partitioned as V_P = fixed effects + V_A + V_R, where fixed effects were sex, age, and litter size, V_A was additive genetic variance, and V_R was residual variance. Heritability was estimated as $h^2 = V_A / (V_A + V_R)$. Each tested block used the full pedigree but only phenotypic information from individuals within the block. We tested an alternate model for each block using truncated pedigrees wherein the first generation of each block was assumed to be unrelated, but found similar results.

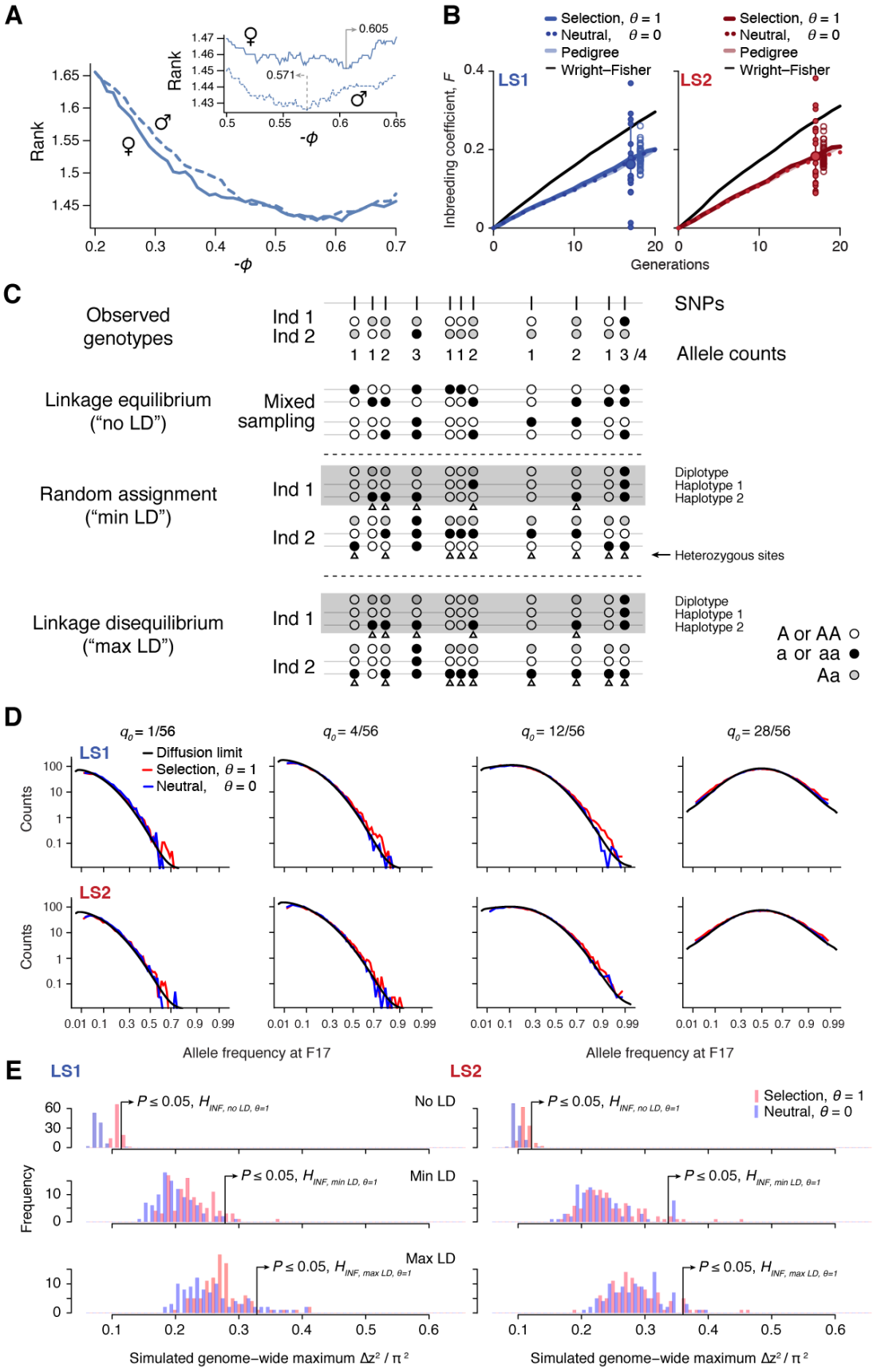

A
Rank
-φ
♀
♂
0.605
0.571
B
Selection, θ = 1
Neutral, θ = 0
Pedigree
Wright–Fisher
LS1
LS2
Inbreeding coefficient, F
Generations
C
Observed genotypes
Ind 1
Ind 2
SNPs
Allele counts
Linkage equilibrium ("no LD")
Mixed sampling
Random assignment ("min LD")
Diplotype
Haplotype 1
Haplotype 2
Heterozygous sites
Linkage disequilibrium ("max LD")
A or AA
a or aa
Aa
D
q_0 = 1/56
q_0 = 4/56
q_0 = 12/56
q_0 = 28/56
Diffusion limit
Counts
Allele frequency at F17
E
No LD
Min LD
Max LD
Frequency
$P \le 0.05$, $H_{INF,\ no\ LD,\ \theta=1}$
$P \le 0.05$, $H_{INF,\ min\ LD,\ \theta=1}$
$P \le 0.05$, $H_{INF,\ max\ LD,\ \theta=1}$
Simulated genome–wide maximum $\Delta z^2 / \pi^2$

Fig. 1 – figure supplement 2. Simulating selection on pedigrees. This figure summarizes the results from our analyses to determine parameters used during the simulation. For full detail, see Supplementary Notes, section "*Major considerations in constructing the simulations*". (**A**) Finding the correct ϕ value for the composite trait $ln(\boldsymbol{TB}^{\phi})$. In each simulated family, offsprings are split by sex and ranked by their composite trait. Due to occasional use of back-up crosses, the average rank of actual breeders is greater than 1. We vary ϕ to find the value where actual breeders in the LS lines have the best (lowest) rank. We find ϕ = -0.571 to show the best match for males and 0.605 for females. For subsequent analyses we set ϕ to be -0.57. (**B**) 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 chromosomes. 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 (*46*). 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. (**C**) 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"). (**D**) Simulated vs. expected allele frequency shifts. The distribution of minor allele frequencies q_0 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. (**E**) 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 **C** 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. Δz^2 is 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. 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.

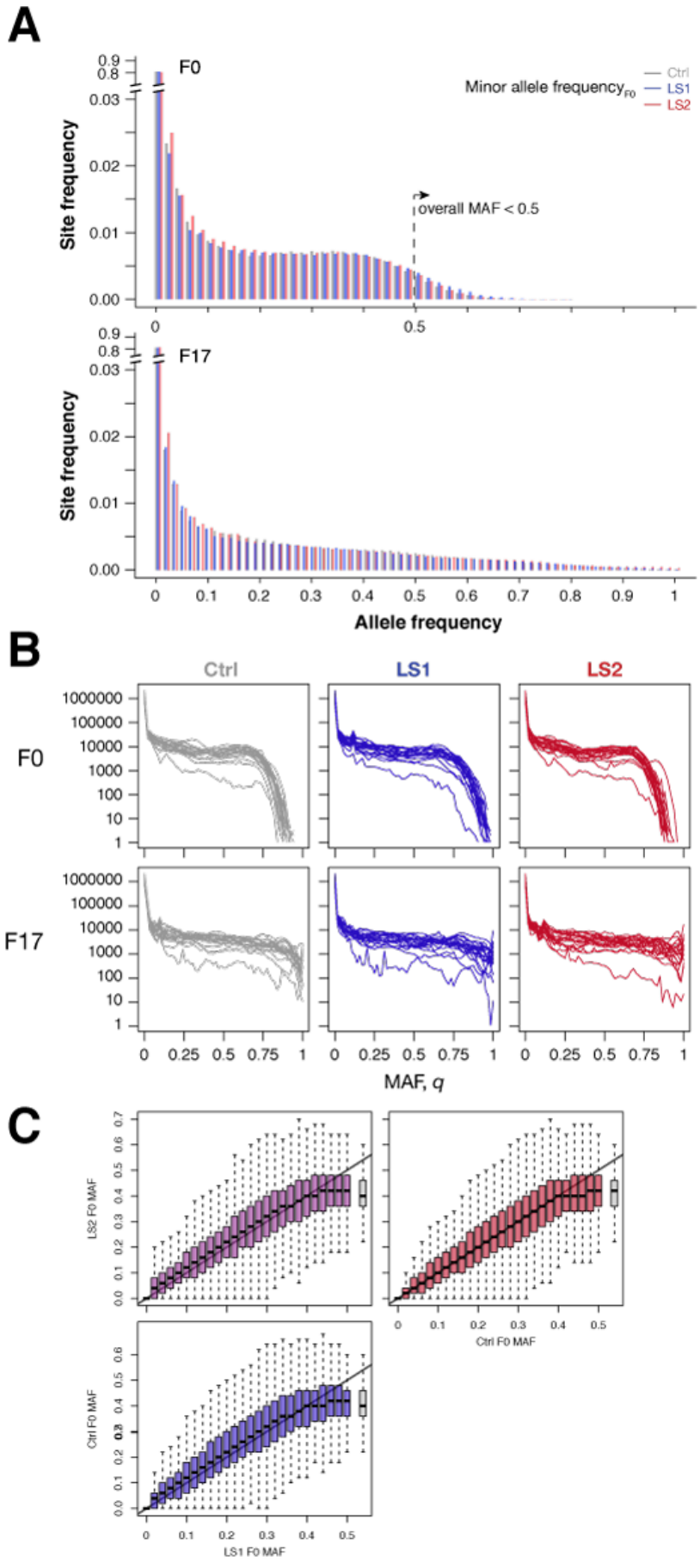

A
F0
Minor allele frequency$_{F0}$
Ctrl
LS1
LS2
overall MAF < 0.5
Site frequency
F17
Allele frequency
B
Ctrl
LS1
LS2
F0
F17
MAF, q
C
LS2 F0 MAF
Ctrl F0 MAF
LS1 F0 MAF

Fig. 2 – figure supplement 1. 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 within each generation. The Control population was mostly intermediate in the decay in the rarer alleles. After 17 generations, the same alleles were generally more spread out, leading to more broadly distributed spectra. There was again little overall difference between the Longshanks and Control lines. (**B**) Variations between chromosomes (separate same-colored lines) shown in each population 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. (**C**) 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 (blue), LS2 (red); 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 in panel A, 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.

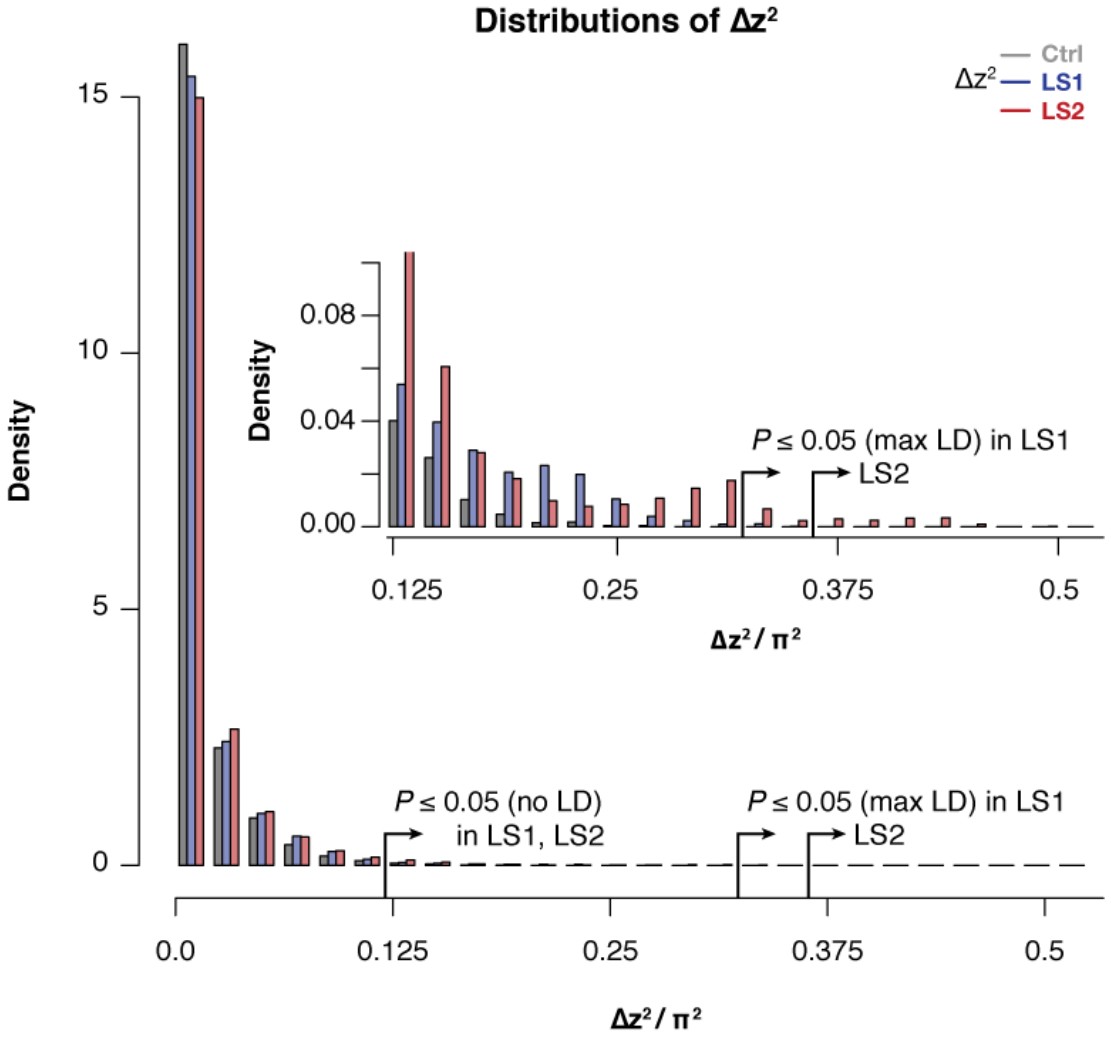


Fig. 2 – figure supplement 2. Selected lines showed more extreme values of Δz^2 than the Control line. Histogram of within-line Δz^2 values in 10 kbp windows across the genome in the LS1, LS2, and Control. Overall similarity is high across all 3 lines, but there was an excess of large Δz^2 value starting 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.

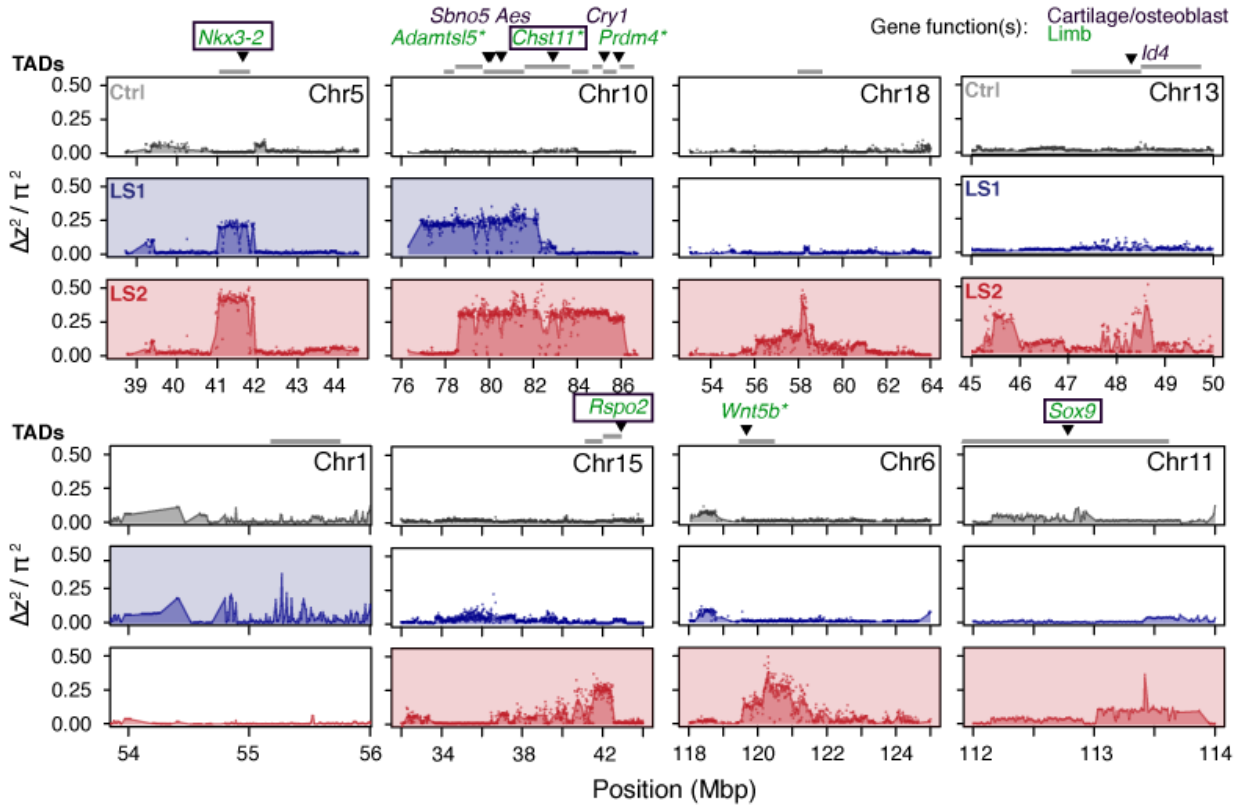


Fig. 2 – figure supplement 3. Detailed Δz^2 profiles at the 8 Longshanks significant loci. For each significant locus, Δz^2 profiles are shown for Ctrl (grey), LS1 (blue) and LS2 (red). Plots are shaded if the locus is significant in a given line. TADs within 250 kbp of the significant signals are shown as grey bars above each locus. Above the TADs are highlighted genes whose knockout phenotypes belong to the following categories: “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), bone development (purple) or both (boxed). Gene symbols marked by asterisks (*) have specifically 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 Notes, section *“Enrichment for genes with functional impact on limb development”* 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.

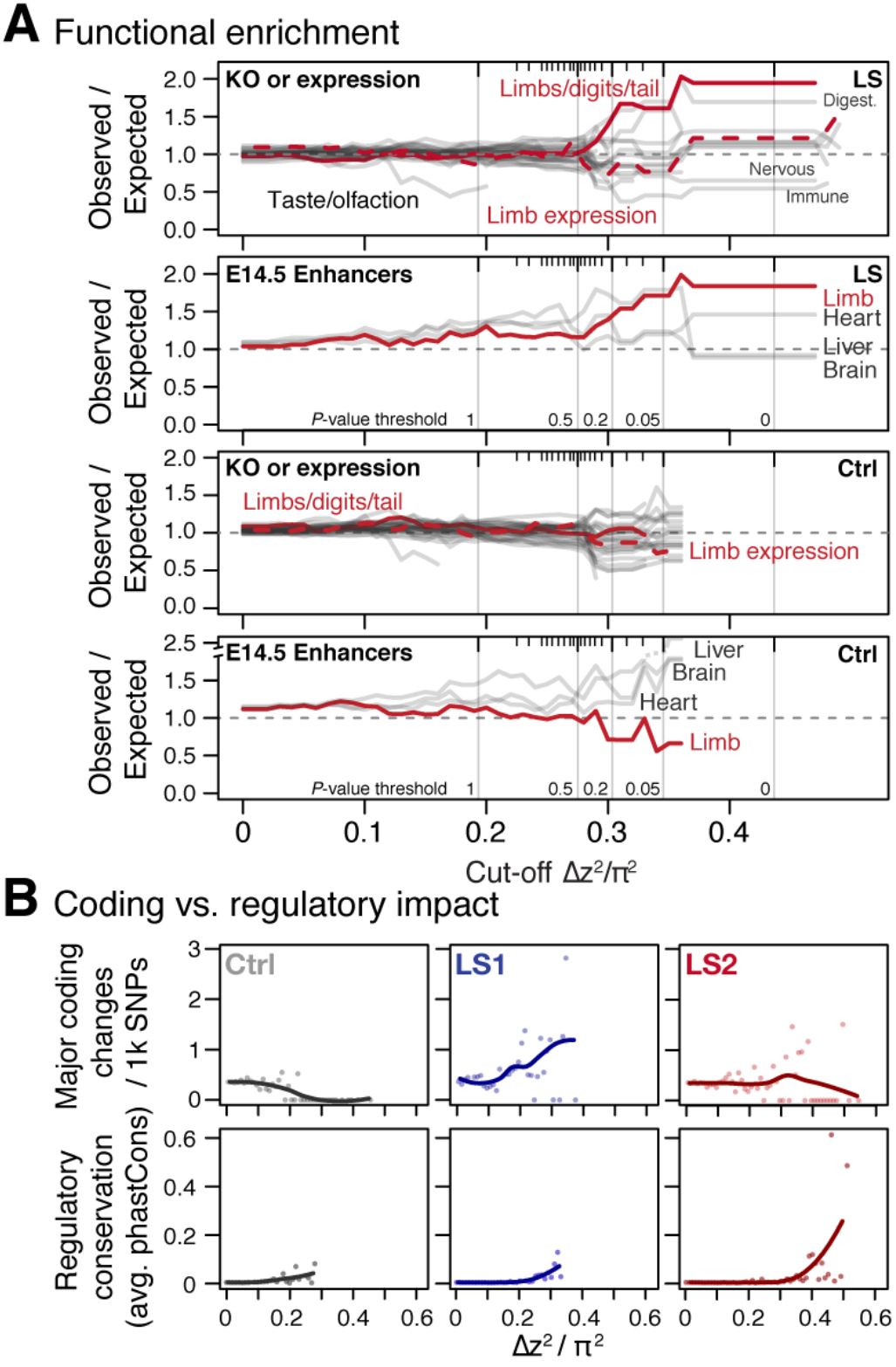

A Functional enrichment
KO or expression
Limbs/digits/tail
LS
Digest.
Nervous
Immune
Taste/olfaction
Limb expression
E14.5 Enhancers
LS
Limb
Heart
Liver
Brain
P-value threshold 1 0.5 0.2 0.05 0
KO or expression
Ctrl
Limbs/digits/tail
Limb expression
E14.5 Enhancers
Ctrl
Liver
Brain
Heart
Limb
P-value threshold 1 0.5 0.2 0.05 0
Observed / Expected
Cut-off Δz²/π²
B Coding vs. regulatory impact
Ctrl
LS1
LS2
Major coding changes / 1k SNPs
Regulatory conservation (avg. phastCons)
Δz² / π²

Fig. 3 – figure supplement 1. 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 topologically associating domains (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.

