Sex-specific phenotypic effects and evolutionary history of an ancient deletion polymorphism of the human growth hormone receptor

Authors Resendez, S.[1γ], Saitou, M.[1γ], Parisi, L.R.[2], Wu, F.[3], Nakagome, S.[4], Satta, Y.[5], Atilla-Gokcumen, G.E.[2], Mu, X.[3]*, Gokcumen, O.[1]*

[γ] Equal contribution
*Corresponding authors

Correspondence:
O.G.: omergokc@buffalo.edu and X.M.: xmu@buffalo.edu

Affiliations:

1. Department of Biological Sciences, University at Buffalo, Buffalo, NY

2. Department of Chemistry, University at Buffalo, Buffalo, NY

3. Department of Ophthalmology/Ross Eye Institute, Jacobs School of Medicine and Biological Sciences, University at Buffalo, Buffalo, NY

4. Trinity College Dublin, School of Medicine, Dublin, Ireland

5. SOKENDAI (Graduate University for Advanced Studies), Department of Evolutionary Studies of Biosystems, Kanagawa Prefecture, Japan

Running title: Evolutionary and functional impact of *GHRd3*

Keywords: Copy number variation, adaptation, human evolution, metabolism, development, mTOR pathway, Neanderthal, Denisovan

ABSTRACT

The deletion of the third exon of the growth hormone receptor (*GHRd3*) is one of the most common genomic structural variants in the human genome. This deletion has been linked to response to growth hormone, placenta size, birth weight, growth after birth, time of menarche, adult height, and longevity. However, its evolutionary history and the exact mechanisms through which it affects phenotypes remain unresolved. While the analysis of thousands of genomes suggests that this deletion was nearly fixed in the ancestral population of anatomically modern humans and Neanderthals, it underwent a paradoxical adaptive reduction in frequency approximately 30 thousand years ago, a demographic signature that roughly corresponds with the emergence of multiple modern human behaviors and a concurrent population expansion. Using a mouse line engineered to contain the deletion, pleiotropic and sex-specific effects on organismal growth, the expression levels of hundreds of genes, and serum lipid composition were documented, potentially involving the nutrient-dependent mTORC1 pathway. These growth and metabolic effects are consistent with a model in which the allele frequency of *GHRd3* varies throughout human evolution as a response to fluctuations in resource availability. The last distinctive prehistoric shift in allele frequency might be related to newly developed technological buffers against the effects of oscillating resource levels.

The third coding exon of the growth hormone receptor gene (*GHR*) is deleted in approximately half of all human genomes, as well as all sequenced Neanderthal and Denisovan genomes. This is surprising given that this receptor is highly conserved among mammals and fundamental in many cellular processes, including cell division, immunity, and metabolism, among others (*1*) and given that a loss of function mutation of this gene causes Laron syndrome in its homozygous form (*2*, *3*). Indeed, locus-specific studies have consistently and reproducibly associated *GHRd3* with altered placental and birth weight (*4*), time of puberty onset (*5*), lifespan (*6*), metabolic activity (*7*), and response to growth hormone treatments (*8*). Despite the known relevance of *GHRd3* to human phenotypes that are likely crucial for human evolution, the mechanisms through which *GHRd3* affects cellular and organismal function and the evolutionary forces that maintain *GHRd3* in human populations remain largely unknown.

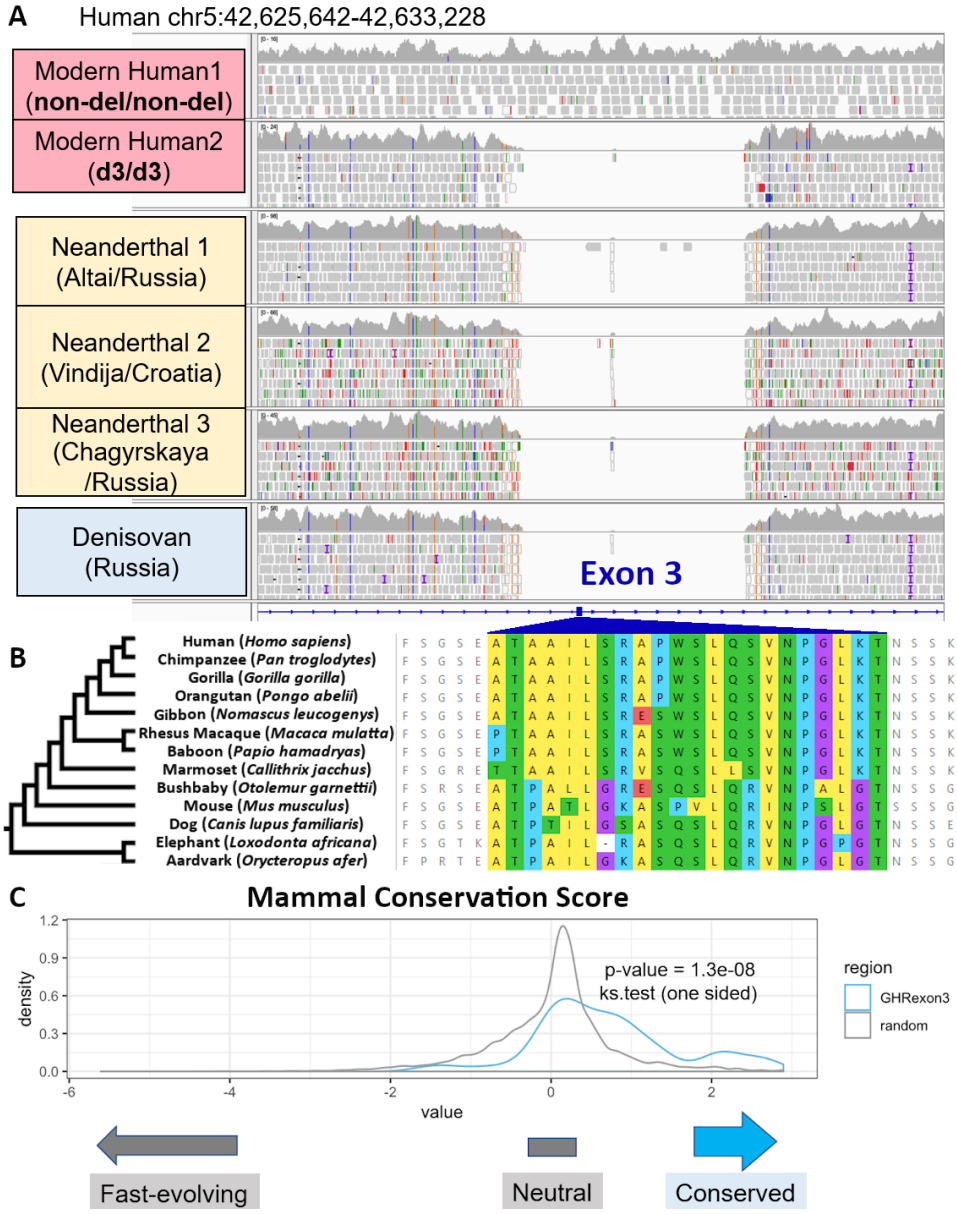


Figure 1. A. *GHRd3* in modern and ancient hominin genomes. These browser snapshots show the genome assembly (Hg19) of a human with the ancestral, homozygous non-deleted genotype and another with a homozygous deleted genotype that shows no reads mapping to the deletion region (top two rows). Similarly, sequences from 3 Neanderthal and 1 Denisovan assemblies were mapped to this region and show a clear signature of the deletion with breakpoints indistinguishable from the deletion observed in modern humans. **B.** The coding sequence alignment of exon 3 among mammals, which indicates near-complete preservation of the amino-acid sequence in primates and clear conservation across mammals. **C.** A conservation score comparison of *GHR* exon 3 sequences with 20,000 randomly chosen sites from chromosome 5.