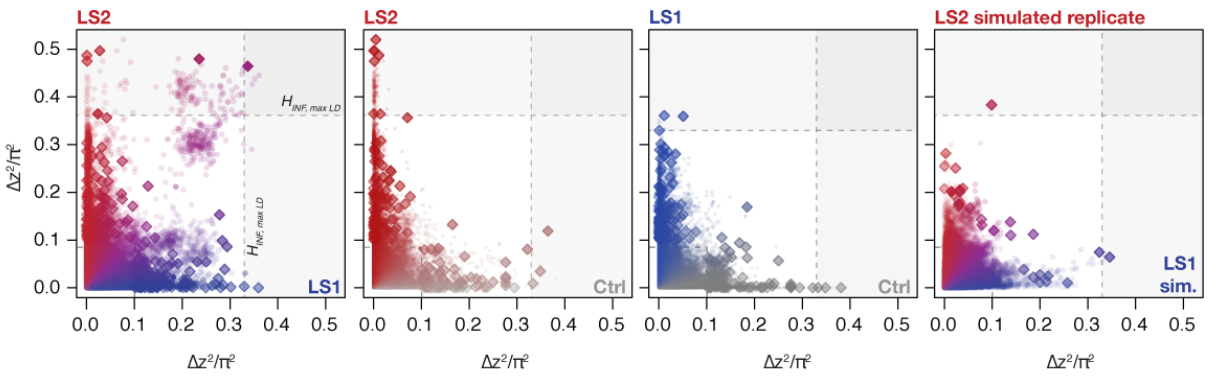
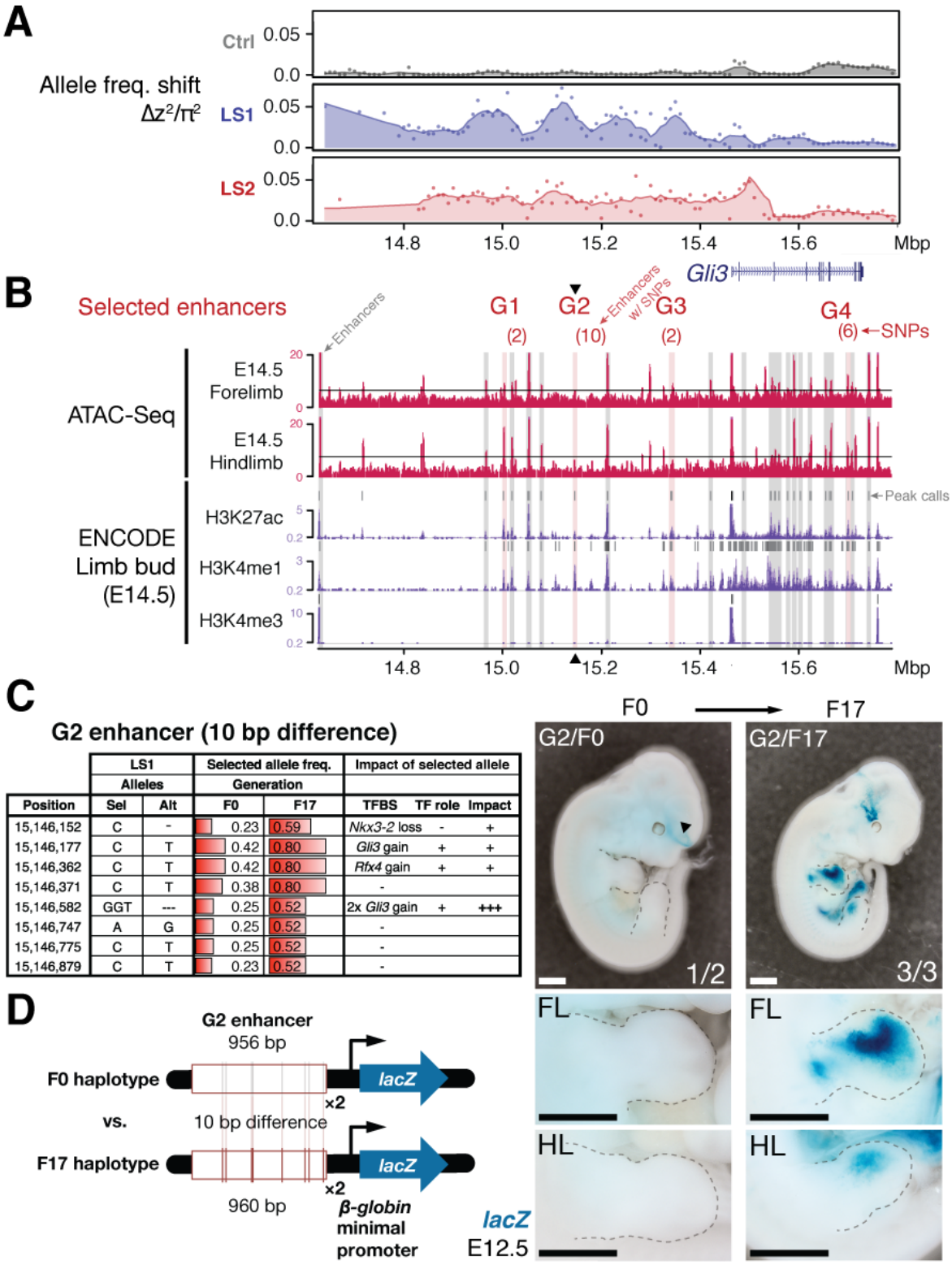


Fig. 3 – figure supplement 2. Changes in Δz^2 across lines. Shown are changes in Δz^2 in individual 10 kbp windows (all windows: circles; peak windows: diamonds). Generally there were no clear differences in Δz^2 along the axes except a slight skew toward higher values in LS2. When taken as a joint LS1–LS2 comparison, however, we observed that many windows show shifts in both LS1 and LS2 (left panel; in purple). In contrast, 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.



	LS1 Alleles		Selected allele freq. Generation		Impact of selected 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	-		

Fig. 4 – figure supplement 1. An enhancer in chromosome 13 boosts *Gli3* expression during limb bud development. (**A**) LS1 showed elevated Δz^2 in the intergenic region containing *Gli3*. (**B**) Putative limb enhancers (grey bars) were identified through peaks from ATAC-Seq (top) and histone modifications (bottom tracks, data from ENCODE project). Four of the enhancers contain mutations (in brackets) with significant allele frequency shifts between F0 and F17 in LS1 (red shading). One of the enhancers located close to the peak Δz^2 signal (G2, arrowhead) containing 10 bp differences was chosen for transgenic reporter assay. (**C**) Analysis of individual mutations showed an average increase of 0.33 in allele frequency, with 6 mutated positions affecting predicted binding of transcription factors 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 variants together drove robust and consistent *lacZ* reporter gene expression at E12.5, recapitulating *Gli3* expression in the developing fore- and hindlimb buds (right; see also Fig. 4 – figure supplement 2). 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 bars: 1 mm for both magnifications.

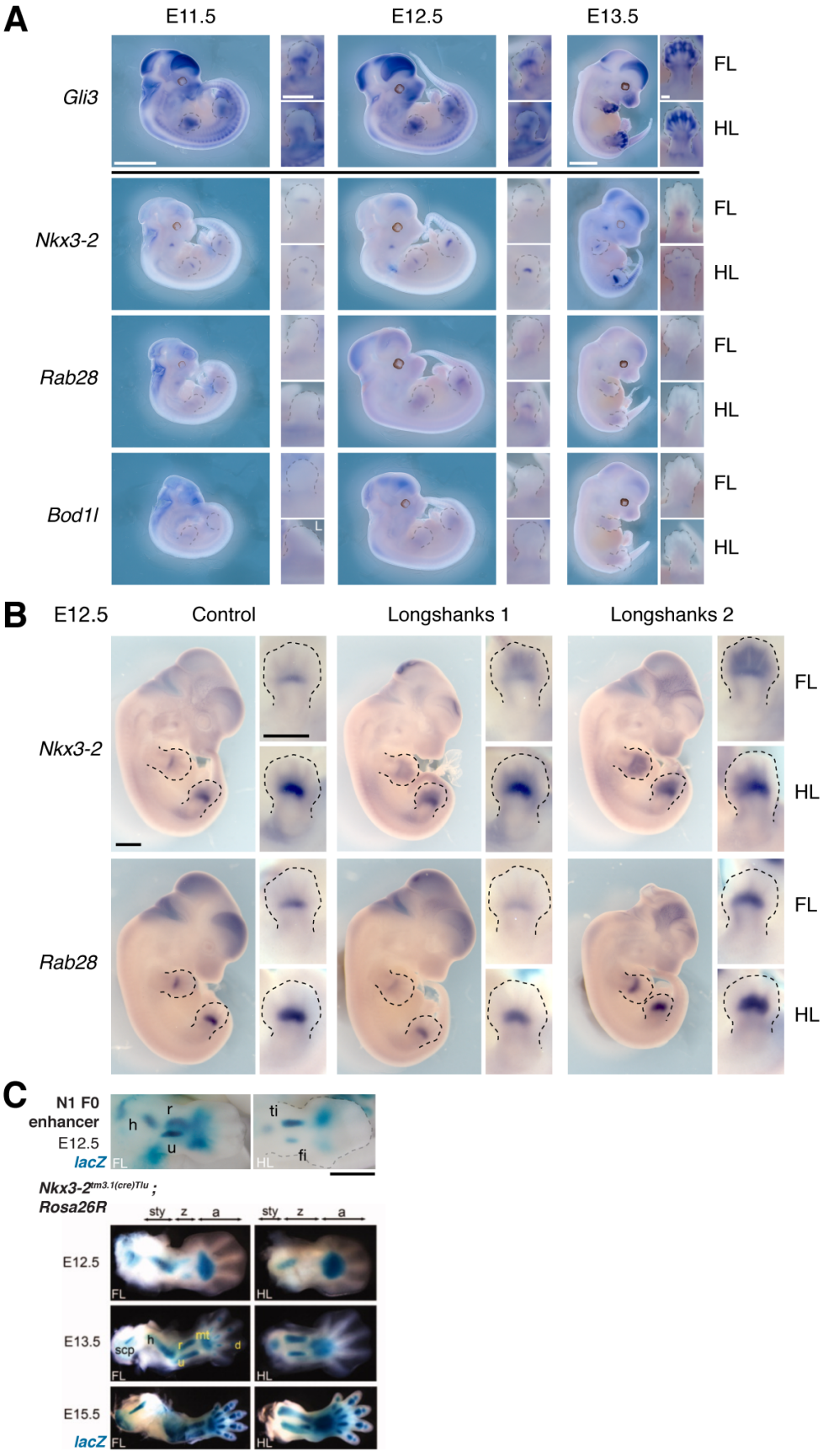

A
E11.5
E12.5
E13.5
Gli3
Nkx3-2
Rab28
Bod1l
FL
HL
B
E12.5
Control
Longshanks 1
Longshanks 2
Nkx3-2
Rab28
FL
HL
C
N1 F0
enhancer
E12.5
lacZ
Nkx3-2tm3.1(cre)Tlu ;
Rosa26R
sty
z
a
E12.5
E13.5
E15.5
lacZ
FL
HL
h
r
u
ti
fi
scp
mt
d

Fig. 4 – figure supplement 2. 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 (*30*) 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. FL, forelimb; HL, hind limb; unless otherwise indicated by "L", all images were taken from the right 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 (*30*). 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 (*30*), 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.

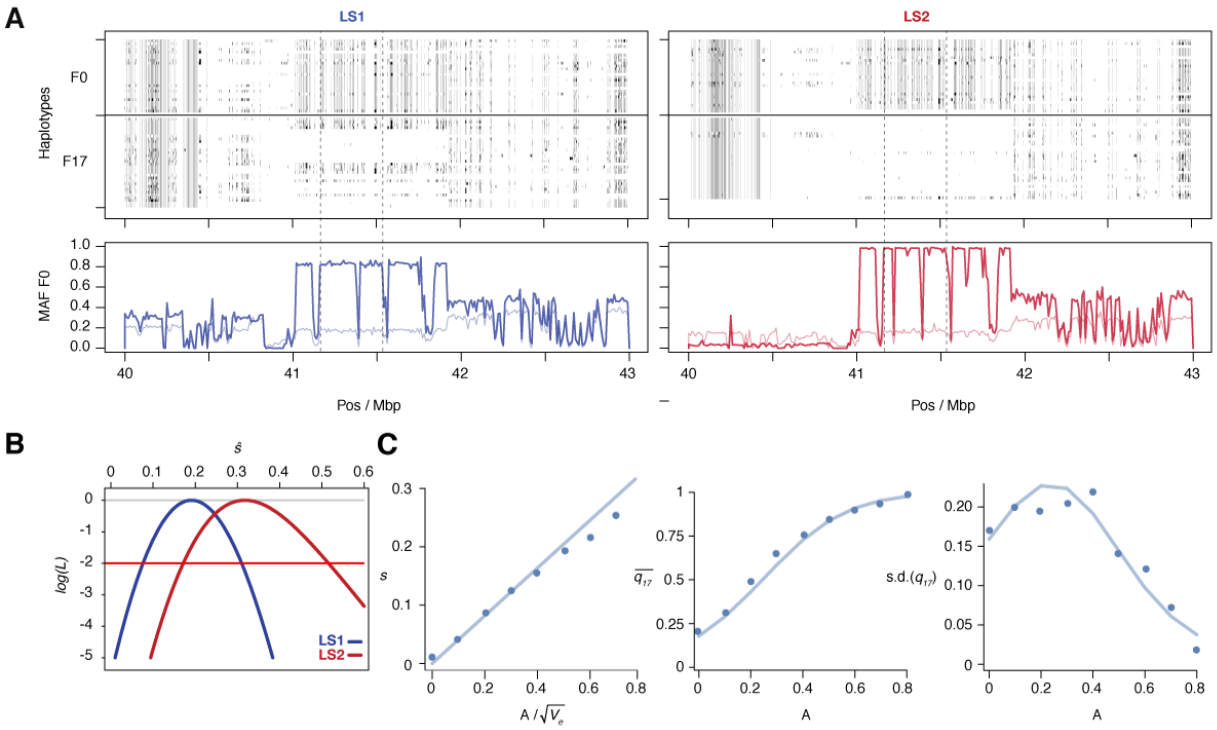


Fig. 5 – figure supplement 1. Selection at the *Nkx3-2* locus. (**A**) Raw genotypes from the F0 and F17 generations from LS1 (left) and LS2 (right) are shown, clearly indicating the area under the selective sweep. The genotype classes are shown as C57BL/6J homozygous (BL6, white), heterozygous (black) 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 the same originally rare allele was segregating at in both founder populations and 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**) The log likelihood of the selection coefficient, *s*, for LS1 and LS2 (blue and red, respectively), 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. (**C**) 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.

2.2.12 Supplementary Tables:

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 sum	**169**	**11,577,782,270**	**1609.9**	**589.5**	**2.91**

Supplementary Table 1. 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 inbred females and 2 inbred 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	6,642,764	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	6,772,383	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	7,121,979	7,062,183	7,214,508	6,795,919	5,813,162	0.17631
LS2 F17	7,218,921	7,155,580	7,239,774	6,832,125	6,778,518	5,749,742

Supplementary Table 2. 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 at least 100-fold among lines by generation F17 (above diagonal, orange boxes). Within-line F_{ST} is intermediate in this respect, reaching about half of the differentiation observed between lines.

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, Uqcr11, 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*

Rk	Chr	Span (Mbp)	Peak	Genes	Core span	TAD span	Genes	s	Type	Candidate genes*	All genes
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	120.3	108	**120.30–120.31**	**119.45–120.49**	*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**	**111.54–113.62**	*2*	0.12	LS2-specific	*Sox9**	*Btbd17, Cdc42ep4, Cog1, Cpsf4l, D11Wsu47e, D11Wsu99e, Dnaic2, Gpr142, Gprc5c, Kif19a, **Rpl38*†**, Sdk2, Slc39a11, **Sox9*†**, Sstr2, Ttyh2*

Rk: ranking by *s*
Chr: Chromosome
Mbp: megabase pairs
s: selection coefficient

*	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

Supplementary Table 3. 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.

				SNP Freq. change		Amino acid change					
Gene symbol	**Chrom**	**dbSNP ID**	**SNP position**	**LS1**	**LS2**	**LS1**	**Dir.**	**PhastCons**	**KO phenotype**	**KO lethal?**	**KO overall assessment**
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	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	0.2	0.52	T357A	PUR	0.001	short limbs		Mice homozygous for an in-frame deletion display **neonatal lethality**, respiratory failure, and severe chondrodysplasia.
		rs32148105	63,225,427	0.3	0.64	N451D	DIV	0			
		rs32529591	63,225,479	0.3	0.6	P468L	DIV	0.001		**Y**	
		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			

				SNP Freq. change		Amino acid change					
Gene symbol	**Chrom**	**dbSNP ID**	**SNP position**	**LS1**	**LS2**	**LS1**	**Dir.**	**PhastCons**	**KO phenotype**	**KO lethal?**	**KO overall assessment**
Fuz	7	rs46684086	44,898,988	0.38	-0.42	L264M	DIV	0.637	abnormal apical ectodermal ridge morphology, polydactyly	**Y**	Mice homozygous for a gene trapped allele exhibit neural tube closure defects, abnormal craniofacial morphology, abnormal skeletal morphology, polydactyly, anopthalmia, pulmonary hypoplasia, and cardiac outflow tract defects.
		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			
Acan	7	rs262365125	79,096,707	0.4	-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.
Kif7	7	rs211772880	79,699,167	0.46	0.04	N1220T	DIV	0.002	abnormal limb bud morphology, polydactyly, postaxial polydactyly, preaxial polydactyly, syndactyly	**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**, polydactyly, and abnormal neural tube development and neurogenesis.
		rs31859391	79,699,507	0.42	0.06	W1137R	PUR	0.992			
		rs222106773	79,708,493	0.58	0.01	S570G	PUR	0.152			

				SNP Freq. change		Amino acid change					
Gene symbol	**Chrom**	**dbSNP ID**	**SNP position**	**LS1**	**LS2**	**LS1**	**Dir.**	**PhastCons**	**KO phenotype**	**KO lethal?**	**KO overall assessment**
Dot1l	10	rs51353227	80,783,290	0.78	0.89	V371A	PUR	0.075	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.
		rs234091350	80,786,131	0.78	0.89	L756P	PUR	0.198			
		rs46367861	80,789,444	0.63	0.69	P1175L	DIV	0.367			
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, fused phalanges, syndactyly	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 persistence of hyaloid vascular system.
Fbn2	18	rs13464230	58,023,239	0.1	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.

Chrom: Chromosome
KO: knock-out
Dir.: directionality
PUR: purifying
DIV: diversifying

Supplementary Table 4. 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	5	41,698,405	41,625,451	785	For	AGGTGGCAAGATGTTGGATAAATAC
					Rev	GATCATCAAAGCTTGGAGCAGC
Nkx3-2	5	41,763,877	41,762,039	579	For	GCGATCCTCAACAAGAAAGAGGA
					Rev	GCGCTTCTTTCGCGGTTTAG
Bod1l	5	41,832,764	41,828,797	873	For	GATGCCATGTCAATCTTGGAAACC
					Rev	CACTGTGAGTTCGTCATCAGAATC

Chr: Chromosome

For: forward

Rev: reverse

Supplementary Table 5. Oligonucleotides for *in situ* hybridization probes.