GHRd3 generates a unique isoform of a key developmental gene and has a complex evolutionary history in humans. Our previous work determined that *GHRd3* is one of 17 exonic, polymorphic human deletions shared with the Altai Neanderthal and Denisovan genomes (*9*). This study now extends this observation to include the Vindija (*10*) and Chagyrskaya Neanderthal genomes (**Fig. 1A**). This suggests that *GHRd3* evolved prior to the divergence of the human and Neanderthal lineages and remains polymorphic in humans, but may have been fixed in the Neanderthal and Denisovan lineages since none of their sequenced genomes contain the non-deleted variant. The deleted coding sequence is conserved among primates (**Fig. 1B**). On a broader scale, exon 3 is significantly more conserved when compared to random sites on the same chromosome ($p = 1.3x10^{-8}$, Kolmogorov–Smirnov test), suggesting that this exon is not evolving under neutral conditions. (**Fig. 1C**). When these data are observed in conjunction with the existing literature associating *GHRd3* with multiple human traits, it suggests that adaptive forces may have shaped the allele frequency of this variant in modern human populations.

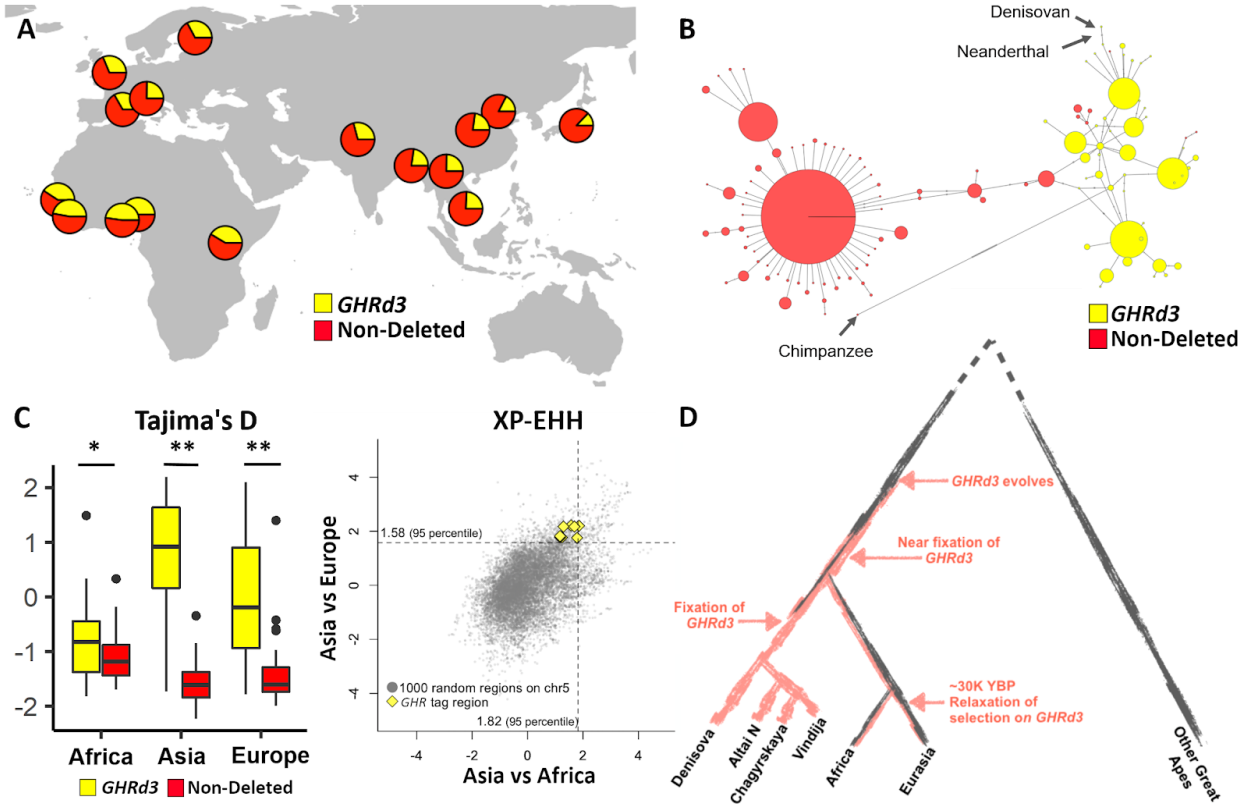


Figure 2. A. The geographic distribution of the *GHRd3* polymorphism. The deletion status is color-coded for the entire figure. **B.** A network of 2,504 human haplotypes, the Altai Neanderthal genome, the Denisovan genome, and the chimpanzee reference genome, calculated from variations that are in high LD upstream of the *GHRd3* location (Hg19: chr5: 42624748-42628325, see **Fig. S1**). **C.** Tajima's D and XP-EHH values calculated for the *GHRd3* upstream region. Tajima's D values are calculated for the deleted and non-deletion alleles in three 1000 Genomes meta-populations; XP-EHH values are calculated for the *GHRd3* tag SNPs ($r^2 > 0.8$) and compared to distributions calculated for 1000 similarly sized regions (shown in grey). **D.** A plausible model for the evolution of *GHRd3.*