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

VP: Viewpoint

Chr: Chromosome

Supplementary Table 6. 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 (*98*). 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	5	41,165,777	41,166,605	829	For	TCAGCTGATATCTTCAGCTGTCGACGCTGAACATTAAAGCCATTACAGGAGCTTTCC
					Rev	TCAGCTGATATCTTCAGCTCTCGAGCTGCCAACAGGCTCCTTCCTATAAACC
N3	5	41,536,216	41,537,263	1048	For	TCAGCTGATATCTTCAGCTGTCGACCACGAGACTTAGCCAAACCTGAGGTG
					Rev	TCAGCTGATATCTTCAGCTCTCGAGCTTCCTCCCAGTGCATATCTCCCAAC

Enh: Enhancer

Chr: chromosome

For: forward

Rev: reverse

Supplementary Table 7. Oligonucleotide primers for amplifying the enhancers at the *Nkx3-2* locus. 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 Major	5	41,536,402	41,536,520	119	For	GCAAATCTTGAAAAGTCAGGGAATAAAATA
					Rev	GCTAAATTTTGCCTCCTGCTTTT
N3-F17 Selected	5	41,536,402	41,536,520	119	For	GCAAATCTTGAAAAGTCAGGGAATAAAATG
					Rev	GCTAAATTTTGCCTCCTGCTTTG

Chr: Chromosome

For: forward

Rev: reverse

Supplementary Table 8. 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.

Chapter 3: Genetic mapping of species differences via *in vitro* crosses in mouse embryonic stem cells

3.1 Declaration of Contributions:

Authors: Stefano Lazzarano, Marek Kučka, **João P. L. Castro**, Ronald Naumann, Paloma Medina, Michael N. C. Fletcher, Rebecka Wombacher, Joost Gribnau, Tino Hochepied, Marc Van Montagu, Claude Libert, and Yingguang Frank Chan

João P. L. Castro contributions: Designed, tested and analyzed data from loss-of-heterozygosity markers; performed stem cell culture; prepared and analyzed mitotic chromosome spreads; aided manuscript review and editing.

Significance of author contributions to the collective work: Data obtain from loss-of-heterozygosity screens proved the occurrence of mitotic recombination in BL6/CAST stem cells. Analysis of mitotic chromosome spreads allowed the assessment of the impact of ML216 on karyotype stability.

Co-author contributions:
Stefano Lazzarano: designed, performed and analyzed all stem cell experiments; manuscript writing, editing and review. **Marek Kučka:** prepared genomic libraries for sequencing**;** performed cloning of targeting constructs; manuscript writing, editing and review. **Ronald Naumann:** performed laser assisted morula injection. **Paloma Medina, Michael N. C. Fletcher and Rebecka Wombacher:** performed initial stem cell experiments. **Joost Gribnau**, **Tino Hochepied, Marc Van Montagu, Claude Libert:** provided F1 hybrid stem cell lines, manuscript writing, editing and review.
Yingguang Frank Chan: designed, performed and analyzed all stem cell experiments and genomic data; performed microCT scans and respective data analysis; manuscript writing, editing and review.

3.2 Full article

Genetic mapping of species differences via "*in vitro* crosses" in mouse embryonic stem cells

Authors: Stefano Lazzarano[a], Marek Kučka[a], João P. L. Castro[a], Ronald Naumann[b], Paloma Medina[a,1], Michael N. C. Fletcher[a,2], Rebecka Wombacher[a,3], Joost Gribnau[c], Tino Hochepied[d,e], Marc Van Montagu[f,g,4], Claude Libert[d,e], and Yingguang Frank Chan[a,4]

Affiliations

[a]Friedrich Miescher Laboratory of the Max Planck Society, 72076 Tübingen, Germany; [b] Max Planck Institute of Cell Biology and Genetics, 01307 Dresden, Germany; [c]Department of Reproduction and Development, Erasmus University Medical Center, 3015 CE Rotterdam, The Netherlands; [d]Department of Biomedical Molecular Biology, Ghent University, 9052 Ghent, Belgium; [e]Center for Inflammation Research, VIB, 9052 Ghent, Belgium: [f]Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052 Ghent, Belgium; [g] Center for Plant Systems Biology, VIB, 9052 Ghent, Belgium

Corresponding authors:

Marc Van Montagu (marc.vanmontagu@ugent.vib.be)

Yingguang Frank Chan (frank.chan@tue.mpg.de)

PNAS (Proceedings of the Natural Academy of Science)

Received: 11 October 2017

Accepted: 26 February 2018

Published: 03 April 2018

Reviewing editors: Frank W. Albert, Martien Kas, and Joseph Schacherer

3.2.1 Abstract:

Discovering the genetic changes underlying species differences is a central goal in evolutionary genetics (1). However, hybrid crosses between species in mammals often suffer from hybrid sterility, greatly complicating genetic mapping of trait variation across species. Here we describe a simple, robust and transgene-free technique to generate "*in vitro* crosses" in hybrid mouse embryonic stem (ES) cells by inducing random mitotic crossovers with the drug ML216, which inhibits *Bloom syndrome* (*BLM*) (2). Starting with an interspecific F1 hybrid embryonic stem cell line between the *Mus musculus* laboratory mouse and *Mus spretus* (~1.5 million years of divergence) (3, 4), we demonstrate the feasibility of mapping the genetic basis of drug resistance to the anti-metabolite tioguanine to hypoxanthine-guanine phosphoribosyltransferase (*Hprt*) in as few as 6 days through "flow mapping", by coupling *in vitro* crosses with fluorescence-activated cell sorting (FACS). We also show how our platform can enable direct study of developmental variation by re-deriving embryos carrying recombinant hybrid genomes. We demonstrate how *in vitro* crosses can overcome major bottlenecks in mouse complex trait genetics and address fundamental questions in evolutionary biology that are otherwise intractable through traditional breeding due to high cost, small litter sizes and/or hybrid sterility. In doing so we describe an experimental platform towards studying evolutionary systems biology in mouse and potentially in human and other mammals, including cross-species hybrids.

3.2.2 Keywords: Interspecific hybrids; mitotic recombination; genetics

3.2.3 Significance Statement:

How species differ from each other is a key question in biology. But genetic mapping between species often fail because of sterile hybrid crosses. Here we have developed a novel technique called *in vitro* recombination to circumvent breeding. We induce genetic reshuffling through mitotic recombination with the drug ML216 and mapped trait variations in a dish. Starting with hybrid embryonic stem cells between the *Mus musculus* laboratory mouse and *Mus spretus* spanning ~1.5 million years of divergence, we show it is possible to map the gene responsible for differential resistance to the drug tioguanine in as few as 6 days. Our technique opens up novel experimental avenues in genetic mapping across mouse species and potential applications in personalized medicine.

3.2.4 Introduction:

Discovering the genetic changes underlying species differences is a central goal in evolutionary genetics (1). However, hybrid crosses between even recently diverged species in animals often suffer from hybrid sterility (1, 4), greatly complicating genetic mapping of trait variation, especially in mammals. On the other hand, within-species genetic mapping has been tremendously successful in linking genetic polymorphisms to trait variations in innumerable organisms since the early twentieth century (5–7). Genetic mapping typically involves breeding and analyzing mapping panels to identify genetic loci controlling trait variations, or quantitative trait loci (QTL). The ability to disentangle linked genetic associations determines mapping resolution and is largely dependent on the number of meiotic crossovers. Accordingly, researchers are driven to create ever-larger mapping populations and/or accumulating recombination over at least two, often many generations to achieve gene-level mapping resolutions (8–10). In the mouse, genetic studies are complicated by the relatively long generation times and small litter sizes, which often decline with increased inbreeding. Consequently, compared to yeast, worms and *Arabidopsis* (8–10), genetic mapping in the mouse requires far greater resources, yet relatively few traits have been mapped to the gene level (but see landmark

studies identifying *Tlr4* and *Prdm9*) (11, 12). This challenge was particularly acute for crosses at or beyond the species level, where hybrid sterility often makes it impossible to generate a panel in the first place. Nonetheless, the potential to reveal unique biology occurring at the species boundaries in mammalian evolution makes such panels worthy attempts, even allowing for lower mapping resolution (13–17). This is because evolutionary changes in trait architecture (even at the cellular or tissue level) can reveal much about the underlying evolutionary process. Should genetic exchange in hybrid animal genomes become feasible, direct genetic mapping of species differences would become routinely possible.

We set out to establish a universal method that allows genetic mapping in mammals without breeding, even across divergent species with viable but otherwise sterile hybrids. We opted to use mouse ES cells to take advantage of the full spectrum of genetic manipulations available in tissue culture. A minimal system will have the two following features: an ability to induce on-demand extensive genetic exchange; and genetic (and trait) variation such as those found in F1 hybrid ES cells, ideally between species.

Intriguingly, the technique to create genetic variation through recombination has been in broad use in the mouse genetics community, albeit never explicitly in F1 hybrid ES cells with the goal of genetic mapping. In 2004, two independent groups showed that recessive, biallelic mutants could be reliably recovered in mouse ES cells without breeding by suppressing *Bloom Syndrome* (*Blm*; Fig. 1*A*) (18, 19). Yusa and coworkers showed that these recessive mutants arose via mitotic recombination between homologous chromosomes (20). We reasoned that the same mechanism could be leveraged to generate genome-wide random *mitotic recombination*. This mechanism enabled the creation of panels of arbitrary size carrying recombinant genomes, while avoiding the limitations of hybrid sterility or inbreeding depression (Fig. 1*B*).

3.2.5 Results:

To test if BLM inhibition could lead to elevated homologous recombination rates in mitosis, we inhibited BLM in a number of mouse ES cell lines using a recently discovered small molecule inhibitor ML216 (Fig. 1*C*) (2). As a first test, we started with F1 ES cells between the laboratory mouse strains C57BL/6J ("BL6" in short) and 129 that carried a targeted transgene as a hemizygous allele at the *GtRosa26* locus on distal Chromosome 6. We estimated homologous recombination by counting colony survival under fialuridine (FIAU) treatment, which selected against the transgene consisting of hygromycin phosphotransferase–thymidine kinase (HyTK) and a green fluorescent protein (GFP; Fig. S1). We found that BLM inhibition led to highly elevated rates of homologous recombination, as revealed by increased numbers of FIAU-resistant colonies (Fig. 1*C*; *in vitro* recombination rate: 2.9×10^{-4} per cell per generation) and the appearance of mosaic GFP expression within a colony (Fig. 2*A*, right panels). This is consistent with previously reported rates under *Blm* suppression or disruption (targeted tetracycline inhibition or knockout alleles: $2.3–4.2\times10^{-4}$; vs. wildtype rates between 8.5×10^{-6}–2.3×10^{-5}) (18, 19). The small molecule BLM inhibitor ML216 offers unique experimental advantages, because its application is simple, rapid and reversible, eliminating the use of transgenes against *Blm* (18, 19) or repeated transfections of small interfering RNA to achieve continued suppression of *Blm*. Importantly, elevated homologous recombination under BLM inhibition is not associated with increased aneuploidy (N=154 metaphase spreads; Mann-Whitney U test, W=1871, h_1>0, n.s.; Fig. S2*A*). Further, ML216-treated ES cells retained robust expression of NANOG, a key stemness marker (Fig. S3).

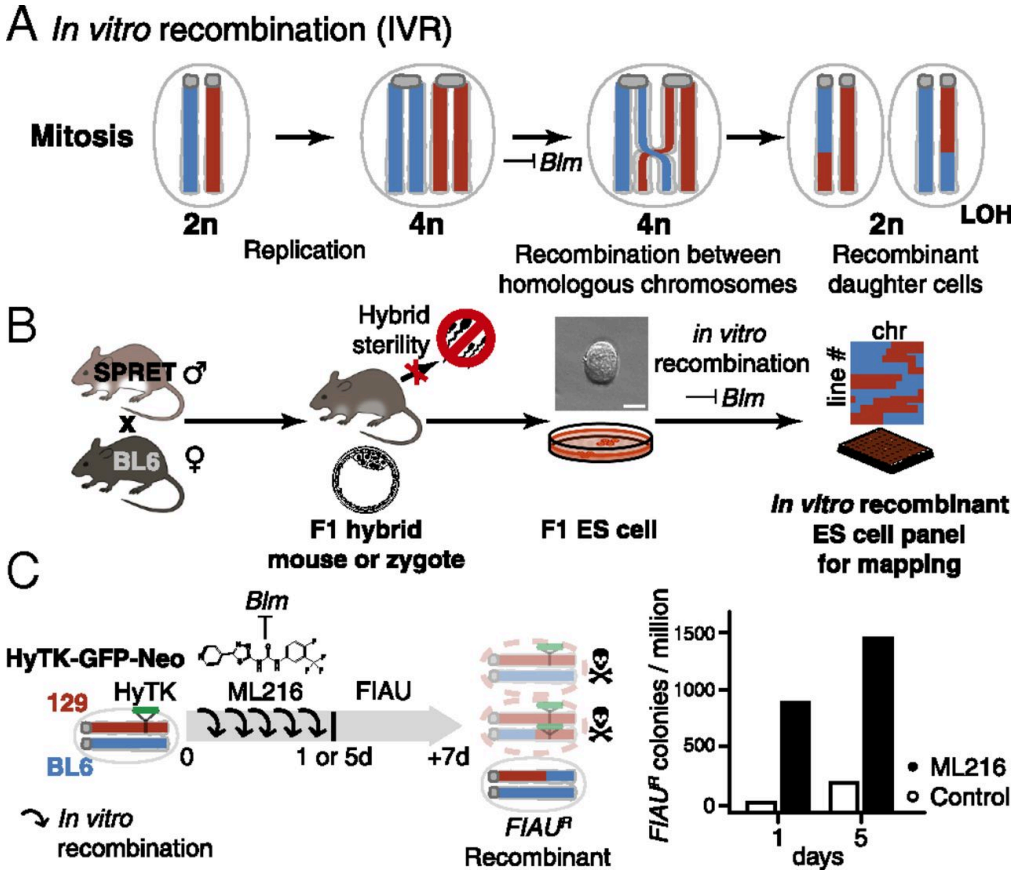


Fig. 1. ***In vitro* recombination via *Bloom syndrome* suppression.** (*A*) *Bloom syndrome* (*Blm*) encodes a helicase normally active during mitosis. Loss of *Blm* activity leads to increased improper sister chromatid exchange as well as recombination between homologous chromosomes. Mitotic recombination can give rise to recombinant diploid daughter cells with loss of heterozygosity (LOH) between the breakpoint and the telomeres. (*B*) *In vitro* recombination (IVR) allowed the circumvention of hybrid sterility in crosses between the laboratory mouse, e.g., C57BL/6J (BL6) and a murine sister species *Mus spretus* (*SPRET*). (BL6 x *SPRET*)F1 hybrid mice were viable and allowed derivation of F1 ES cells despite male sterility (3). Applying IVR to F1 ES cells allowed rapid and efficient generation of recombinant ES cell panels for genetic mapping. Scale bar = 50 μm. (*C*) Efficiency of IVR was estimated by colony survival assay. We estimated recombination rate between homologous chromosomes with cells hemizygous for a dominant selectable

marker (hygromycin phosphotransferase-thymidine kinase, abbreviated HyTK, green). We induced IVR by adding a small molecule BLM inhibitor ML216 (2) to the culturing medium for 1 or 5 days (d). Under fialuridine (FIAU) negative selection, cells having undergone mitotic recombination to become homozygous for the wildtype BL6 alleles (blue) survived; while non-recombined cells or recombinant cells retaining the HyTK transgene metabolized FIAU, resulting in cell death due to misincorporation of toxic nucleotide analogues (top and middle cells with red chromosomes). Under ML216 treatment (25 μM), IVR rate was estimated to be 2.9×10^{-4} per cell per generation, yielding 800–1500 FIAU-resistant colonies per million following treatment.

To determine the frequency and distribution of mitotic crossovers under ML216-mediated BLM inhibition, we sequenced and compared the genomes of 11 clones that survived ganciclovir selection (a FIAU alternative; Fig. 2*B*). We also treated F1 hybrid ES cells derived from BL6 and *Mus castaneus* (diverged ~1 million years ago; CAST/EiJ, abbreviated to *CAST*) (21) with ML216 but otherwise grown *without selection*. Using the transgene-free (BL6 x *CAST*)F1 line (21), we screened 46 randomly-picked ML216-treated clones for LOH recombinants and recovered recombinants in both BL6/BL6 and *CAST*/*CAST* directions on Chromosome 1. Sequencing of representative clones revealed conversion from F1 heterozygous genotypes towards both homozygous genotypes at the telomeres (Fig. 2*C*, clones 21 and 50, note also additional recombination on Chromosome 13). In contrast, control non-recombinant clones retained heterozygosity at the telomeres (clone 54 and 56). But even here we discovered a single clone carrying additional internal recombinants on Chromosome 1 (Fig. 2*C*).

Genome-wide sequencing of the recombinants revealed several striking patterns. First, crossover breakpoints were distributed along the entire chromosome (Fig. 2*B*), suggesting recombinants can be used for interval mapping. Second, ganciclovir selection significantly enriched for recombinant chromosomes compared to unselected conditions (N=11 out of 11 vs. 9 out of 826; Fisher Exact Test, $P < 2.2\times10^{-16}$), with the recombination map biased strongly by the location of the selection cassette (all 11 crossovers were centromeric to Chromosome 6, 113 Mbp, Fig. 2*B*). This is presumably because chromosome segments telomeric to the

cassette had no effect on selection. In our subsequent experiments on natural variations, this centromeric bias became a useful signature for trait mapping. Third, crossovers created by mitotic recombination usually occurred only on one or few chromosomes at a time (Fig. 2 *B* and *C*; Fig. S4)(20,22), unlike in meiosis with typically one crossover per chromosome arm. Taken together, the data show that BLM inhibition efficiently generated *in vitro* recombination (IVR) across wide evolutionary distance and IVR ES cell panels may constitute genetically distinct lineages suitable for genetic mapping.

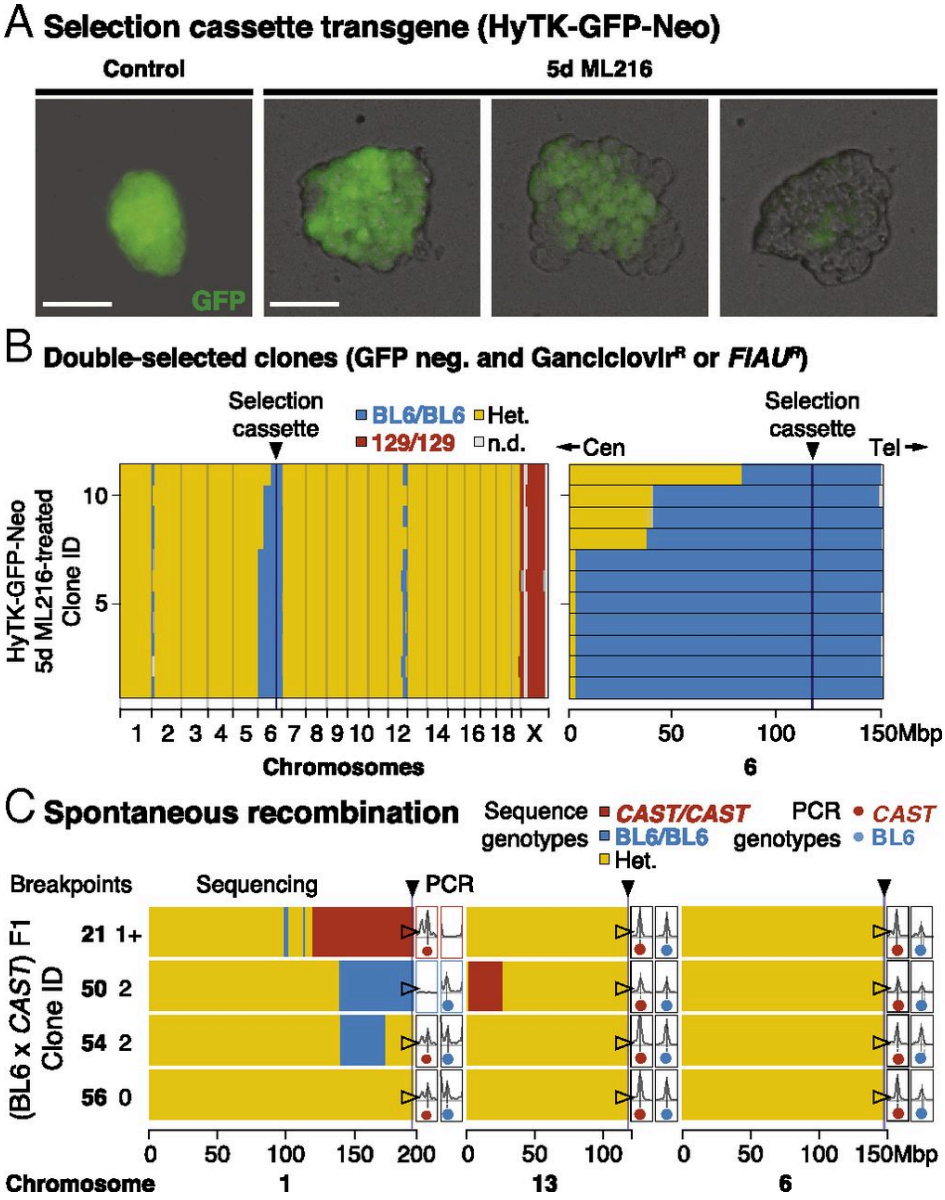


Fig. 2. Widespread *in vitro* recombination across a range of evolutionary divergence. (*A*) ES cell colonies displayed mosaic GFP expression within a colony when cultured with ML216, but not under control conditions, consistent with homologous recombination and loss of GFP through IVR. Recombination between homologous chromosomes could result in daughter cells with two wildtype (BL6 allele, dark) or transgenic copies (129 allele, bright). Early recombination events followed by random cell loss during clonal expansion could produce completely dark colonies. Scale bar = 100 μm. (*B*) After expansion under negative selection against the transgene (both ganciclovir and FIAU kill cells expressing HyTK), 11 ganciclovir-resistant and GFP-negative colonies were whole genome sequenced. Selection favoured loss of transgene (homozygous BL6/BL6 genotypes) at distal Chromosome 6. In contrast to normal meiotic recombination (averaging 1 or more crossovers per chromosome pair), mitotic recombination typically affected only a single chromosome pair: much of the genome remained heterozygous (yellow), with the exception of the transgene-carrying chromosome 6 (mostly BL6/BL6, blue) and the single 129 Chromosome X (male, 129, red). Mitotic recombination events converted genotypes telomeric to the breakpoint towards homozygosity (LOH, yellow to blue). (*C*) IVR also occurred in cells carrying divergent genomes with no transgenes. (BL6 x *CAST*)F1 hybrid ES cells were treated with ML216 and screened by PCR genotyping at diagnostic telomeric markers. Selected clones (two recombinant and control clones each) were whole genome sequenced, showing recombination events towards both homozygous genotypes, consistent with PCR genotype screening results (total breakpoints per clone ranged from 0–2). Additional recombination events were also recovered, even though the Chromosome 1 telomeric marker remained heterozygous (clone 54). These clones also carried non-recombined chromosomes (e.g., Chromosome 6, fully heterozygous, yellow).