To understand the evolutionary forces shaping the maintenance and distribution of *GHRd3* within human populations, the linkage disequilibrium (LD) architecture around *GHRd3* was resolved. The deletion is in strong LD (**$r^2 > 0.75$**) with a small number of nearby single nucleotide

variants. These variants help construct a short haplotype block that is flanked by known recombination spots (Hg19, chr5: 42625443-42628325, **Fig. S1**). This specific haplotype tags the *GHRd3* allele. Based on these tag-SNPs, as well as direct genotyping made possible through the 1,000 Genomes dataset, it was found that *GHRd3* is the major allele in most African populations, but the allele frequency decreases to less than 25% in the majority of Eurasian populations (**Fig. 2A, Fig. S2**). The resolution of this upstream haplotype block allowed for the construction of a network tree where one can visualize the clear separation of haplotypes that carry the *GHRd3* allele and those that do not (**Fig. 2B).** One unexpected observation is that the haplotypes with the deletion are more variable and more similar to the chimpanzee haplotype than those without the deletion (**Fig. S3**). This raises the possibility that *GHRd3* was almost fixed in the human lineage, meaning the haplotypes that harbor the ancestral, non-deleted allele rebounded in frequency only recently.

To test if selection is acting on this locus, the single nucleotide variation data within the *GHRd3* haplotype block was used to calculate Tajima's D (*11*) and XP-EHH (*12*) values. These statistics measure deviations from expected allele frequency spectra and extended homozygosity differences between populations, respectively. A significantly lower Tajima's D value was observed in the East Asian population as compared to randomly selected regions on chromosome 5 (**$p < 0.05$, MW test**). Moreover, the non-deleted haplotypes showed lower Tajima's D values when compared to the deleted haplotypes in the Eurasian population (**$p < 0.01$, MW test, Fig. 2C, Fig. S4**). Concordant with the Tajima's D analysis, when the *GHRd3* locus of the Han Chinese (CHB) population was compared with that of the West European (CEU) and Yoruban (YRI) populations, XP-EHH values that were higher than genome-wide expectations were noted, potentially suggesting a sweep of the existing non-deleted haplotypes in the East Asian population (**Fig. 2C**).

To summarize these results, a plausible model can be constructed where *GHRd3* was nearly fixed in the human-Neanderthal lineage and remained the major allele throughout most of human history, but only recently reduced in allele frequency due to relaxation of the selection favoring *GHRd3* (**Fig. 2D)**. This would be observed as a sweep on the existing haplotypes harboring the non-deleted allele. A method that first combines site frequency spectrum analyses with LD calculations and then compares the results with simulated expectations (*13*) was used to further investigate this model. There was a significant deviation from the simulated neutral expectations in the East Asian population, showing selection starting approximately 27 thousand years ago (**$p < 0.01$, Supplementary Material**). To strengthen this analysis, a recent approximate Bayesian computation method (*14*) that combines the site frequency spectrum and haplotype structure of a locus was utilized. This method then compares these data with different models (**Fig. S5**, **Supplementary Material**) and found similar results to the first method. It suggested a recent sweep of the non-deleted haplotypes in East Asia, dating the onset of this sweep to 29,811 years ago, while putting the frequency of the non-deleted allele at ~11% before the sweep ($\log_{10}$-scaled approximated Bayes factor, $\log_{10}(aBF) = 5.250$ against neutral; $\log_{10}(aBF) = 13.257$ against hard-sweep) (**Fig. S6**). Collectively, the evolutionary analyses of this locus renders neutral evolution a less likely explanation for the high allele frequencies of *GHRd3* and instead points to population-specific adaptive forces that likely vary across time.