Our experiments to determine IVR rate demonstrated that the collective location of recombination breakpoints could reveal the position of the selectable transgene (HyTK or GFP), with the major difference that under mitotic recombination the critical interval was defined only on the centromeric side. To further illustrate the potential of this approach, we used IVR to map naturally-occurring variations. One

classical polymorphism is the 25 to 75-fold increased activity of the *Mus spretus* “a” allele of hypoxanthine-guanine phosphoribosyltransferase (*Hprt*^{a}) compared to the laboratory mouse *Hprt*^{b} allele (23). Importantly, HPRT metabolizes the anti-metabolite tioguanine (6-TG) and causes cytotoxicity. It should be noted that beside the known *Hprt* polymorphism, tioguanine susceptibility itself has not been previously mapped genetically within or between mouse species. Here, we expected ES cells carrying *Hprt*^{a} to be highly susceptible to 6-TG treatment, whereas *Hprt*$^{b/b}$ or *Hprt*$^{-/-}$ ES cells should survive far higher 6-TG concentrations (Fig. S5). We set out to map the QTL for differential 6-TG susceptibility using a bulk segregant assay simply by comparing allele frequencies across the genome between pools of 6-TG susceptible and resistant ES cells.

We first confirmed the absence of chromosome-scale rearrangements between the parental strains that could preclude mapping using the *de novo* assembled genomes of the parental strains made available by the Wellcome Trust Sanger Institute (BL6 and SPRET/EiJ, abbreviated to *SPRET* here) (24, 25). We generated IVR panels by treating a female (BL6 x *SPRET*)F1 hybrid ES cell line (“S18”) (3) with ML216 over 5, 10 and 21 days (d; Fig. 3*A*). The use of a female ES cell line, which carried two active X chromosomes prior to the onset of X inactivation during differentiation (26), allowed direct selection on the alternative *Hprt*^{a} and *Hprt*^{b} alleles. After confirming biallelic *Hprt* expression in S18 cells using quantitative PCR, we treated control and IVR S18 cells with 6-TG and determined cell viability via a 4’,6-diamidino-2-phenylindole (DAPI) exclusion assay. Damaged cells with ruptured membrane exhibited rapid uptake of DAPI, a feature unaffected by ML216 treatment, and were distinguishable by fluorescent-activated cell sorting (FACS; “Live” proportions under ML216 treatment vs. “Live” proportions under 6-TG treatment, N = 5 paired treatments; Kruskal-Wallis test, χ^2 = 13.17, d.f.= 1, P < 0.0003; Fig. 3*A*; Fig. S6). We separately recovered and sequenced each “Resistant” (*6-TG*^{R}) and “Susceptible” (*6-TG*^{S}) pool (Fig. 3*A*). Under both 5d and 21d ML216 treatment, a large skew towards enriched *SPRET* coverage was observed on Chromosome X in the *6-TG*^{S} relative to the *6-TG*^{R} pool (Fig. 3 *A* and *B*). This was in stark contrast to the genomic background, which showed little deviation from equal *SPRET* and BL6 contributions (normalized *SPRET* coverage for Chromosome X: 1.10, 95% confidence interval: 1.02–1.19; autosomes: 0.998, conf. int.: 0.986–1.01). The genome-wide peak *SPRET* enrichment window was found on Chromosome X, 53

Mbp, and it contained the *Hprt* gene itself (normalized *SPRET* coverage in *6-TG*^{S} pool, 1 Mbp window: 1.19, conf. int.: 1.09–1.28). Here, our forward genetic mapping for 6-TG susceptibility clearly identified a single locus, suggesting that 6-TG susceptibility depended only on *Hprt* genotypes. To confirm the role of *Hprt* in mediating differential 6-TG susceptibility beyond bulk sequencing, we also sequenced 46 individual *6-TG*^{R} IVR clones after 10d ML216 treatment to determine their recombination breakpoints (Fig. 3*C*). Echoing the skewed crossovers patterns centromeric to the HyTK selection cassette (Fig. 2*B*), we observed more *SPRET*-to-BL6 than BL6-to-*SPRET* centromeric recombinants (N = 35 vs. 8, $P \leq 2 \times 10^{-5}$, exact binomial test, $h_1 \geq h_0$) and also ruled out aneuploidy or deletion of *Hprt*^{a} as major contributors to 6-TG resistance. We note, however, that despite the strongly skewed ratio of 27 BL6/BL6 homozygous clones at the *Hprt* locus out of 46 total recovered clones, we still observed 9 heterozygotes and 10 *SPRET*/*SPRET* homozygous clones (BL6/BL6 58.6%; Chi-squared test using observed allele counts, $\chi^2 = 13.17$, d.f. = 2, $P \leq 0.002$). Close examination of our flow cytometry data suggests that the original DAPI-only FACS gating may not be sufficiently exclusive (as opposed to gating using additional channels, Fig. 3*A* and S6). Other alternative explanations may be a quantitative, rather than absolute allelic difference in susceptibility to 25 μM 6-TG treatment (Fig. S5); or other mutation(s) at *Hprt* or elsewhere leading to 6-TG resistance (18). Taken together, we conclude that we were able to perform forward genetic mapping using IVR and recover *Hprt* as the gene underlying 6-TG susceptibility differences between BL6 and *SPRET*.

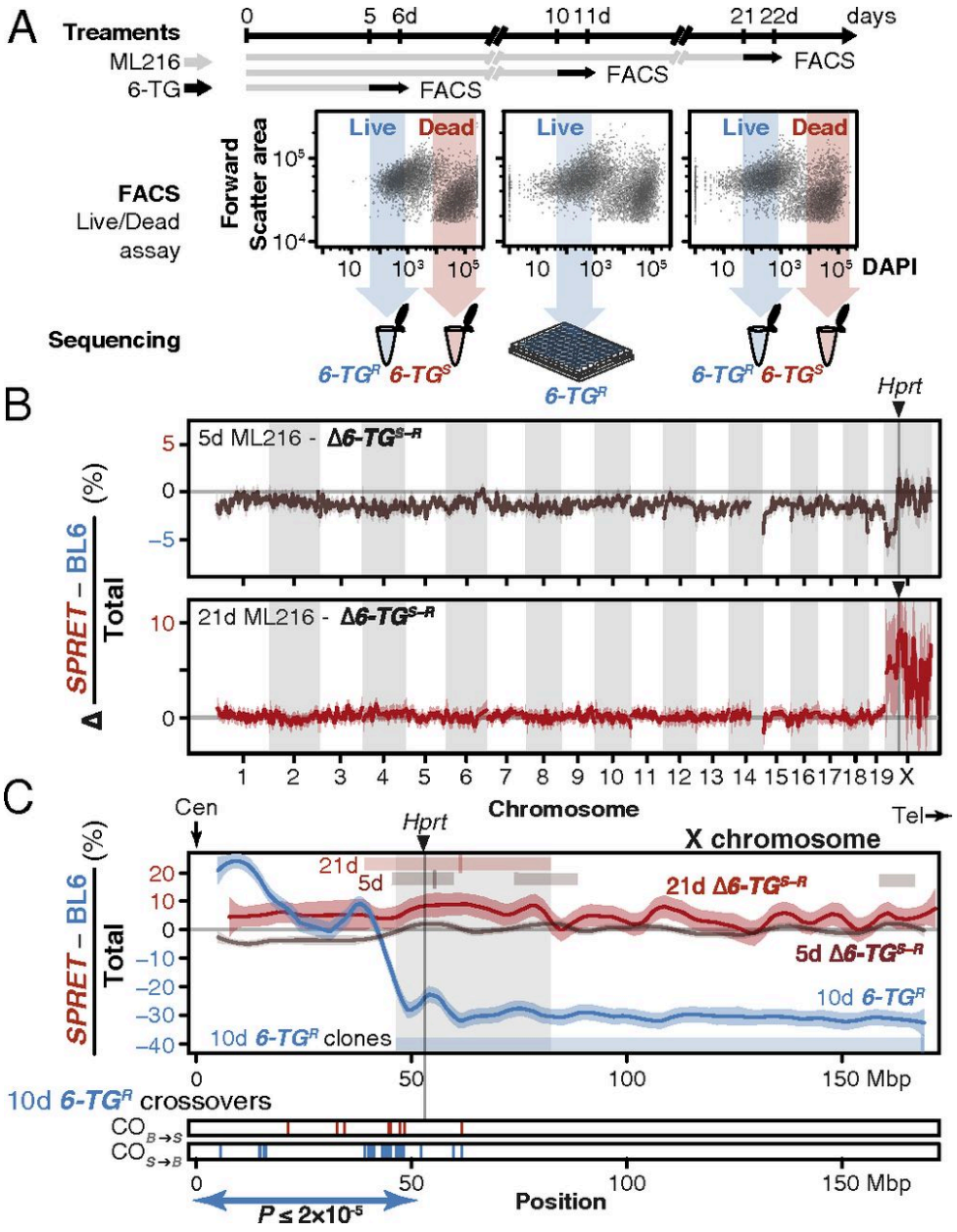


Fig. 3. ***In vitro*** **genetic mapping of variation in tioguanine susceptibility between divergent species.** (*A*) A female ES cell line S18 derived from a *Mus spretus* and C57BL/6N F1 interspecific hybrid was treated with ML216 (25 μM) and subjected to the anti-metabolite tioguanine (6-TG) for 1d prior to fluorescence-activated cell sorting (FACS). ES cells were evaluated for viability based on 4',6-diamidino-2-phenylindole (DAPI) exclusion. Resistant and susceptible (*6-TG*^{R} and *6-TG*^{S}) sub-populations were gated conservatively (shaded arrows) and pooled for sequencing. Individual clones from the 10d ML216 treatment were cultured and whole genome sequenced. (*B*) Skewed allelic contributions between the *6-TG*^{R} and

6-TG^{S} pools suggested that the *SPRET* allele on Chromosome X conferred 6-TG susceptibility. Allele frequencies were normalized against *6-TG*^{R} sample as an internal ML216 treatment control. Plotted are per megabase mean *SPRET* allele frequencies ± s.e.m. after 5 d and 21 d ML216 treatment. In both cases, the genome-wide peak window contains the *Hprt* gene with the *SPRET* allele showing significantly increased susceptibility. (*C*) Individual *6-TG*^{R} clones following 10 d ML216 treatment were sequenced to determine recombination breakpoints. Crossovers in clones surviving 6-TG treatment recombined significantly more likely in the *SPRET*-to-BL6 direction (*S*>*B* = 37; *B*>*S* = 5; $P \leq 2\times10^{-5}$) between the centromere and *Hprt*, consistent with strong selection favouring the BL6 *Hprt*^{b} allele. In contrast, only 3 additional crossovers were detected telomeric to *Hprt*. At *Hprt*, most 6-TG surviving clones are homozygous for the *Hprt*^{b} allele (27 vs. 9 heterozygotes and 10 *Hprt*^{a} homozygotes).

The ability to easily circumvent hybrid sterility in evolutionarily divergent murine species led us to ask what developmental phenotypes may arise from such otherwise inaccessible genetic configurations (*M. spretus*–laboratory mouse hybrid males are sterile, following Haldane's rule. Backcrosses using female hybrids are possible but extremely challenging) (17). Assaying developmental phenotypes from evolutionarily divergent hybrid ES cells is non-trivial, because hybrid sterility makes conventional re-derivation impossible due to the lack of germ line transmission in the chimera generation. Instead, we directly produced fully ES cell-derived founder animals using laser-assisted morula injection (27) with two karyotypically normal but genetically distinct IVR ES cell lines along with control, non-recombined S18 cells (IVR 1 and 2; Fig. S2; Fig. S7; Movies S1–3). We succeeded in obtaining multiple embryos per line at embryonic (E) 14.5d of development (N=36, 24 from IVR lines vs. N=9 untreated S18 line). Using high-resolution micro-computer tomography (microCT), we observed that the embryos from the untreated clones showed uniformly normal development, whereas embryos from both IVR lines showed both normal development and dramatic craniofacial and neural tube closure defects (2 abnormal embryos out of 4 scanned embryos in IVR line 1; 2 out of 7 in line 2; and 0 out of 6 from the original S18 line; Fig. 4; Fig. S8; Movies S1–3). Neural tube and craniofacial defects are among the most common developmental defects due to the

complex coordination of cell migration and cell–cell communications, which may be impaired due to novel genetic combinations in the IVR lines (Fig. S7). Besides major developmental defects, we also made 3D measurements from various organs, including sub-regions of the brain, the heart and the liver, in individual embryos from each ES cell line. This approach illustrates for the first time the potential and feasibility of characterizing, or given an expanded IVR panel, mapping the genetic basis of evolutionary developmental variation using recombinants from divergent species.

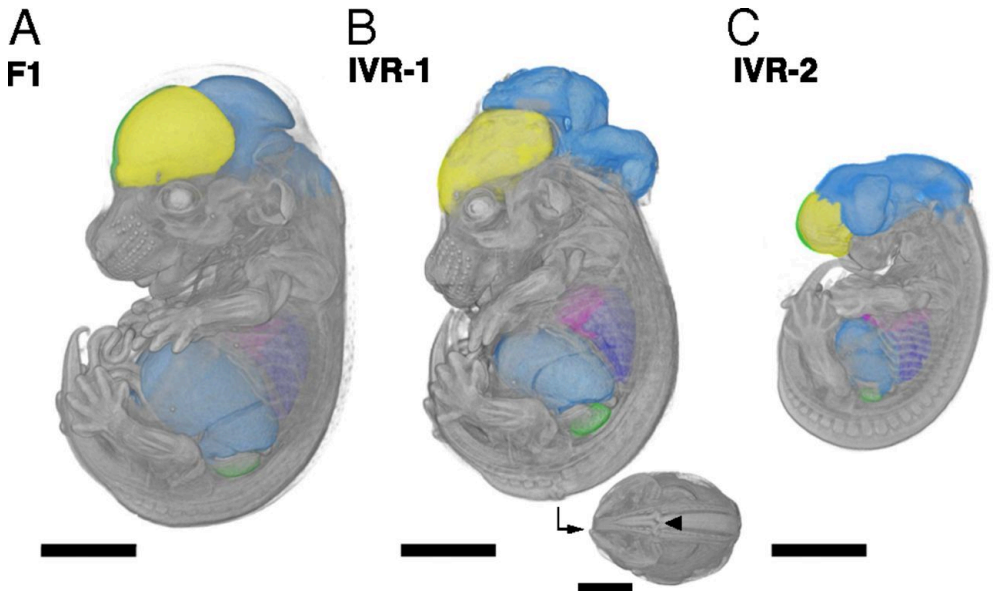


Fig. 4. Accessing developmental phenotypes in recombinants between evolutionarily divergent species. Embryos at mid-gestation (14.5 d after fertilization) with nearly exclusive ES cell contribution were derived from non-recombinant F1 S18 ES cells (*A*) and IVR lines 1 (*B*) and 2 (*C*)**.** Embryos were dissected, contrast-stained and scanned using X-ray micro-computer tomography at 9.4 µm resolution. The high scanning resolution allowed identification and precise measurements of individual organs (colorized here). Major developmental craniofacial and neural tube closure defects were observed in the IVR lines (*B*, caudal view with arrowhead indicates neural tube lesion). Scale bar = 200 µm.

3.2.6 Discussion:

A central goal of evolutionary genetics is to identify how mutations arose during evolution and influenced phenotypes. For many organisms, a major barrier has been the inability to reliably generate diverse and large mapping panels of sufficient evolutionary diversity. Here we describe a simple and robust method to make "*in vitro* crosses", resulting in functionally intercross panels from otherwise sterile interspecific hybrid crosses. Being able to bring forth genetic diversity in a petri dish creates the unique opportunity to conduct mouse genetic mapping at unprecedented speeds with "flow mapping" (similar to "X-QTL" in yeast (28)) or arbitrarily large panels unmatched by most other model organisms, except possibly yeast (22, 28). As renewable stem cells, IVR panels can be expanded, archived and shared, offering a cellular resource with many of the advantages sought from traditional community resources such as recombinant inbred line panels. Further, we have shown that our IVR method works in a broad range of ES cells. With millions of potentially recombinant (thus genetically distinct) ES cells in a petri dish, we demonstrated how IVR enabled mapping of QTLs for drug resistance in as few as 6 days (with an estimated total of 5 doublings over 5 days). Putting this in context, such an experiment using traditional mouse crosses would have taken 450 days, based on the typical mouse generation time of 90 days, assuming that hybrid sterility could be overcome and allowing for selfing.

Rather than a challenge to current organismal genetic mapping methods, we see IVR as a complementary extension. In the mouse, the largest organismal recombinant inbred (RI) panel BXD contains "only" ~160 lines (with most published work based on the ~35 original BXD strains) (29) and attempts in generating panels incorporating greater divergences encountered enormous challenges (30). Nevertheless, mouse RI resources still represented some of the most powerful tools available to dissect system genetics in the mouse, the prime biomedical model organism (31). Seen in this light, IVR represents an alternative approach that could greatly extend the available renewable resources, not least because the genotype combinations between divergent species are hitherto impossible to obtain in the first place.

Recently, Sadhu and coworkers have also achieved a major advance in genetic mapping using CRISPR/*Cas9*-mediated mitotic recombination in yeast (22). In contrast to CRISPR targeting, our transgene-free approach offers the simplicity of inducing genome-wide recombinants by the simple addition of a single inexpensive small molecule to the tissue culture medium. Going forward, we envision a combined, complementary approach to IVR: using BLM inhibition for mapping panel generations and efficient QTL identification, then switching to targeted transgene-based screening or CRISPR/*Cas9*-based IVR for fine-scale mapping.

We are currently working on improving the efficiency of IVR panel creation from hybrid cell lines and developing robust phenotyping protocols, in order to make mapping interspecific trait differences in mouse or other mammals using *in vitro* recombination routine. In addition to the traits we have investigated, *Mus spretus* and the *Mus musculus* laboratory mouse differ in a number of distinct traits, such as longevity and telomere lengths (32), cancer and inflammation resistance (33, 34) and metabolism (35). Many of these traits have tissue or cellular models suitable for IVR mapping panels or flow mapping through fluorescent detection of specific proteins or metabolites. Future experiments may also probe even greater evolutionary divergence: early work has shown that F1 hybrids spanning as much as 6 million years between *Mus musculus* and *Mus caroli* was viable (36). Given active development in single-cell genomics and disease modeling from patient-specific induced pluripotent stem cells (iPSC), including organoids or organ-on-a-chip microfluidics systems, we anticipate that the *in vitro* recombinant platform can be broadly applied to mouse, human or even other species to accelerate the identification of the genetic basis of many traits and diseases.

3.2.7 Materials and Methods:

Animal Care and Use

All animal experiments have been approved by the applicable animal welfare ethics committees: Faculty of Sciences, Ghent University, Belgium (reference number 06/022); and Landesdirektion Sachsen, Germany, permit number 24-9168.11-9/2012-5.

Cell Culture

Table S1 summarizes the ES cell lines used in this study. Unless otherwise stated, ES cell lines were cultured on SNL76/7-4 feeder cells in Attachment Factor (AF)-coated plates under 2i medium, supplemented with LIF (see SI Methods for details).

BLM Inhibition using ML216

BLM inhibition was performed using 25 μM ML216 (Sigma Aldrich) in 2i/LIF media on inactivated feeders.

Generation of HyTK-EGFP-Neo cell line

Starting with G4 ROSALUC B12 ES cells (37), we replaced the cassette at the *GtRosa26* locus with a cassette carrying two selectable markers, HyTK and enhanced green fluorescent protein (EGFP, selectable in fluorescence-assisted cell sorting; Fig. S1; see SI Methods for details). Successful replacement of the cassette was confirmed through selection using geneticin (G148; ThermoFisherScientific) and genotyping.