As described earlier, *GHRd3* has been associated with multiple human traits. However, these associations were found in locus and phenotype-specific studies, which are often prone to type I errors. Moreover, genome-wide association studies (GWAS) have rarely found associations with

GHRd3 and any other trait. It is likely that the inability of GWAS studies to pick up signals from this deletion is due to limitations in SNP-array platforms, which rely on imputation to genotype *GHRd3*. To remedy these issues, the UK Biobank dataset phenotypes (*15*, *16*) were examined for associations with a single nucleotide variant that tags the deletion. It was found that the *GHRd3* haplotype (represented by rs4073476) is strongly associated with standing height in this cohort (**nominal $p<10^{-8}$, in a PheWAS of 742 haplotypes**) (**Fig. S7**). This replicated findings from a previous, locus-specific study that found an association with height in a different cohort (*6*). That same study also suggested sex-specificity of the effect of *GHRd3* on longevity, so the UK Biobank dataset was examined further to test if *GHRd3* is correlated with any phenotypic effects in a sex-specific manner. The top 10 traits that are at least nominally associated with *GHRd3* were investigated and it was found that grip strength has a greater correlation with the deletion in males (p-values - left: 1.29e-05; right: 1.40e-04) than in females (p-values - left: 3.57e-01; right: 1.48e-01). This effect was noted independently for both the left and right hands. The sex-specific phenotypic effects of *GHRd3* in human cohorts remain an important area of future research.