Colony Survival Assay

HyTK-EGFP-Neo cells were seeded at a density of 5×10^5 per 10 cm AF/feeder plate, followed by 25 μM ML216 treatment for 1 or 5 d. Prior to the start of negative selection, cells were re-plated at 2×10^5 per 10 cm AF/feeder plate and FIAU (0.2 μM, Sigma Aldrich) or ganciclovir (10 μM, Sigma Aldrich) selection was applied for 5 d. Colonies were stained with the Alkaline phosphatase kit (EMD Millipore, Billerica, MA, USA), photographed and counted to determine survival rates. The entire procedure was repeated multiple times and the survival colony counts averaged to determine IVR rate.

Screening for Spontaneous Recombinant ES Cell Colonies

Cells were treated with ML216 for 2 d at 5 μM, then 3 d at 25 μM. Cells were then re-plated and cultured for 5 d in 2i/LIF without ML216. One hundred and eighty-nine colonies were randomly picked (without selection), of which 136 were screened with multiplexed genotyping.

Multiplexed genotyping for detection of loss-of-heterozygosity (LOH)

Diagnostic markers between BL6, *CAST* and *SPRET* strains at tips of each chromosome were designed to track the presence of each allele. The markers were amplified with fluorescence-tagged primers as proposed in (38) (see SI Methods and Table S2 for details). The PCR reactions were pooled at equimolar proportions and analyzed with a 3730xl DNA Analyzer capillary sequencer (ThermoFisher Scientific, Germany) and scored for conversion from heterozygous into homozygous genotypes (LOH) at the tips of each chromosome.

6-TG treatment and DAPI exclusion assay

Prior to the main experiments, killing curves for 6-tioguanine (Sigma Aldrich) was performed using WST-1 assay (Roche) according to the manufacturer's instructions (Fig. S5). For the main experiments, S18 ES cell line was cultured for 5, 10 or 21 d with 25 μM ML216. Following the designated ML216 treatment, the cells were re-plated and treated with 25 μM 6-TG in 2i/LIF. We determined "live/dead" cell viability using 4',6-diamidino-2-phenylindole (DAPI) staining (1 μg/mL, Sigma Aldrich) after 1 d of 25 μM 6-TG treatment.

Fluorescence-Activated Cell Sorting (FACS)

Flow cytometry was performed using an Aria II Cell Sorter (Becton Dickinson GmbH, Heidelberg, Germany). We defined the *6-TG*^{R} and *6-TG*^{S} populations using the DAPI exclusion assay in reference ES cell populations. In sorting experiments, ML216-treated or control *6-TG*^{R} and *6-TG*^{S} population were recovered for sequencing. For quantification, we performed post-hoc analysis using the R Bioconductor package flowCore (39), principally by clustering using the forward scatter area (FSC-A) and DAPI/Pacific Blue-A channels into "Live" and "Dead" clusters using mclust v5.2 (40, 41) in 6-TG-treated experiments, considering ML216-treated and controls separately (Fig. S6).

Sequencing and analysis pipeline

Sequencing libraries for high-throughput sequencing were generated using Nextera DNA Library Prep Kit (Illumina, Inc., San Diego, USA) according to manufacturer's recommendations or equivalent purified *Tn5* transposase as described in (42). Each sample (FACS-sorted clones, single colonies or pooled cells)

was barcoded through PCR extension using an i7-index primer (N701-N763) and the N501 i5-index primer. Pooled libraries were sequenced by a HiSeq 3000 (Illumina) and analyzed using a custom pipeline (see SI Methods for details). We performed genotyping based on allelic coverage per megabase using known informative variants between the BL6, *CAST* and *SPRET* strains (Mouse Genomes Project version 3 dbSNP v137 release)(25). Scripts are available at: https://github.com/evolgenomics/HybridMiX .

Laser-assisted morula injection

Fully ES cell-derived embryos were obtained through injection into 8-cell stage embryos (morulae) as described in (27). At 14 d after the injection and subsequent embryo transfer into surrogates (approximating developmental stage E14.5), the gestation was terminated and embryos were dissected and fixed with 4% paraformaldehyde for microCT scanning.

Micro computer-tomography (microCT)

Soft tissue X-ray contrast staining was done via 4 d perfusion in 25% Lugol's, or iodine potassium iodide solution. Then the embryos were rehydrated, mounted in 1% agarose and scanned with a Skyscan 1173 instrument (Bruker Corporation, Billerica, MA, USA) at 9.96 micron resolution (0.5 mm aluminium filter, energy at 70 kV and 110 µA). Image analysis, segmentation and visualizations were performed using Amira v6.2.0 (FEI, Hillsboro, OR, USA) with the XImagePAQ extension 6.2.

3.2.8 Acknowledgements:

We thank Felicity Jones for experimental design, helpful discussion and input, and for improving the manuscript, Caroline Schmid for animal husbandry, Sebastian Kick for microCT scanning, the Chan and Jones Labs members for support, scientific input and improving the manuscript, Christa Lanz and Ilja Bezrukov for assistance with high-throughput sequencing and associated data processing; Andre Noll for high-performance computing support; Cornelia Grimmel and Stella Autenrieth for technical assistance with FACS; RV-L3-HyTK-2L and pCAG-Flpo were gifts from

Geoff Wahl and Massimo Scanziani (Addgene plasmids # 11684 & # 60662). The G4 ROSALUC ES cell line was a gift from Jody J. Haigh. We thank Hua Tang, David M. Kingsley, Karsten Borgwardt and Detlef Weigel for input and discussion on experimental design. J.P.L.C. is supported by the International Max Planck Research School “From Molecules to Organisms”. P.M. was supported by the Fulbright US Student Program. T.H. and C.L. are supported by Ghent University. Y.F.C. is supported by the Max Planck Society and a European Research Council Starting Grant #639096 “HybridMiX”.

3.2.9 References:

1. Darwin C (1859) *On the Origin of Species by Means of Natural Selection* (John Murray, London).
2. Nguyen G, et al. (2013) A small molecule inhibitor of the BLM *helicase* modulates chromosome stability in human cells. *Chem Biol* 20:55-62.
3. Hochepied T, et al. (2004) Breaking the species barrier: derivation of germline-competent embryonic stem cells from *Mus spretus* x C57BL/6 hybrids. *Stem Cells* 22:441-447.
4. Dejager L, Libert C, Montagutelli X (2009) Thirty years of *Mus spretus*: a promising future. *Trends Genet* 25:234-241.
5. Lander ES, Botstein D (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199.
6. Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10:565-577.
7. Orr HA (2001) The genetics of species differences. *Trends Ecol Evol* 16:343-350.
8. Churchill GA, et al. (2004) The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 36:1133-1137.
9. Nicod J, et al. (2016) Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. *Nat Genet* 48:912-918.
10. Parker CC, et al. (2016) Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nat Genet*

48:919-926.

11. Poltorak A, et al. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in *Tlr4* gene. *Science* 282:2085-2088.
12. Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J (2009) A mouse speciation gene encodes a meiotic histone *H3 methyltransferase*. *Science* 323:373-375.
13. Turner LM, White MA, Tautz D, Payseur BA (2014) Genomic networks of hybrid sterility. *PLoS Genet* 10:e1004162.
14. White MA, Ikeda A, Payseur BA (2012) A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. *Mamm Genome* 23:454-466.
15. White MA, Stubbings M, Dumont BL, Payseur BA (2012) Genetics and evolution of hybrid male sterility in house mice. *Genetics* 191:917-934.
16. Forejt J (1996) Hybrid sterility in the mouse. *Trends Genet* 12:412-417.
17. Burgio G, et al. (2007) Interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species: a powerful tool to dissect genetic control of complex traits. *Genetics* 177:2321-2333.
18. Guo G, Wang W, Bradley A (2004) Mismatch repair genes identified using genetic screens in *Blm*-deficient embryonic stem cells. *Nature* 429:891-895.
19. Yusa K, et al. (2004) Genome-wide phenotype analysis in ES cells by regulated disruption of *Bloom's syndrome* gene. *Nature* 429:896-899.
20. Yamanishi A, et al. (2013) Enhancement of microhomology-mediated genomic rearrangements by transient loss of mouse *Bloom syndrome* helicase. *Genome Res* 23:1462-1473.
21. Barakat TS, Rentmeester E, Sleutels F, Grootegoed JA, Gribnau J (2011) Precise BAC targeting of genetically polymorphic mouse ES cells. *Nucleic Acids Res* 39:e121.
22. Sadhu MJ, Bloom JS, Day L, Kruglyak L (2016) CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science* 352:1113-1116.
23. Johnson GG, Chapman VM (1987) Altered turnover of *hypoxanthine phosphoribosyltransferase* in erythroid cells of mice expressing *Hprt a* and *Hprt b* alleles. *Genetics* 116:313-320.
24. Mouse Genomes Project, Wellcome Trust Sanger Institute (2017) Data from Genome Evaluation Browser, SPRET_EiJ. (http://mice-

geval.sanger.ac.uk/SPRET_EiJ_R20150909/Info/Index)

25. Keane TM, et al. (2011) Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477:289-294.
26. Barakat TS, Gribnau J (2010) X chromosome inactivation and embryonic stem cells. *Adv Exp Med Biol* 695:132-154.
27. DeChiara TM, et al. (2010) Producing fully ES cell-derived mice from eight-cell stage embryo injections. *Methods Enzymol* 476:285-294.
28. Ehrenreich IM, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* 464:1039-1042.
29. Williams RW, Williams EG (2017) Resources for Systems Genetics. *Methods Mol Biol* 1488:3-29.
30. Collaborative Cross Consortium (2012) The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics* 190:389-401.
31. Andreux PA, et al. (2012) Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* 150:1287-1299.
32. Ding H, et al. (2004) Regulation of murine telomere length by *Rtel*: an essential gene encoding a helicase-like protein. *Cell* 117:873-886.
33. Mahieu T, et al. (2006) The wild-derived inbred mouse strain SPRET/Ei is resistant to LPS and defective in *IFN-β* production. *Proc Natl Acad Sci USA* 103:2292-2297.
34. Puimège L, et al. (2015) Glucocorticoid-induced *microRNA-511* protects against *TNF* by down-regulating *TNFR1*. *EMBO Mol Med* 7:1004-1017.
35. Song Y, et al. (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Curr Biol* 21:1296-1301.
36. West JD, Frels WI, Papaioannou VE, Karr JP, Chapman VM (1977) Development of interspecific hybrids of *Mus*. *J Embryol Exp Morph* 41:233-243.
37. Haenebalcke L, et al. (2013) Efficient *ROSA26*-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem Cell Rev* 9:774-785.
38. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18:233-234.
39. Hahne F, et al. (2009) flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics* 10:1.

40. Fraley C, Raftery AE (2002) Model-based clustering, discriminant analysis, and density estimation. *J Am Stat Assoc* 97:611-631.
41. Fraley C, (2012) *mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation.* (University of Washington, Seattle).
42. Picelli S, et al. (2014) *Tn5* transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24:2033-2040.

3.2.10 Supplementary Text:

Major differences between mitotic and meiotic mapping

As a novel mapping system, we observed a number of key differences between IVR and conventional meiotic genetic mapping. First, recombination tends to occur as loss-of-heterozygosity between the breakpoint and the telomeres. Unlike conventional breeding with random assortment, under IVR in F1 hybrids, outcrosses are not possible. As a result, we tend to observe only heterozygous genotypes near the centromere, with informative crossovers almost always found between the centromere and a selectable QTL but not on the telomeric side. This asymmetry often led to a plateau in the association profiles from the QTL towards the telomeres on a given chromosome (Fig. 2*B* and 3*C*), an effect also reported elsewhere (21). As a consequence, interval mapping in IVR analogous to those in meiotic panels yields excellent genetic resolution on the centromeric side but poor resolution on the telomeric side (see distribution of crossover directions and breakpoints in Fig. 3*C* and (21). Second, access to common tissue culture methods under IVR greatly mitigates typical concerns such as panel sizes and power calculation in generating meiotic mapping panels. Since it is trivial to freeze samples and introduce selectable markers at any given locus or targeted chromosome breaks with a CRISPR/Cas9 panel (21) with ES cells under tissue culture conditions, refinement of mapping resolution under IVR no longer depends on the diminishing return of breeding and screening for increasingly rare informative recombinants. To underscore this point, our flow mapping experiment for 6-TG susceptibility identified the single region containing the known causal *Hprt* gene within weeks. Third, while it is true that mitotic recombination as used in IVR depends on error-prone repair of double-strand breaks that could affect phenotype through chromosome rearrangements and new mutations at breakpoints, two observations from our results may moderate this concern. One, we did not observe elevated aneuploidy under ML216 treatment, suggesting that IVR did not elevate rates of chromosome rearrangements (Fig. S2). Two, millions of variants already exist in our (BL6 x *CAST*) F1 or (BL6 x *SPRET*) F1 lines. These variants vastly outnumber any new mutations generated through IVR. Assuming a typical spectrum of mutation effects, these parental variants likely would contribute far more to trait variance than new mutations arising in a specific line.

Since genetic mapping depends on testing for different genotypic effect of an allele across all lines carrying the same genotype at loci that are typically megabases away from a random double-strand breakpoint, it is reasonable to expect that the mutagenic effect of mitotic recombination should have a rather limited impact on genetic mapping. Under the flow mapping design, the mutagenic effect of mitotic recombination is further diluted, because millions, if not tens of millions of cells in bulk population cultures are phenotyped and sequenced as pools. This conclusion is supported by our ability to locate and map various transgenes or QTLs in this current study. We are nonetheless in the process to formally characterize the relative contribution to trait variation due to the mutagenic effects of IVR and that of the parental genomes.

3.2.11 Supplementary 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 Faculty of Sciences, Ghent University, Belgium, (reference number 06/022); 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. Sequence data have been deposited in the GEO database under accession number PRJNA390071.

Cell Culture

Unless otherwise stated, murine stem cell lines have been cultured on Attachment Factor Protein (AF) (ThermoFisherScientific, Schwerte, Germany) coated cell culture dishes on inactivated SNL 76/7-4 feeder cells (“feeder” plates; SCRC-1050, ATCC, Middlesex, United Kingdom) and using 2i/LIF media as follows: KnockOut Serum Replacement (ThermoFisherScientific), KnockOut DMEM (ThermoFisherScientific), 2-Mercaptoethanol, 1000x, 55 mM

(ThermoFisherScientific); MEM Non-Essential Amino Acids Solution, 100x (ThermoFisherScientific); GlutaMAX Supplement, 100x (ThermoFisherScientific); 3 μM GSK-3 inhibitor CHIR99021 (Sigma Aldrich, Munich, Germany); 1 μM MEK inhibitor PD0325901 (Sigma Aldrich); insulin solution, human (Sigma Aldrich), 1000 U/mL recombinant mouse LIF (expressed in-house). Unless otherwise stated, cell culture media was replaced daily.

BLM inhibition using ML216

BLM inhibition was performed using 25 μM ML216 (Sigma Aldrich) in 2i/LIF media on inactivated feeders. Killing curves for ML216 was performed using the WST-1 assay (Roche, Basel, Switzerland) according to the manufacturer's instructions.

Plasmid construction

The pMK11 plasmid was constructed by blunt-end ligation of the pRMCE-DV1 plasmid's backbone, after excision of its chloramphenicol–*ccdB* cassette between the EcoRV and SbfI sites, and replacing it with the HindIII-excised hygromycin phosphotransferase-thymidine kinase cassette (HyTK) from the RV-L3-HyTK-2L plasmid (43) (Plasmid # 11684, Addgene, Cambridge, USA). The final pMK11 construct contained flanking *FRT wt* and *FRT mutant* sites for recombinase-mediated cassette exchange detailed below.

Generation of HyTK-EGFP-Neo cell line

G4 ROSALUC B12 ES cells (37) were co-transfected with pMK11 described above and *FLP* mRNA (StemMACS *Flp* Recombinase, Miltenyi Biotec, Bergisch Gladbach, Germany) or pCAG-Flpo (44) (Plasmid # 60662, Addgene) using Lipofectamine 2000 (ThermoFisherScientific). We replaced the cassette at the *GtRosa26* locus with a cassette carrying two selectable markers, HyTK and enhanced green fluorescent protein (EGFP, selectable in fluorescence-assisted cell sorting; Fig. S1). Successful replacement of the cassette with a re-activated neomycin resistance gene was selected for with 200 μg/mL geneticin (G148; ThermoFisherScientific). Resistant colonies were picked after 7 days (d) of selection and further expanded. Correct replacement was confirmed by junction PCR with primers SA_loxP_Rev: 5'–GCGGCCTCGACTCTACGATA–3' and

ROSA26_3HA_F_BamHI: 5'–GCGGGATCCCCTCGTGATCTGCAACTCC–3'. The presence of an intact BL6 wildtype allele was confirmed by an alternative reverse primer oIMR8545 5'–AAAGTCGCTCTGAGTTGTTA–3'. PCR was performed as a quantitative PCR reaction. See "RNA Extraction, Reverse Transcription and "Real Time PCR" section below for more details.

Colony Survival Assay

HyTK-EGFP-Neo cells were seeded at a density of 5×10^5 per 10 cm AF/feeder plate. Eight hours (h) following the plating, 25 μM ML216 treatment was initiated and continued for 1 or 5 d. Prior to the start of negative selection, c**ells were re-plated at** 2×10^5 per 10 cm AF/feeder plate and FIAU (0.2 μM, **Sigma Aldrich) or ganciclovir (10 μM, Sigma Aldrich) selection was initiated after 1d and** continued for 5d.

In order to determine the plating efficiency after ML216 treatment, cells were plated at 1×10^3 per 6 cm AF/feeder dish. Colonies were stained with the Alkaline phosphatase kit (EMD Millipore, Billerica, MA, USA). Before the application of negative selection, 20 random views of each plate were taken using an EVOS FL Cell Imaging System (ThermoFisherScientific) and counted using Fiji v2.0.0-rc-54/1.51h (45).

Screening for spontaneous recombinant ES cell colonies

Cells were plated at a density of 1×10^5 per 3.5 cm AF plate. Treatment with 5 μM ML216 was initiated 16 h after plating, continued for 2d and then followed by 3d of 25 μM ML216 treatment. Cells were then re-plated on a 10 cm AF plate and cultured for 5d in 2i/LIF without ML216. One hundred and eighty-nine colonies were randomly picked and 153 were expanded for multiplexed genotyping.

Multiplexed genotyping for detection of loss-of-heterozygosity (LOH)

Diagnostic insertions or deletions (indels) between BL6, *CAST* and *SPRET* strains that are greater than 20bp in length and located within the most distal 10Mbp of each chromosome were filtered from the publicly available variant panel from the Mouse Genomes Project made available by the Wellcome Trust Sanger Centre (v5 dbSNP v142 release) (23, 24) using VCFtools v0.1.14 (46). Automated primer design was carried out with Primer3 v.1.1.3 using the following parameters: PRIMER_OPT_SIZE=20; PRIMER_MIN_SIZE=18; PRIMER_PRODUCT_OPT_SIZE=300;

PRIMER_PRODUCT_SIZE_RANGE=250-400; PRIMER_MAX_SIZE=23; PRIMER_NUM_NS_ACCEPTED=1; PRIMER_LEFT_MIN_TM=58 PRIMER_LEFT_MAX_TM=62; PRIMER_RIGHT_MIN_TM=58; PRIMER_RIGHT_MAX_TM=62; PRIMER_MAX_DIFF_TM=2; PRIMER_MIN_GC=45.0; PRIMER_MAX_GC=85.0; PRIMER_MAX_POLY_X=3; PRIMER_SELF_ANY=4. Among indels with successfully designed primer pairs, the most telomeric amplicons were chosen and an extension was added to either the forward (M13F) or reverse (M13R-46) oligonucleotide to allow for easy fluorophore incorporation as described (38). The amplicon sizes were further optimized following pilot capillary sequencer runs to avoid amplicon size overlap in a multiplexed run. All primers and expected fragment sizes are listed in Table S1.

For genotyping of cell colonies, primers pairs were pooled into 4 multiplexed PCR reactions. Group 1 (Chr1, Chr7, Chr13, Chr14 and Chr18) and Group 2 (Chr3, Chr6, Chr16, Chr17 and Chr19) primer mixes contained 2 and 4 µM of forward and reverse primers, respectively, for each listed chromosome plus 20 µM of the universal FAM-labeled M13F_FAM primer. Chromosomes 13 and 17 primers were mixed at 6 and 12 µM concentration. For Group 3 (Chr2, Chr4, Chr5, Chr11, ChrX) and Group 4 (Chr8, Chr9, Chr10, Chr12, Chr15) mixes, the forward and reverse primers were mixed at 4 and 2 µM concentration, along with the HEX-labelled M13R-46_HEX primer at a concentration of 20 µM. QIAGEN Multiplex PCR *Plus* Kit (Qiagen, Hilden, Germany) was used according to manufacturer's recommendations (including the addition of 5× Q-Solution) at 10 µL final reaction volume with 3 to 10 ng of DNA per PCR reaction. The PCR program used was: 95°C for 15 min, then 52 cycles of 94°C for 30 s; Group-specific annealing temperature for 2.5 min; and 72 ºC for 1 min; followed by a final extension at 72°C for 30 min and hold at 4 ºC. The group-specific annealing temperatures were: Group 1: 63°C; Group 2: 63.8°C; Group 3: 57°C; and Group 4: 64°C. Then the PCR reactions were pooled at equal 1 µL proportions and analyzed with a 3730xl DNA Analyzer capillary sequencer (ThermoFisher Scientific, Germany) using the fragment analysis program with the G5-RCT Dye Set. Electropherogram traces were analyzed with the Microsatellite Plugin in Geneious v7.1.9 (47).