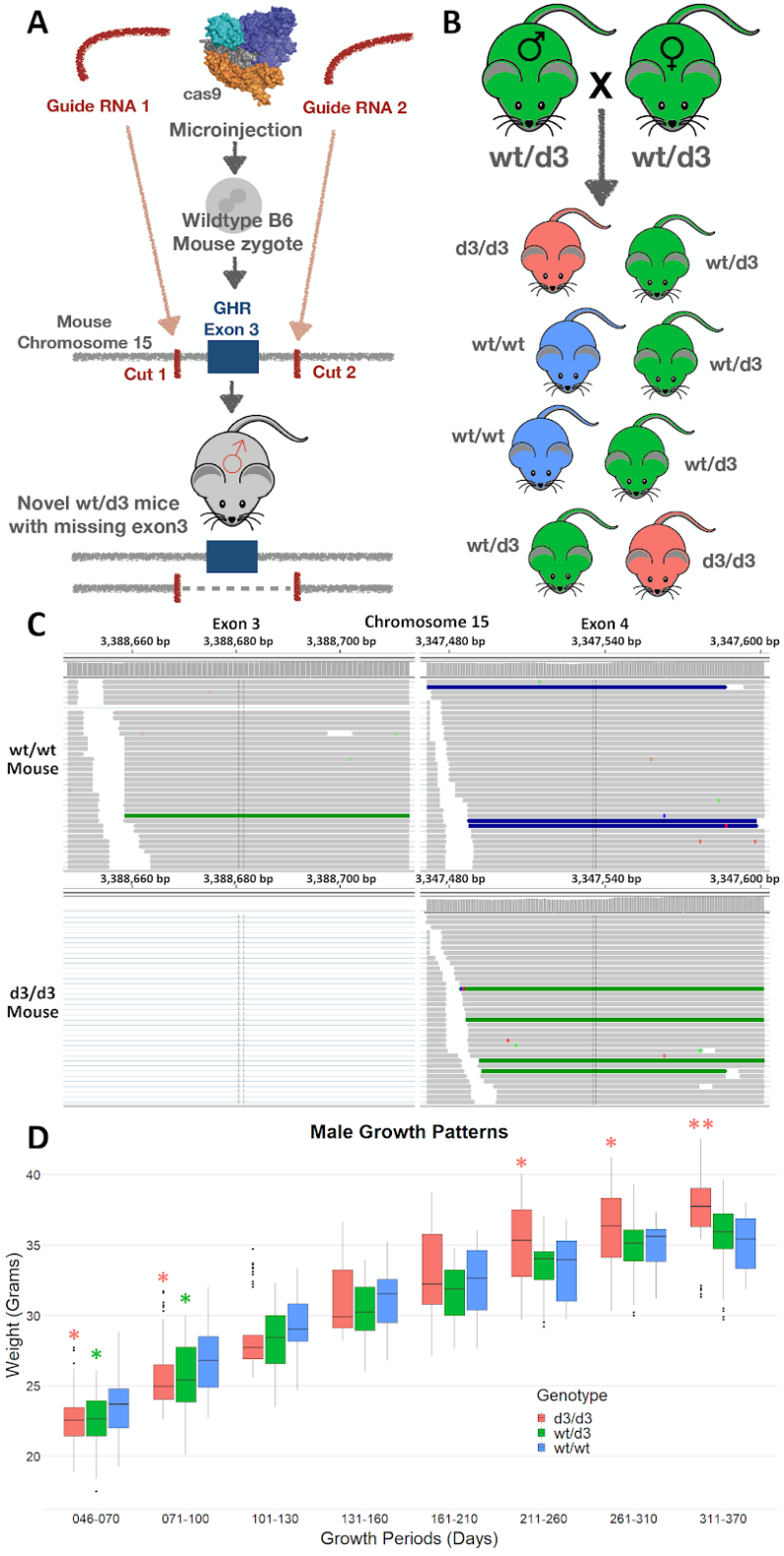


Figure 3. A. The design of the mouse model using a CRISPR/Cas9 approach with sgRNAs targeting each side of the deletion. **B.** Experimental breeding design used for weight measurements, transcriptomics, and lipidomics analyses. Parental bias was limited by breeding only heterozygous parents. **C.** RNAseq exonic read-depth from a wt/wt and d3/d3 mouse for exons 3 and 4. Each horizontal grey line represents a

single read. The d3/d3 mouse expresses *GHR* exon 4 in similar levels as compared to the wt/wt mouse, but does not express exon 3. **D.** The weight measurements for males across a year period. Periods with significant (**p<0.05, MW test**) differences between d3/d3 or wt/d3 and wt/wt mice are indicated by a single asterisk; p-values below 0.01 are indicated with two asterisks.

To further investigate the biological effects of *GHRd3,* a mouse model was developed by generating a ~2.5kb deletion, mimicking the human variant and removing the otherwise conserved exon 3 ortholog from the C57BL/6 genome using CRISPR/Cas9. This procedure involved editing the genome of mouse embryos using custom-designed single guide RNAs (sgRNAs) that complement the sequences flanking the third exon of the growth hormone receptor gene (*Ghr*) in the mouse genome (**Fig. 3A, Supplementary Methods**). The resulting founder mouse was a heterozygous (wt/d3) male, which was then used to generate offspring with *Ghrd3*. No obvious phenotypic effects or deviations from Mendelian expectations were observed in the colony. The size (gross weight) of 41 individuals with heterozygous (wt/d3) parents (**Fig. 3B**) was then measured over the first year of their lives (**Table S1**). We confirmed the mouse genotypes using both gel electrophoresis and expression analyses of exon 3 in the *Ghrd3* mice **(Fig. 3C, Supplementary Methods)**. When comparing the different genotypes, a significant difference in the trajectory of weights was found in males, but not in females. Specifically, it was observed that during early development, male d3/d3 mice were significantly lighter, but became significantly heavier than their wt/wt relatives in late adulthood (**Fig. 3D**, **p<0.05, MW test)**. Such differences were not observed in females (**Fig. S8**). These findings provide experimental evidence that supports human association studies which showed that placenta size and size at birth are negatively correlated with the presence of *GHRd3 (4).*

To understand the mechanistic underpinnings of the biological effects of *GHRd3*, a comparative transcriptomics analysis of 28 to 34-day-old **(Supplementary Methods)** liver tissues from wt/wt and d3/d3 mice was conducted. At this age range, the mice have reached a point of sexual maturity, but are still in a period of rapid growth and development (*17*). Since growth hormone secretion patterns differ based on sex, one might expect to observe sex-specific effects on the transcriptome caused by *Ghrd3.* To account for this, the transcriptome data were compared independently in 13 female (7 wt/wt and 6 d3/d3) and 13 male mice (6 wt/wt and 7 d3/d3). The differences observed were female-specific, contrasting with the male-specific growth effects (**Fig. S9**, **Table S2**). In fact, no statistical differences in gene expression were found when comparing male wt/wt and d3/d3 livers, except for a single Y-chromosome, non-coding transcript, Mid1-ps1 (p_{adj}<0.05, Wald test). In contrast, the female analysis revealed 474 genes that show significantly altered expression levels (p_{adj}<0.05) (**Fig. 4A, Table S2**). This stark difference in the impact of *Ghrd3* on transcriptomes is likely caused by an interaction between this variant and the overall, sex and age-specific hormonal environment.

The sex-specific effects of *Ghrd3* were observed independently in two replicate experiments involving mice from different litters, granting increased confidence in these results (**Table S2**). Specifically, these results are compatible with a scenario in which the effect of *Ghrd3* is dependent on the sex-specific release patterns of growth hormone. It is well documented that growth hormone is secreted in a pulsatile fashion in males, while its secretion is more continuous in females (*18*). Thus, one would expect to observe the effect of *Ghrd3* more clearly when comparing individuals with sustained secretion levels of growth hormone, as is the case with the female mice. In contrast, the variable growth hormone secretion patterns observed in male mice would introduce additional noise. As expected, dramatically increased standard error

values of transcript fold-changes were found in male mice as compared to female mice (**Fig. S10**), especially for the genes with significantly altered expression levels in the female livers (**Fig. S11**). This explains why the effect of *Ghrd3* was not observable in the male transcriptomes. These findings illustrate the complex hormone-gene interactions and sex-specificity involved with Ghr activity and show the value and need for future studies involving additional developmental stages and organ tissues.

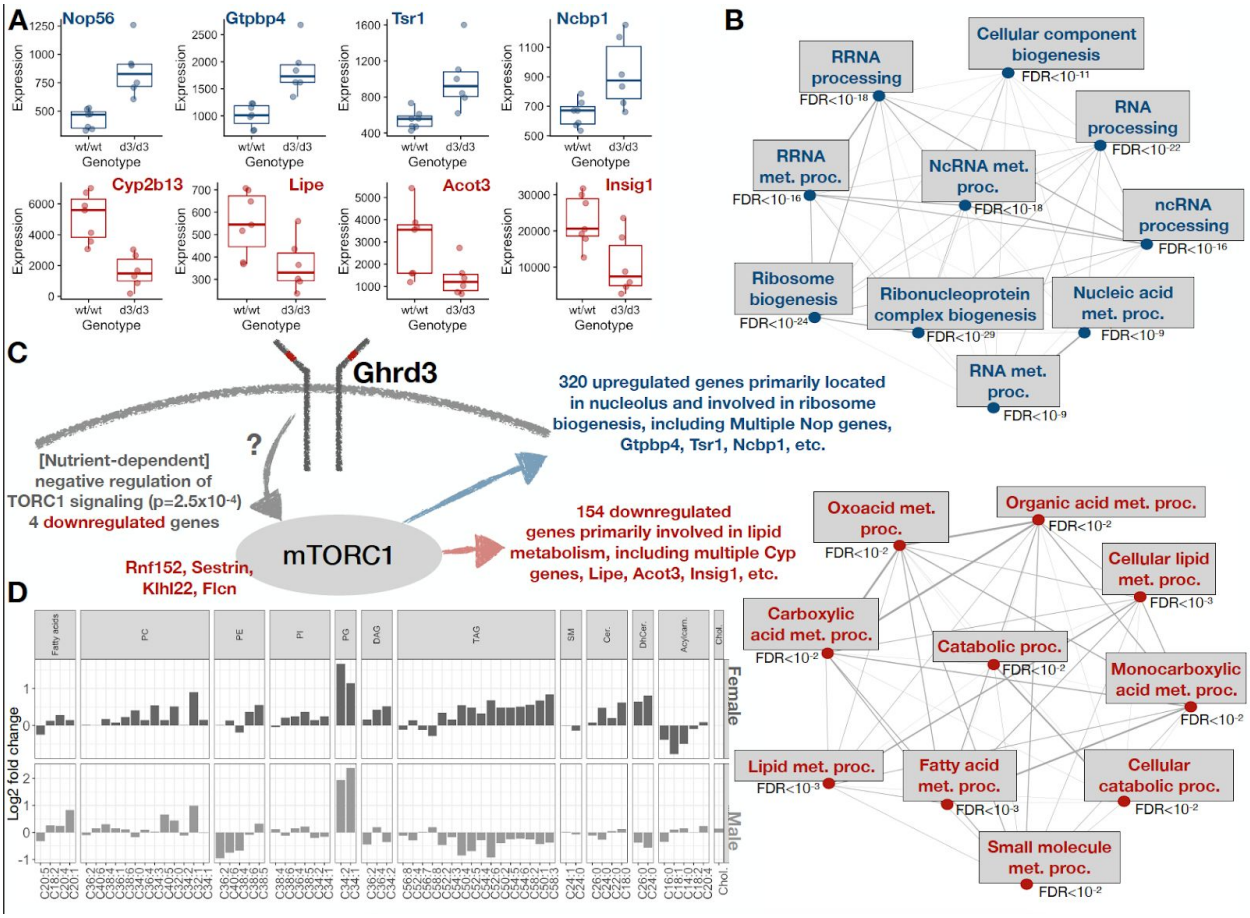


Figure 4. A. A sample of genes with significant expression differences between wt/wt and d3/d3 females. **B.** Top ten biological process GO enrichment categories for upregulated (blue) and downregulated (red) genes. **C.** A general model for the pleiotropic impact of *Ghrd3.* **D.** Log_2 fold changes between wt/wt and d3/d3 female mice (top) and wt/wt and d3/d3 male mice (bottom), showing sex-specific impacts of *Ghrd3* on lipid composition of blood serum in females and males. FDR=False Discovery Rate, PC=phosphatidylcholine, PE=phosphatidylethanolamine, PI=phosphatidylinositol, PG=phosphatidylglycerol, DAG=diacylglycerol, TAG=triacylglycerol, SM=sphingomyelin, Cer.=ceramide, DhCer.=dihydroceramide, Acylcarn.=acylcarnitine, Chol.=cholesterol.