6-TG treatment and DAPI exclusion assay

Prior to the main experiments, killing curves for 6-tioguanine (Sigma Aldrich) was performed using WST-1 assay (Roche) according to the manufacturer's instructions (Fig. S5). For the main experiments, S18 ES cell line was cultured for 5, 10 or 21d with 25 µM ML216 prior the treatment with 25 µM 6-TG in 2i/LIF starting from an initial seeding concentration of 1×10^5 cell per 3.5 cm AF plate. To avoid overcrowding, at day 3 of the ML216 treatment colonies were dissociated using Accutase Cell Dissociation Reagent (ThermoFisherScientific) and re-seeded on a 10 cm AF-plate. At day 5, the cells were moved to a 15 cm AF plate prior to 6-TG treatment. After 16 h, 6-TG in 2i/LIF was added at a concentration of 25 µM. From each plate 2.5×10^4 cells were used to continue the experiment until day 10 or 21. 4',6-diamidino-2-phenylindole (DAPI) staining (1 µg/mL, Sigma Aldrich) was employed for "live/dead" cell viability determination after 1d of 25 µM 6-TG treatment. Briefly, ES cells were treated with ML216 and/or 6-TG to induce IVR and cell death, respectively. Colonies were dissociated using Accutase and re-suspended in phosphate buffered saline (PBS) within 1 h of analytical or preparative fluorescence-activated cell sorting (FACS). For details on FACS see below.

Fluorescence-Activated Cell Sorting (FACS)

Flow cytometry was performed at the University Clinic Tübingen Dermatology Clinic FACS Core Facility using an Aria II Cell Sorter (Becton Dickinson GmbH, Heidelberg, Germany). To determine cell viability, we performed the DAPI exclusion assay. After excluding cell aggregates, we defined the *6-TG^R* and *6-TG^S* populations using conservative interval gates based on evaluating the data from reference flow experiments with 6-TG-treated DAPI-stained ES cells. For cell population evaluations, flow cytometry data was exported from BD FACSDiva Software v8.0.1 (Becton Dickinson). We carried out basic data handling and $\log_{10}$ transformation using the R Bioconductor package flowCore (39). Since live and dead cells cluster also in other measurements, we took both forward scatter area (FSC-A) and DAPI into account for our quantification, rather than using a simple interval gate on the DAPI/Pacific Blue-A channel. We defined data-driven "Live" and "Dead" clusters using mclust v5.2 (40, 41) in 6-TG-treated experiments, considering ML216-treated and controls separately. We then classified each cell in to the "Live" and "Dead" clusters, applying a 5% uncertainty cut-off. "Live" and "Dead" proportions were then

calculated from the confidently assigned cells. Data was visualized using the package flowViz (48) (Fig. S5).

RNA Extraction, Reverse Transcription, and Real Time PCR

RNA was isolated using TRIzol Reagent (ThermoFisherScientific) with a single-step method following (49). Complementary DNA (cDNA) was generated using High-Capacity cDNA Reverse transcription kit (ThermoFisherScientific) with 500 ng of RNA per reaction according to the manufacturer's instructions. The newly synthesized cDNA (20 µl reaction) was diluted 5-fold and quantitative PCR (qPCR) was performed with SYBR-select Master Mix for CFX (ThermoFisherScientific) using a CFX384 Real-Time PCR system instrument (BioRad, Munich, Germany). We used the following primers for allele-specific amplification and detection:

Hprt^{a} (*SPRET*) forward: 5'–CAAAGCCTAAGAGCATGAGCGC–3',

reverse: 5'–CAGAGGGAACTGATAGGCTGGC–3', amplicon size: 229bp;

Hprt^{b} (BL6) forward: 5'–GCCAAATACAAAGCCTAAGATGAGCG–3',

reverse: 5'–CCAGCCTACCCTCTGGTAGATTG–3', amplicon size: 236bp. The standard CFX mode for Tm ≥ 60°C was used, with the following thermocycling program: 50ºC for 2 min, 95ºC for 2 min, followed by 35 cycles of 95ºC for 15 s, 60ºC for 1 min. Melting curve analysis over 80 steps of 0.5ºC increments was performed and curves inspected to ensure uniform annealing.

Immunofluorescence staining

ES cells were cultured for 3d on 12 mm cover glasses pre-coated with AF and feeder layer. Cells where then fixed 10 min in 4% paraformaldehyde, permeabilized 10 min in 0.25% Triton X, blocked in 5% serum for 1 h at room temperature. ES cell colonies were stained with anti-*Nanog* (1:100, rabbit, Cat # ab80892, Abcam, Cambridge, UK) antibodies for 2 h at room temperature and conjugated secondary antibody (1:400, anti-rabbit Alexa 467) for 1 h at room temperature. Nuclei were counter-stained for 5 min with DAPI at 1 µg/mL, mounted with ProLong Diamond Antifade Mountant (ThermoFisherScientific) and imaged using an AXIOVERT 200M inverted microscope (Zeiss, Oberkochen, Germany)

Karyotyping

Metaphase spreads were prepared from Control and ML216-treated ES cells under 2i conditions (5d culture for treatment on the original S18 background, 2d for the IVR lines 1 and 2; see Cell Culture above for detailed description of culturing conditions). Metaphase spreads were prepared essentially as described in (50) with the following modifications. Cells were initially plated at a density of 2×10^5 cells per 10 cm AF-coated culture dish. Spreads were mounted with ProLong Diamond Antifade Mountant (ThermoFisherScientific). Metaphase chromosomes were imaged with a 63x objective in a Zeiss APOTOME AXIO Imager.Z1 (Zeiss) equipped with an Orca-flash4.0 digital camera (C11440-22CU, Hamamatsu, Herrsching am Ammersee, Germany) and coupled to HCImage v4.3.5.8 image acquisition software. Chromosomes were anonymized and independently counted twice manually in Fiji v2.0.0-rc-54/1.51h using the multi-point tool.

Sequencing and analysis pipeline

Sequencing libraries for high-throughput sequencing were generated using 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 (42). Briefly, genomic DNA was extracted from FACS-sorted clones, single colonies or pooled samples 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 "tagmented" by commercial or purified *Tn5*-transposase. Each tagmented DNA sample was then PCR amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs) using barcoded i7-index primer (N701-N763) and the N501 i5-index primer. Pooled libraries were sequenced by a HiSeq 3000 (Illumina) at the Genome Core Facility at the MPI Tübingen Campus. Sequenced data were processed using a custom pipeline consisting of data clean-up, mapping, base-calling and analysis based upon fastQC v0.10.1 (51); trimmomatic v0.33 (52); bwa v0.7.10-r789 (53); GATK v3.4-0-gf196186 modules MarkDuplicates and IndelRealignment (54, 55); samtools v1.2 (56, 57); bcftools v1.2 (58); and R v 3.2.0 (59). Genotype calls were performed against known informative single and multiple nucleotide variants between C57BL/6NJ, CAST/EiJ and SPRET/EiJ strains made available by the Wellcome Trust Sanger Centre (Mouse Genomes Project version 3

dbSNP v137 release) (25). Coverage depths for the reference and alternative alleles were calculated based on the DP4 field in the variant VCF files. For flow mapping, each pool was sequenced to an average of 9.8x coverage of the mouse genome, corresponding to an average of 100272 ± 32692 read-depth over informative SNPs in each of 1 Mbp bin, our analytical unit for flow mapping. For individual clones, each clone were sequenced to an average of 0.95x, yielding on average 1841 ± 1159 reads per 1 Mbp bin. Crossover breakpoints were called by TIGER (60), using default parameters. Custom Perl scripts were used to process files prior to plotting and visualization in R. Scripts have been deposited in the following repository: https://github.com/evolgenomics/HybridMiX.

Statistical analysis

Chromosome loci affecting *6-TG* susceptibility were evaluated using the following statistical methods: smoothing function, window-based analyses. For the smoothing function, we first calculated the raw *SPRET* excess using the following formula: (*SPRET* – BL6) / (SPRET + BL6) fraction in each adjacent megabase window. Then in cases with two contrasting *6-TG*^{R} and *6-TG*^{S} pools (5d and 21d ML216 treatments), a *differential* SPRET excess was calculated by subtraction (*6-TG*^{S} – *6-TG*^{R}). Otherwise the raw *SPRET* excess for the sole pool (*6-TG*^{R} after 10d ML216 treatment) was reported. Then a LOESS smoothing function with 10% span was applied along the chromosome to obtain both the fit and standard error of the estimates. We then reported the regions in the genome with a LOESS fit that is greater than 1 s.e.m. away from 0 as coloured bars in Fig. 3C (a 0 indicates no bias for either *SPRET* or BL6 reads). Between the 5d and 21d ML216 treatments we found only such bias on ChrX, with one of the common region overlapping *Hprt* (5d: 49–60 Mbp; 10d: 42–169 Mbp; 21d: 24–81 Mbp) and additional telomeric regions (Fig. 3C). For the window-based analysis, we broke each one megabase window into up to a hundred 10 kbp bins and estimated the *SPRET* bias. We separately obtained a genome-wide outlier threshold at $P \leq 0.05$ (after applying Bonferroni correction). Any window lying greater than 1 s.e.m. beyond the threshold are considered genome-wide significant (Fig. S6A). All but one outlier windows in any treatment are found on ChrX, with the main cluster falling in the same region containing *Hprt*. Both methods converge towards the same area on ChrX, indicating the known role of *Hprt* in mediating *6-TG* susceptibility.

Laser-assisted morula injection

ES cell-derived embryos were obtained essentially as previously described in (27). Briefly, female C57BL/6NCrl mice were mated and host embryos harvested. ES cells from untreated S18 line and two IVR lines were injected into 8-cell stage embryos (morulae) after perforation of the zona pellucida with a laser pulse. The introduction of excess ES cells were expected to produce embryos with fully ES cell contributions. This was confirmed using coat-colour analysis from previous injections performed under R.N. After incubation for 1.5–2 h, injected embryos were transferred into the oviducts of E0.5 pseudo-pregnant CD1-ICR female foster mice. The host mice were monitored for recovery and development. At 14d after the embryo transfer (approximating developmental stage E14.5), the gestation was terminated and embryos were individually dissected, fixed with 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. Due to sample preparation and fixation, genotyping of microCT-scanned embryos were not performed. Instead, control (unscanned) embryos were dissociated and genotyped at diagnostic loci using Taqman probes designed to specifically anneal to allele-specific polymorphisms on Chr14 and ChrX (Fig. S7 and Table S1). Quantitative real-time PCR was performed on C1000 Touch Thermal Cycler (Bio-Rad) with TaqMan Universal Master Mix II, with UNG (Life Technologies), 20 ng template DNA and 1x PrimeTime qPCR Assay. Genomic DNA controls from *SPRET* and BL6 mice were used in mixed pool of known proportions. At this point we cannot formally rule out that ES cell contributions to the embryo proper can be incomplete due to reasons related to lower ES cell viability (e.g., genetic incompatibilities) compared to typical results from morulae injections involving only laboratory strains (27).

Micro computer-tomography (microCT)

Prior to scanning, embryos were perfused for 4d in 25% Lugol's, or iodine potassium iodide solution. Contrast-stained embryos were rehydrated and mounted in 1% low-melting agarose and scanned with a Skyscan 1173 instrument (Control software version 1.6, Build 15; Bruker Corporation, Billerica, MA, USA) at 9.96 micron (µm) resolution using a 0.5 mm aluminium filter with energy settings at 70 kV and 110 µA. Volume reconstructions were performed using NRecon v.1.6.10.4

(Bruker Corporation) using parameters determined based on fine-tuning for each scan (misalignment correction: 23–30; beam-hardening correction: 25%; ring-artifact correction: 10; no smoothing). Image analysis, segmentation and visualizations were performed using Amira v6.2.0 (FEI, Hillsboro, OR, USA) with the XImagePAQ extension 6.2.

3.2.12 Supplementary References :

43. Wong ET, et al. (2005) Reproducible doxycycline-inducible transgene expression at specific loci generated by *Cre*-recombinase mediated cassette exchange. *Nucleic Acids Res* 33:e147.
44. Xue M, Atallah BV, Scanziani M (2014) Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* 511: 596-600.
45. Schindelin J, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9:676-682.
46. Danecek P, et al., 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156-2158.
47. Kearse M, et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-9.
48. Ellis B, Gentleman R, Hahne F, Meur NL, Sarkar D (2016) flowViz: Visualization for flow cytometry. *R package version 1.36.2*.
49. Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* 162:156-159.
50. Campos PB, Sartore RC, Abdalla SN, Rehen SK (2009) Chromosomal spread preparation of human embryonic stem cells for karyotyping. *J Vis Exp* 31:1512.
51. FastQC. (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; Accessed: August 5, 2016).
52. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114-2120.
53. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-

Wheeler transform. *Bioinformatics* 26:589-595.

54. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297-1303.
55. DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 43:491-498.
56. Li H, et al., 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.
57. Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987-93.
58. "Multiallelic calling model in bcftools (-m)." (https://samtools.github.io/bcftools/call-m.pdf. Accessed 26.Jan.2017.)
59. Team RC, (2015) R: A language and environment for statistical computing. *R: A language and environment for statistical computing*.
60. Rowan BA, Patel V, Weigel D, Schneeberger K (2015) Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3 (Bethesda)* 5: 385-98.

3.2.13 Supplementary Figures:

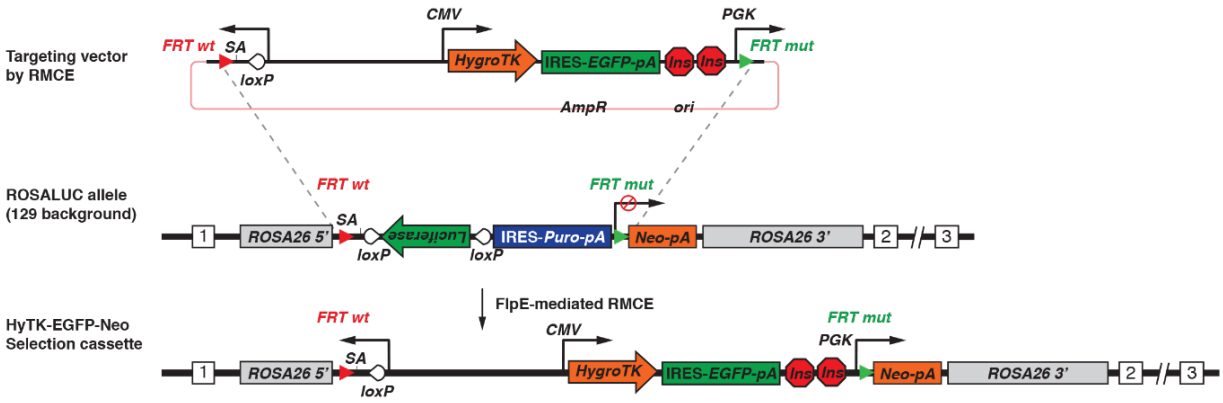


Fig. S1. **Site-specific integration of a versatile selection reporter cassette into the G4 ROSALUC ES cell line.** Utilizing the recombination-mediated cassette exchange (RMCE) technique, the targeting vector was inserted by a *Flp*-recombinase into the ROSALUC allele as previously described (37). The vector introduced the hygromycin phosphotransferase-thymidine kinase (HyTK) fusion selectable marker, the enhanced green fluorescent protein (EGFP) and the phosphoglycerate kinase 1 (PGK) promoter, thus restoring the expression of the latent neomycin resistance gene upon the successful integration of the vector into the ROSALUC allele. Figure modified from (37).

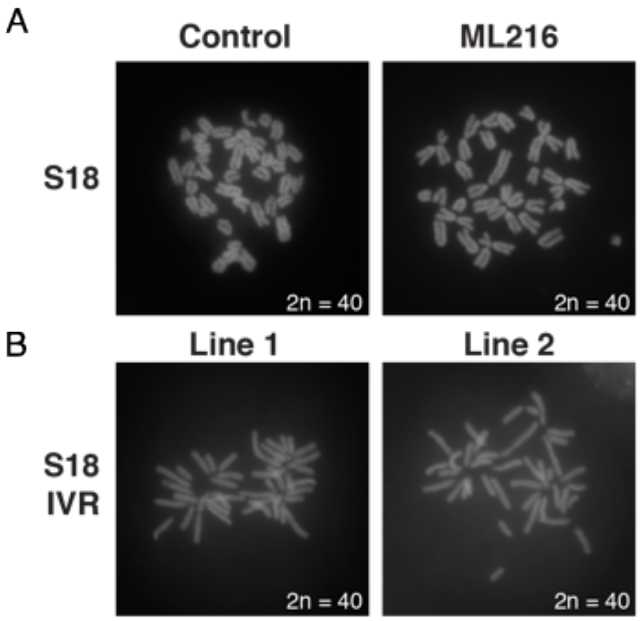


Fig. S2. Normal karyotypes were maintained under culturing and IVR treatment. (A) Representative metaphase spreads from S18 line under control and ML216 treatment show normal karyotype of 2n = 40. **(B)** After confirmed IVR treatment, selected lines 1 and 2 were chosen for re-derivation. The karyotypes of both lines are also normal with high levels of euploidy. Whole embryos derived from laser-assisted morulae injection results showed that the S18 line, and IVR lines 1 and 2 are broadly competent to differentiate into diverse cell lineages (Fig. 4, S7 & S8).

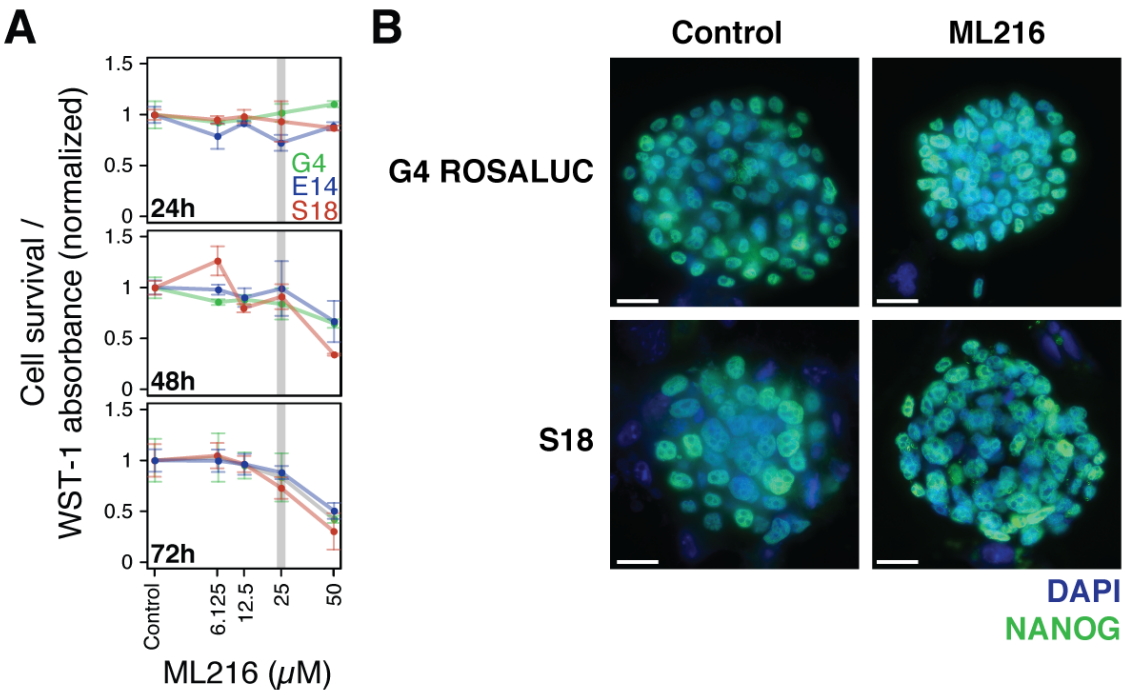


Fig. S3. ML216 treatment is compatible with ES cell culturing. (**A**) Determination of cell survival under culturing with varying ML216 concentrations. It should be noted that rather than a selection treatment or kill curve, here our interest is to determine the highest concentration that is compatible with good survival. Based on the results 25 μM (grey, n.s.) was chosen. (**B**) To determine if ML216 treatment affect ES cell colony viability and maintenance of stemness, we cultured ES cells strains G4 Rosaluc [(BL6 x 129S) F1] and S18 [(BL6 x *SPRET*) F1] under control 2i/LIF and 25 μM ML216 plus 2i/LIF conditions for 3 days. Both control and ML216 treated colonies showed good colony morphology, cell density and robust stem cell marker NANOG expression in both ES cell lines. We concluded that ML216 induction of *in vitro* recombination is compatible with ES cell culturing across considerable evolutionary divergence. Scale bar = 200 μm.

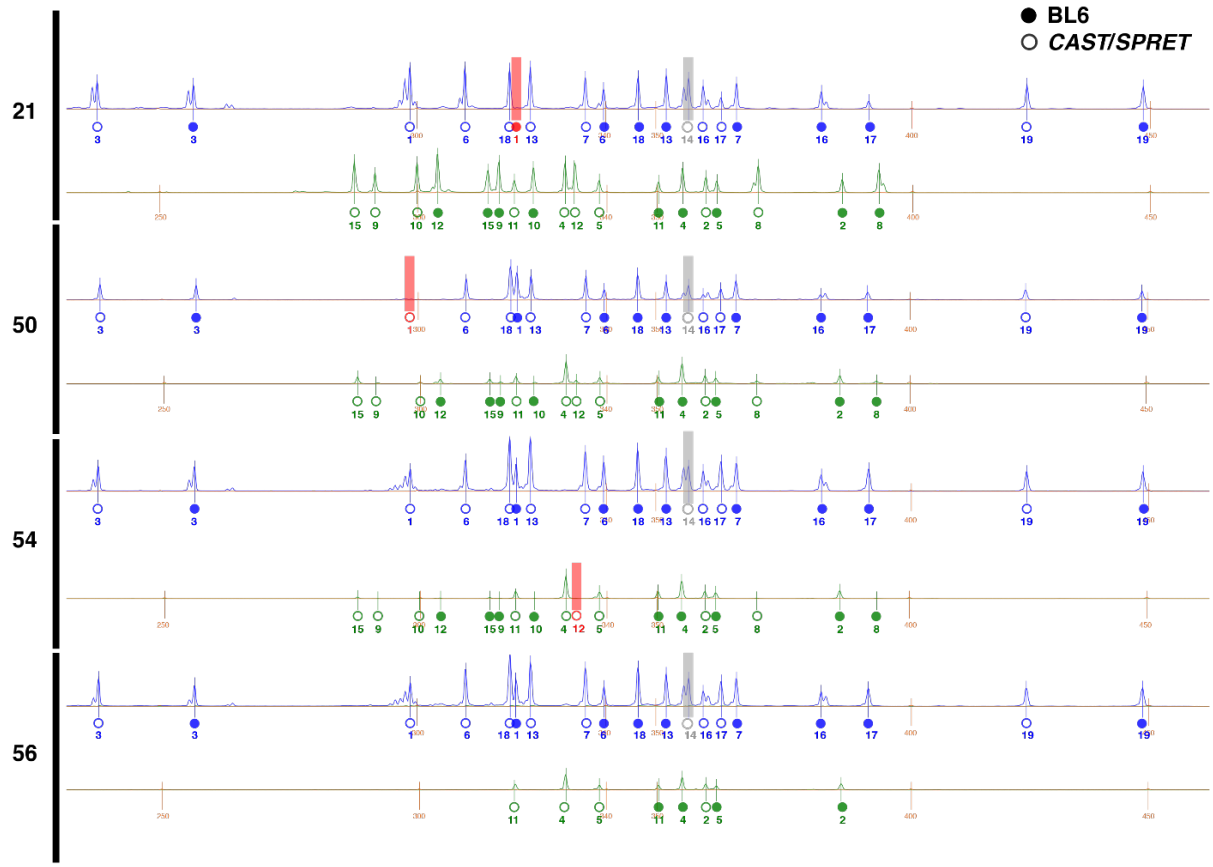


Fig. S4. Multiplexed PCR genotyping screen for spontaneous recombinants. Hybrid ES cells [(BL6 x *CAST*) F1 hybrid line "E14"] were treated with ML216 and screened by multiplexed PCR genotyping at diagnostic markers within the last 10Mbp of each autosome chromosome (see Methods & Table S1). Amplified fragment sizes were determined using a capillary sequencer. The markers were designed such that they show staggered fragment sizes, allowing clear identification using fragment analysis software. Shown above are the electropherogram traces corresponding to the clones shown in Fig. 2, out of 46 total clones. The blue (FAM) and green (HEX) channels are shown separately for each sample, adjusted according to size standards (LIZ, orange, in basepairs. Fluorescence levels are shown on arbitrary units on the Y-axis). Genotype calls corresponding to BL6 (solid circles) and *CAST* or *SPRET* (open circles) alleles for each chromosome are shown underneath the called peaks (markers were designed for both outgroups. Only E14 analyses are included in this study). Missing genotypes indicative of recombination or LOH events are indicated in red. Chromosome 14 calls were removed due to

invariant calls in all samples, including untreated F1 hybrid cells. This approach allowed us to rapidly screen through many colonies to detect possible recombinants. Notably, whole genome sequencing results suggested that in addition to the typical recombination events recovered by this multiplexed fragment analysis, further recombination events may occur elsewhere in the genome.

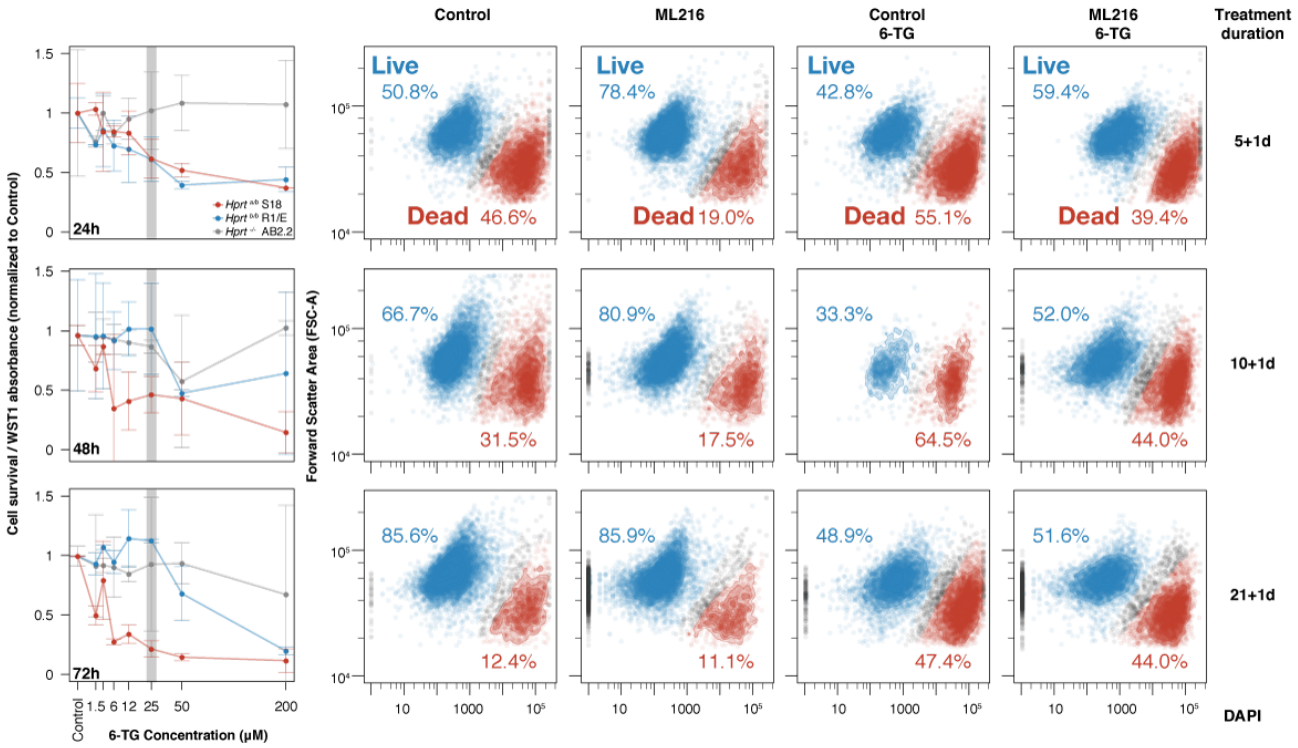
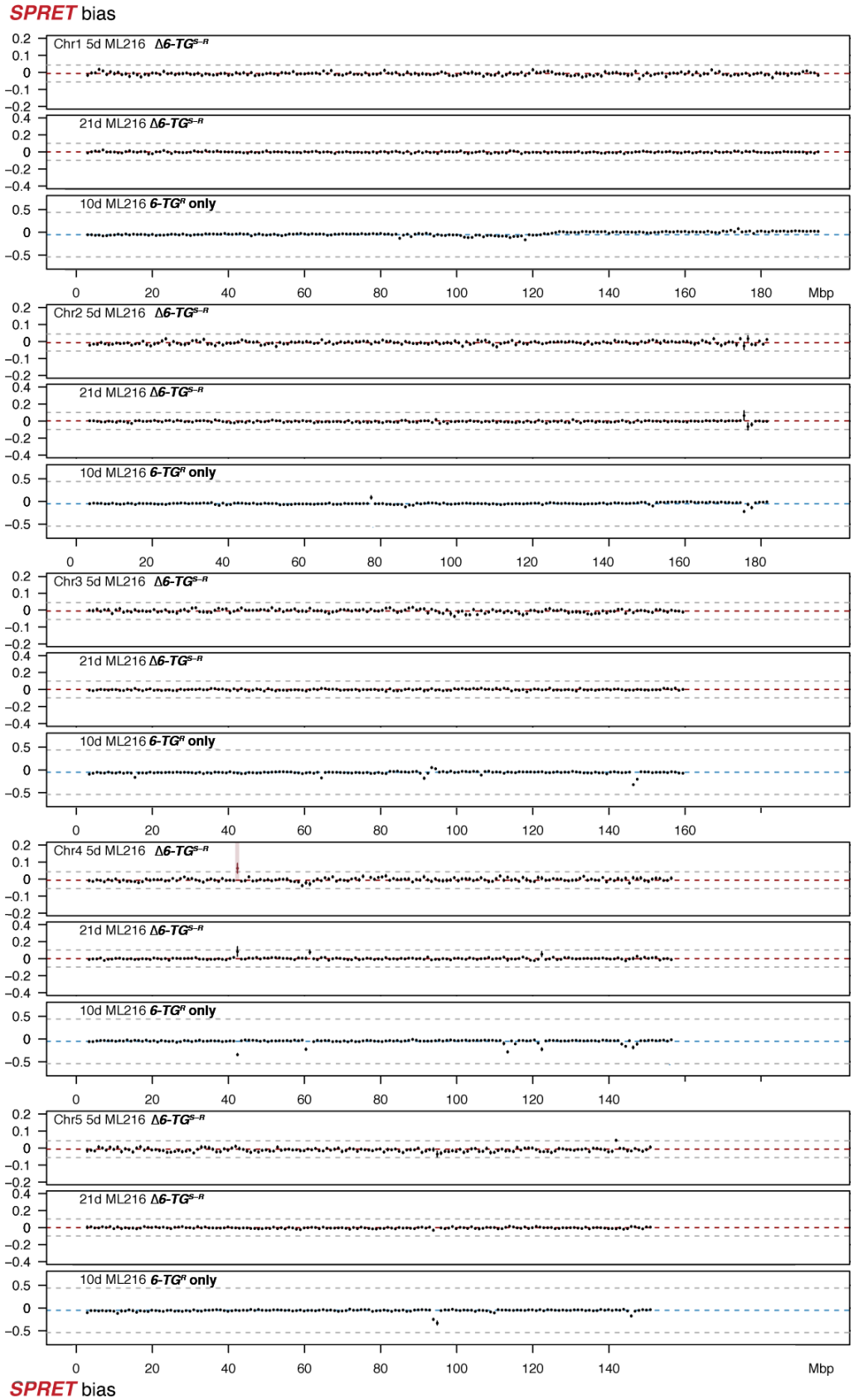
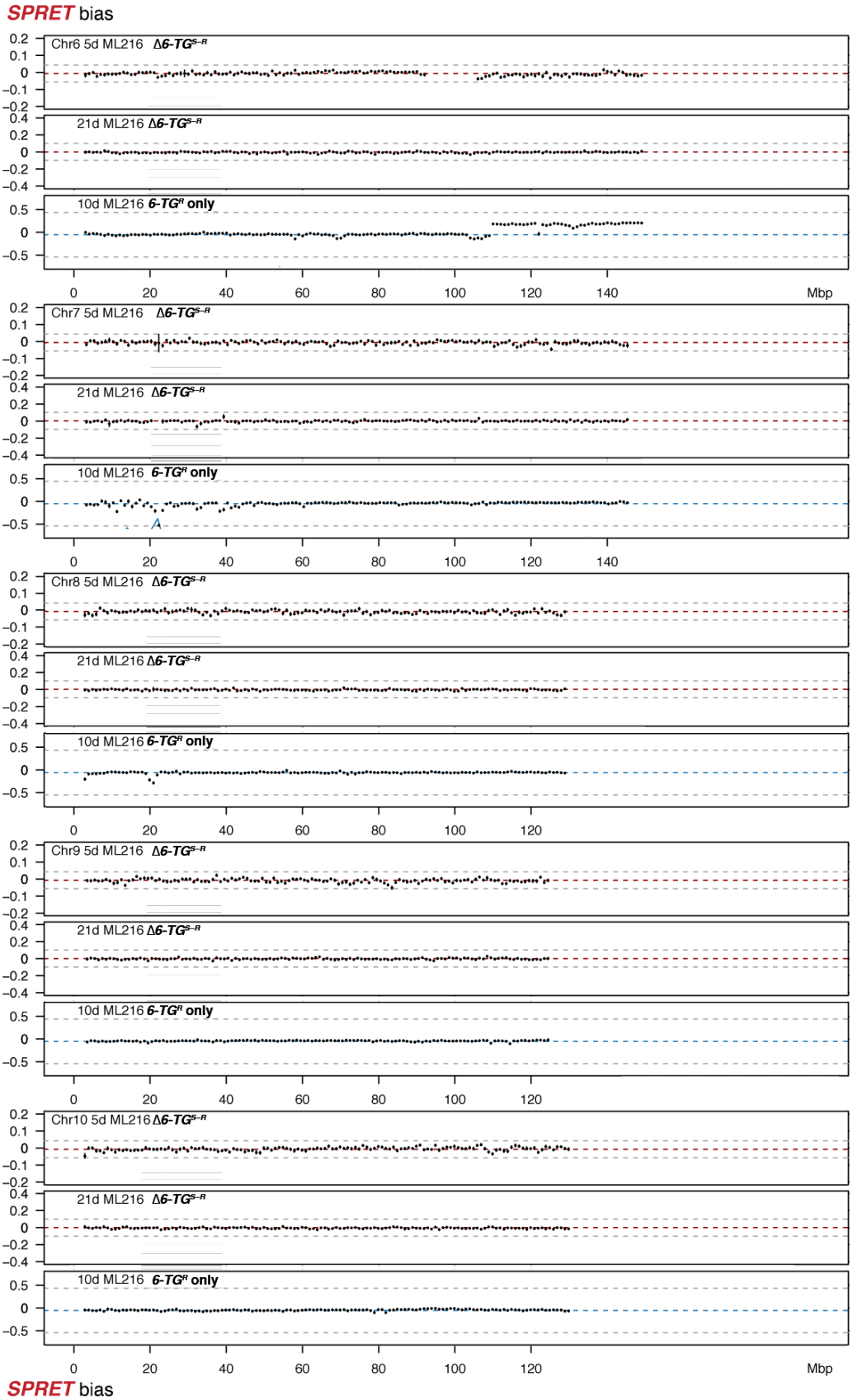


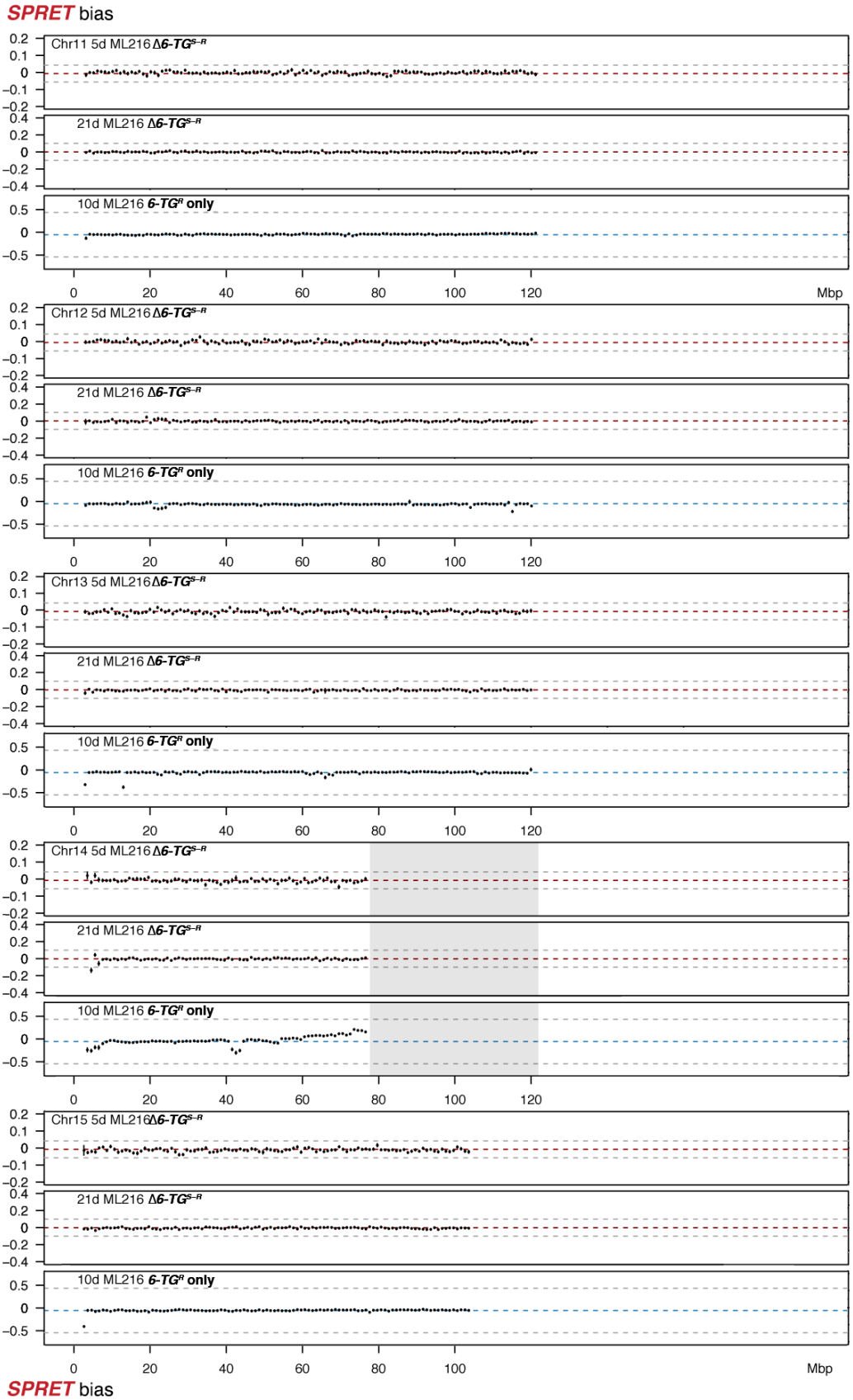
Fig. S5. Optimal 6-TG concentration for differential *Hprt*-dependent cytotoxicity. (**Left**) Concentration for 6-TG treatment was determined by treating ES cells with concentrations ranging from 1.5 to 200 μM. Cell survival were determined by a colorimetric WST-1 absorbance assay. ES cells carrying different *Hprt-a*, *-b* or null alleles on various genetic backgrounds were assayed in duplicates over 24, 48 or 72 hours (*Hprt*$^{a/b}$ on (BL6 x *SPRET*)F1 S18 background: red; *Hprt*$^{b/b}$ on R1/E 129X1/129S1 background: blue; and *Hprt*$^{-/-}$ on AB2.2 129S5 background: grey). Absorbance values were normalized against control treatment of no 6-TG after subtracting blank measurements. We chose 25 μM 6-TG treatment for subsequent experiments for the strong survival difference between cells carrying *Hprt*^{a} and those carrying *Hprt*^{b} or null genotypes after 48 h. To ensure genome

integrity for sequencing in flow mapping, we performed FACS already after 24 h of 25 μM 6-TG treatment together with a more sensitive DAPI exclusion cell viability assay. Plotted values are normalized mean between replicates ± s.d.

(**Right**) S18 cells under various treatments were analyzed by flow cytometry to determine if ML216 (25 μM) treatment induces cell death. Under ML216 treatment alone, cells show robust viability (second column). Only after 1 d 6-TG (25 μM) treatment do the cells exhibit greatly increased cell death, as shown by the increased proportion of the "Dead" population (third column; red). Notably, combined ML216 and 6-TG treatment appears to mitigate cell damage and death, as indicated by the increased proportion of the "Live" population (third vs. fourth column; blue).


SPRET bias
Chr1 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr2 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr3 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr4 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr5 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Mbp
SPRET bias


SPRET bias
Chr6 5d ML216 Δ6-TG^S–R
21d ML216 Δ6-TG^S–R
10d ML216 6-TG^R only
Chr7 5d ML216 Δ6-TG^S–R
21d ML216 Δ6-TG^S–R
10d ML216 6-TG^R only
Chr8 5d ML216 Δ6-TG^S–R
21d ML216 Δ6-TG^S–R
10d ML216 6-TG^R only
Chr9 5d ML216 Δ6-TG^S–R
21d ML216 Δ6-TG^S–R
10d ML216 6-TG^R only
Chr10 5d ML216 Δ6-TG^S–R
21d ML216 Δ6-TG^S–R
10d ML216 6-TG^R only
Mbp
SPRET bias


SPRET bias
Chr11 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr12 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr13 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr14 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
Chr15 5d ML216 Δ6-TG^S-R
21d ML216 Δ6-TG^S-R
10d ML216 6-TG^R only
0 20 40 60 80 100 120
Mbp
SPRET bias

B

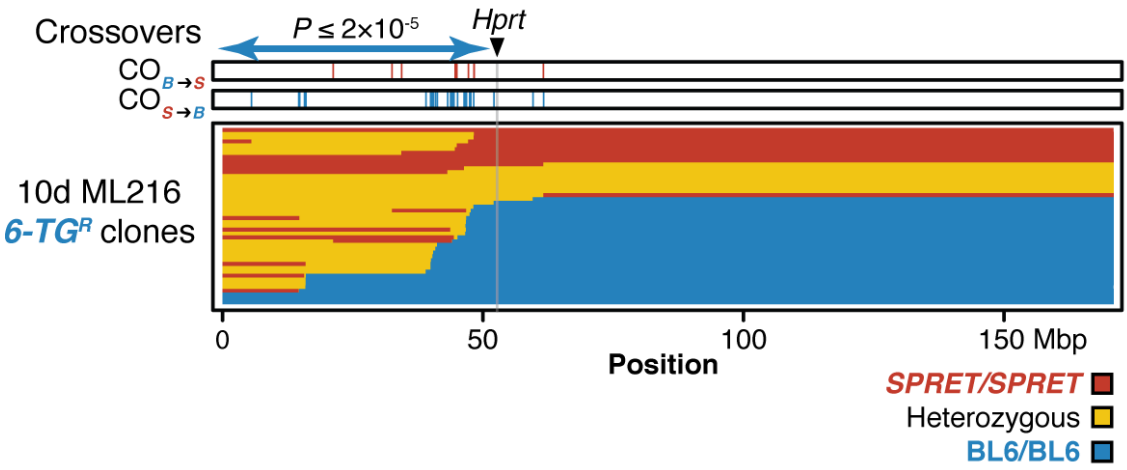


Fig. S6. Flow mapping localization of *SPRET* bias signal to a region on ChrX overlapping *Hprt*. (**A**) Raw sequencing read counts matching the BL6 (reference) or *SPRET* alleles from *6-TG^S and 6-TG^R* pools were binned into 10 kbp windows and the fraction of *SPRET* bias was calculated as the difference between *SPRET* and BL6 read counts out of the total in each window, which are themselves grouped into 1 Mbp windows. For the 5d and 21d ML216 treatments, we plotted the *differential* SPRET bias between the two pools (*6-TG^S* – *6-TG^R*). For the 10 d ML216 treatment we plotted the raw SPRET bias since only one *6-TG^R* pool was present. In both cases the *SPRET* bias estimates are represented by dots and lines showing ± 1 s.e.m. The dots and lines are coloured according to treatment (5d ML216: brown; 21d ML216: red; 10d ML216: light blue;) if they are significantly different from 0 as detailed below. Coloured dashed lines at around 0 show that the genome-wide average estimate in each treatment lie very close to no deviations between the *6-TG^S*, and the *6-TG^R* pools (and from the original 1:1 contribution in each pool in the 5d and 21d treatments, plots omitted for clarity). The outer dash lines indicate the critical values at $P \leq 0.05$ (after applying the Bonferroni correction for multiple testing) given the observed genome-wide distribution of (differential) *SPRET* bias within the sample. Note also that the data from the telomeric end of Chr14 were excluded (grey shading) due to consistent loss of the chromosome end and conversion towards *SPRET* genotypes in the S18 line regardless of ML216 or 6-TG treatment.

Both the 5d ML216 and the 21d ML216 pools show generally a very tight distribution on autosomes around 0. At significant outlier windows, shading on the excess side (*SPRET* or BL6) highlight the region, with the shading extended if nearby windows are also significant. Across all autosomes, there is a single outlier window on Chr4: 40Mbp in the 5d ML216 sample that is not matched by the other time points. However, on ChrX, there were many significant outlier windows, mostly in the 21d ML216 sample but also a few from the 5d ML216 treatment, with the main area of overlap around *Hprt* (53 Mbp, grey vertical line with arrowhead) from approximately 48 Mbp to 68 Mbp, in the direction of expected $Hprt^a$ increase activity (*SPRET* bias). In the 5d ML216 treatment the skew was less evident compared to the LOESS method shown in Fig. 3, due to the more conservative thresholds after correcting for multiple testing. Here, a single window near (but not overlapping) *Hprt* at 62 Mbp was significantly skewed in the *SPRET* direction and showed a frequency estimate consistent with its neighbouring windows, unlike the other more centromeric windows with significant skews. The significant window at 62Mbp falls within the common region across treatments. Due to the small sample size in the 10d ML216 (46 *6-TG*^{R} clones) which can cause greater fluctuations in allele frequencies, we observed no outlier windows. However, from ChrX: 45Mbp onwards the frequency plateaus towards BL6 bias (blue shading). The combined area overlaps *Hprt*.

(**B**) Individual *6-TG*^{R} clones following 10 d ML216 treatment were sequenced to determine recombination breakpoints. Crossovers of the indicated directions as in Fig. 3C are shown here to allow matching with raw data. At *Hprt*, most 6-TG surviving clones are homozygous for the $Hprt^b$ allele (27 vs. 9 heterozygotes and 10 $Hprt^a$ homozygotes).

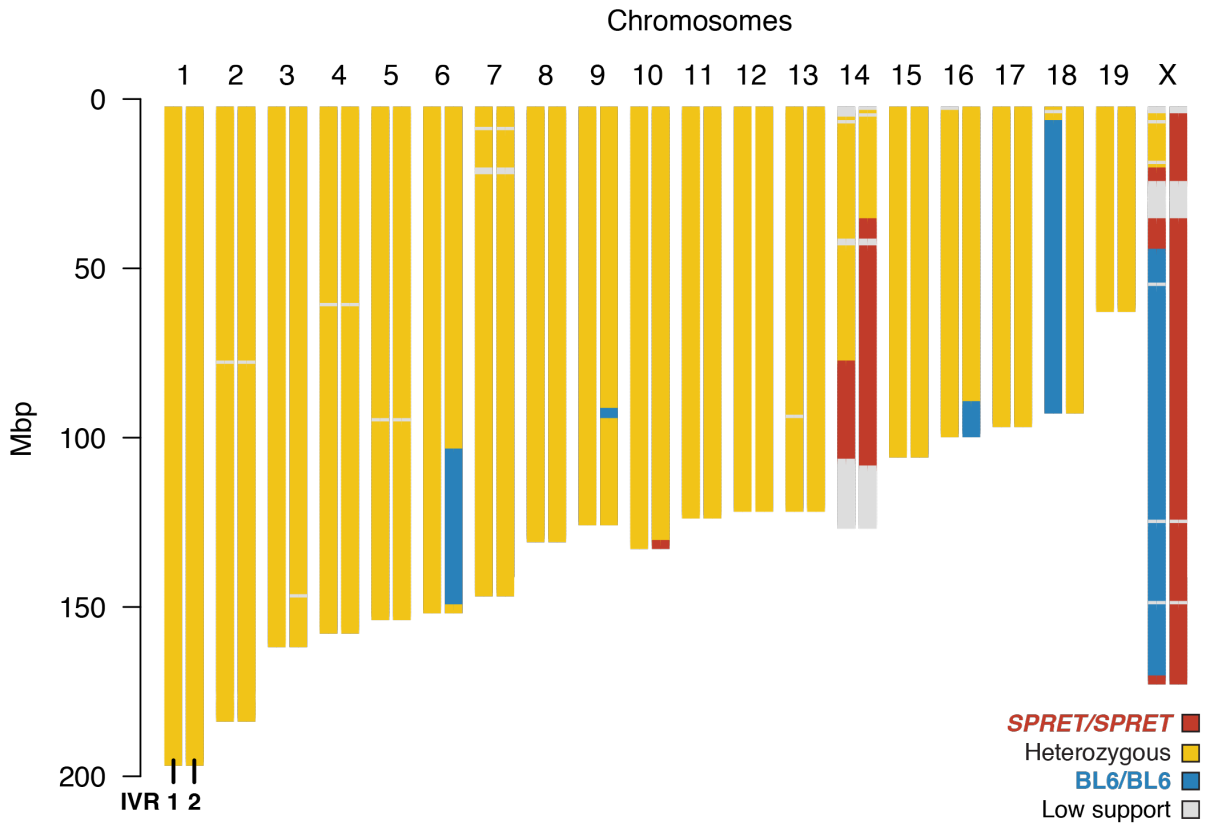


Fig. S7. Genome-wide genotype of the two S18 IVR ES cell lines selected for embryo re-derivation. High-confidence genotypes of each line for each chromosome are plotted as heterozygous (yellow) and the two BL6/BL6 (blue) and *SPRET*/*SPRET* (red) homozygous genotypes. Low-coverage or repetitive regions were considered ambiguous (grey). Both lines 1 and 2 showed substantial proportion of the genome carrying heterozygous genotypes, reflecting their F1 hybrid origin. Because these lines were obtained through *6-TG* selection, much of the observed recombinant genotypes belong to Chromosome X. In addition, we have observed chromosome instability at the distal end of Chromosome 14 (also see Fig. S4). In addition, there are major genotypic differences between IVR lines 1 and 2 on chromosomes 6, 16 and 18, as well as X. Such recombinant genotype would be difficult, if not impossible to obtain under conventional breeding. These results illustrate the potential of applying IVR at expanded scale to investigate the genetic basis of species divergence.

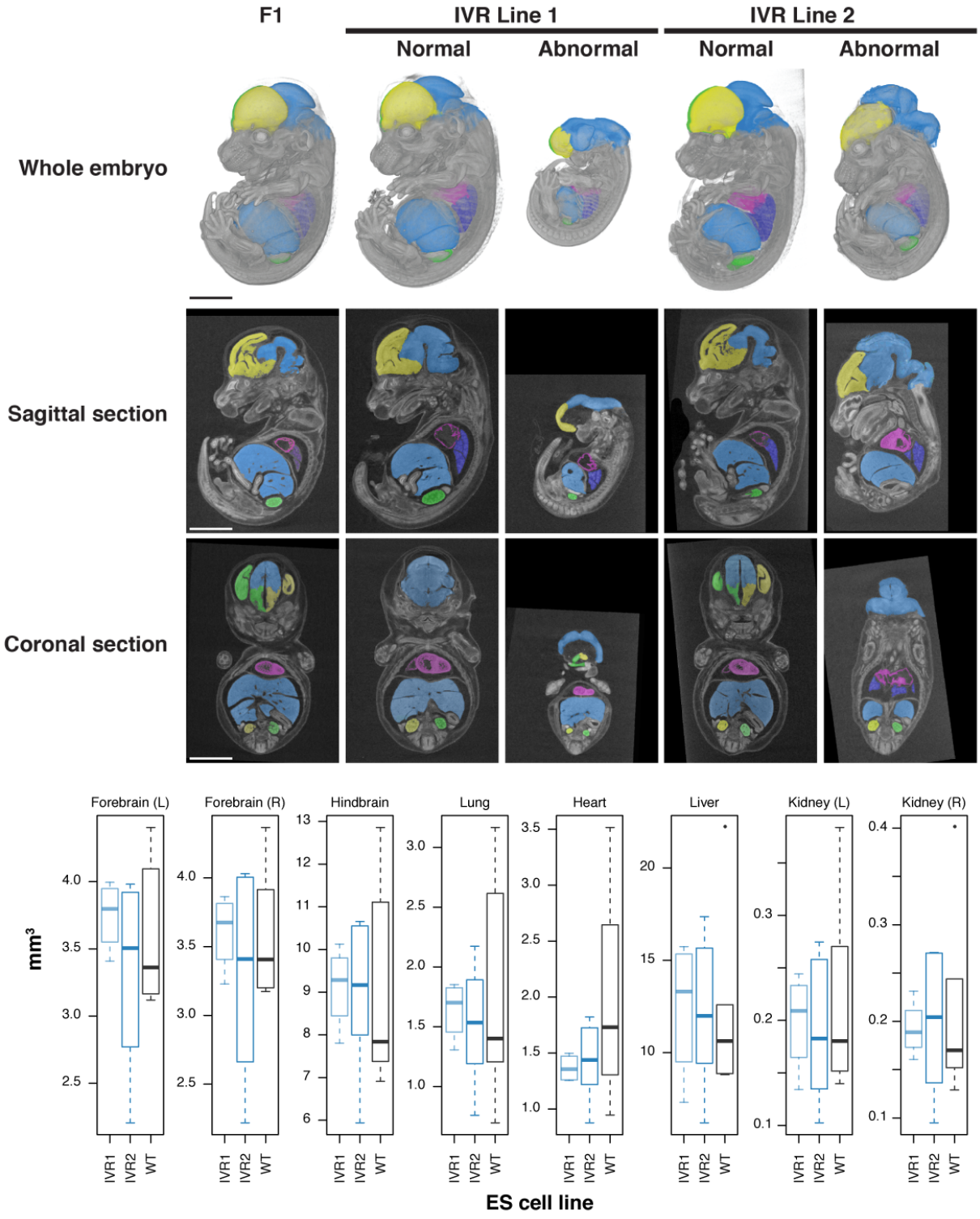

F1
IVR Line 1
IVR Line 2
Normal
Abnormal
Normal
Abnormal
Whole embryo
Sagittal section
Coronal section
Forebrain (L)
Forebrain (R)
Hindbrain
Lung
Heart
Liver
Kidney (L)
Kidney (R)
mm³
IVR1
IVR2
WT
ES cell line

Fig. S8. Whole embryos derived from F1 hybrid S18 non-recombinant and IVR ES cells. (**Top**) Embryos with almost exclusively ES cell contribution could be generated in the founder generation via laser-assisted morula injection. This allowed phenotyping of organismal traits by circumvention of hybrid sterility. Embryos were dissected in mid-gestation stage (approximately 14.5 days post-coitus, or embryonic E14.5), contrast-stained and scanned using X-ray micro-computer tomography (microCT) at 9.4 micron (μm) resolution. The use of contrast staining allowed identification and precise measurements of individual organs (colorized here for clarity). Embryos from non-recombinant S18 ES cells (left column) and two IVR ES cell lines were examined (columns 2–3 and 4–5 respectively). Representative individuals displaying normal and abnormal developmental phenotypes are shown as whole embryos with representative sagittal and coronal sections. In contrast to non-recombinant S18-derived embryos, multiple embryos from each IVR lines showed major craniofacial and neural tube closure defects. Despite a small sample size, such occurrence was highly atypical. Notably, defects in cell migration and cell–cell communication are consistent with hybrid incompatibilities. Following speciation, divergent genotype combinations carried by the same individuals have not been subjected to selection for compatible functions. Consequently, hybrid incompatibilities often result in developmental defects. Derivation of embryos from panels of IVR ES cell lines may allow genetic dissection of developmental variation arising from evolutionary divergences. (**Bottom**) Volume measurements from individual organs were obtained and their distributions shown here, after excluding embryos with defects. There are no statistically significant differences between embryos derived from different ES cell lines. With increased sample size certain organs, such as heart, lung and liver, may be good candidates to screen for consistent trait differences between lines.

3.2.14 Supplementary Tables:

Table S1. Oligonucleotide primers for multiplexed genotyping of sub-telomeric markers. Each pair of primers carry an extension (underlined) to allow easy attachment of a third, universal fluorophore-conjugated primer for fragment analysis in a capillary sequencer as described in (10).

	Position				Amplicon length, bp		
Chr	Start	End	Primer	Sequence, 5′–3′	*BL6*	*CAST*	*SPRET*
1	191,536,961	191,537,264	Forward_M13F	TGTAAAACGACGGCCAGTGCACCAGACCCGCTTAGTGTGTGT	321	296	300
			Reverse	CCTGTGGATGTAGCTCTCAGC			
2	173,989,251	173,989,620	Forward	AGTCCTGGGCTGAAGGAGAT	387	359	359
			Reverse_M13R	GCGGATAACAATTTCACACAGGACTCATTCTCAGTTCTGGCAGACCA			
3	153,694,525	153,694,762	Forward_M13F	TGTAAAACGACGGCCAGTGGCATTTGGACAATCCTGTGTGATG	255	236	231
			Reverse	TGTGACAGGTCTGTTCCGTG			
4	154,084,110	154,084,446	Forward	ACTTGGGTGGAAGCCTTGTC	355	331	331
			Reverse_M13R	GCGGATAACAATTTCACACAGGCTCCTCTCTGTCTTTGCCGGTT			
5	148,049,116	148,049,460	Forward	CCTTTCTGACTACTTTTGGATCAGTCC	361	338	338
			Reverse_M13R	GCGGATAACAATTTCACACAGGGATCCAGAAAAAACAAGTGATCAGACAAGTAG			
6	146,150,310	146,150,633	Forward_M13F	TGTAAAACGACGGCCAGTGCCAAACGTGGTAGTGGAAGCAGG	339	311	310
			Reverse	TGGCCCTTTAGAGTCAGGAC			
7	136,216,727	136,217,074	Forward_M13F	TGTAAAACGACGGCCAGTGCACAAGGAAGGGTGTTGCAGGATG	366	335	335
			Reverse	CTGTCAGAAGCCAGGGAAGG			
8	127,441,747	127,442,124	Forward	GCTTTCATCAGCAGTTAGAGCAG	394	369	369
			Reverse_M13R	GCGGATAACAATTTCACACAGGGATTCCACCATGCACTCTACTTTC			
9	121,912,729	121,913,029	Forward	TTGGTCACGAATCCTCCTGC	317	291	291
			Reverse_M13R	GCGGATAACAATTTCACACAGGCACTCCTTGGCTCTGGTGGT			
10	125,272,539	125,272,846	Forward	GTTTCTGCTTATGACCACCCAA	324	300	300
			Reverse_M13R	GCGGATAACAATTTCACACAGGTCAAGACCTATAGTCTCTCTCAGTGTCTTAT			
11	119,517,460	119,517,790	Forward	GCACTAGGGACTTGGTCACC	350	320	320
			Reverse_M13R	GCGGATAACAATTTCACACAGGTTGCTAGGCTGCCTTCATTAGCT			
12	117,814,241	117,814,555	Forward	CCATTATGGAACTTTCCTAAGAGTGGCA	332	304	304
			Reverse_M13R	GCGGATAACAATTTCACACAGGGGCAGGAAGGGCAGAAGTTT			
13	117,015,681	117,016,015	Forward_M13F	TGTAAAACGACGGCCAGTGCACTGCATGGTTCCATCTGTGCT	351	324	324
			Reverse	GGAACTGCTCTTGTCACCAGTAT			
14	106,293,973	106,294,368	Forward_M13F	TGTAAAACGACGGCCAGTGCCTGGTGGCAACCAACATCTGTCTACATGG	412	388	386
			Reverse	GTCATCATGTCTGAAACATTCCCTGAGCCTTC			
15	94,806,342	94,806,639	Forward	GGCCCTTTCTTCTTCGGGAA	315	288	288
			Reverse_M13R	GCGGATAACAATTTCACACAGGTGTCCCTTAGAAGTGTCCCCTC			
16	93,940,685	93,941,051	Forward_M13F	TGTAAAACGACGGCCAGTGGCGCAGACATATCAAGAGTAGCGTTGTGAAGGCTGG	383	358	358
			Reverse	CGAATGCCACACTCCTTCCTGG			
17	84,166,698	84,167,073	Forward_M13F	TGTAAAACGACGGCCAGTGATACATTTAACCCGTCTCTACGGGGAGC	392	363	363
			Reverse	GCCTTACTTACACCTGAGGCCTCC			
18	82,968,745	82,969,071	Forward_M13F	TGTAAAACGACGGCCAGTGGTACAGTGTTAGTGTGCTCATATCAGC	346	319	319
			Reverse	CTGTGCTCACTGCCTCAAGA			
19	56,856,486	56,856,918	Forward_M13F	TGTAAAACGACGGCCAGTGGTCCATCACTTGCAGAGCCTGCC	449	424	425
			Reverse	GTTTCAGCAGAACAAGGCAGCAACAAGCATG			
			M13F_FAM	5′FAM-TGTAAAACGACGGCCAGTG			
			M13R-46_HEX	5′HEX-TGCGGATAACAATTTCACACAGG			
X	79,444,414	79,444,518	Forward	AGGCTCTACACAGAGTACAG	105	105	79
			Reverse	CAGTAGTGGTGCTTGAGTTCTA			
			BL6 Probe	/56-FAM/ATGCCTGTG/ZEN/GAATGGTCTCCTCTTG/3IABkFQ/			
			SPRET Probe	/56-FAM/ACCCTTTCT/ZEN/GGCTGCCTGTG/3IABkFQ/			
14	76,235,178	76,235,282	Forward	GAAGTGGAAGAAGCACACTAGG	105	73*	73
			Reverse	TGTTACACACAGTGAGAAAGGAC			
			BL6 Probe	/56-FAM/TTAGGTCCC/ZEN/GCTGCACACTGTC/3IABkFQ/			
			SPRET Probe	/56-FAM/TCTAGGAGC/ZEN/AGTAACATAACCAGCTGC/3IABkFQ/			

Each pair of primers carry an extension (underlined) to allow easy attachment of a third, universal fluorophore-conjugated primer for fragment analysis in a capillary sequencer as described in ref. 1. Chr, chromosome.
*Taqman genotyping assay not applicable for CAST due to SNPs located within the annealing sequence for either Taqman probe.

1. Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol* 18:233–234.