A gene ontology enrichment analysis was then performed to identify the biological impacts of the observed expression differences. Among the genes that are upregulated in d3/d3 female livers, there were significant enrichments of genes related to rRNA processing, ribosomal function and biogenesis, and ribosomal localization (**Fig. 4B, Table S3**). For downregulated genes, there were significant enrichments of genes that affect mTORC1 signaling, organelle organization, protein localization, and lipid catabolic processes, particularly concerning fatty acid metabolism. Further analysis showed that the downregulated genes categorized under mTORC1 signaling (i.e. *Rnf152*, *Sestrin*, *Klhl22*, and *Flcn*) are all involved in the nutrient-dependent regulation of mTORC1 (*19*). Multiple downstream targets of mTORC1, such

as *Eif2b3*, *Insig1*, *Plin5*, and *Plin4*, were differentially expressed between *Ghrd3* and wildtype mice. These expression trends suggest a hormone-dependent and sex-specific interaction between *Ghrd3* and the mTORC1 pathway in developing mice, with broad consequences on cellular metabolism and growth (**Fig. 4C**).

To further evaluate the downstream effects of *Ghrd3*, the blood serum lipid composition of 12 homozygous mice (3 from each of the following categories: wt/wt female, d3/d3 female, wt/wt male, d3/d3 male) was analyzed using liquid chromatography-mass spectrometry. Several lipid-species showed consistent negative or positive correlations with the presence of the deletion in both sexes **(Fig. 4D, Table S4)**. For example, two phosphatidylglycerol species detected in the serum had approximately a 4-fold greater abundance in the d3/d3 mice as compared to their wildtype counterparts. The sex-specific trends of the triglycerol species were even more interesting. For females, it was found that 13 of the 16 triglycerols analyzed had a greater abundance, whereas 14 had a lesser abundance in the males. Together with the expression data, the lipid composition analysis suggests a broad, sex-specific influence of *Ghrd3* on the metabolism of growing mice.

In this study, an evolutionary genetics analysis was integrated with a model organism study to observe the formation, maintenance, and functional effects of a common human deletion variant. Both empirical and simulation-based population genetic analyses suggest the presence of non-neutral evolutionary forces maintaining *GHRd3* in the human lineage since before human-Neanderthal divergence. The reanalysis of genome-wide association study datasets allowed for the replication of a previous association of *GHRd3* with adult standing-height. A novel mouse model of the *GHR* exon 3 deletion was also created through CRISPR/Cas9. Using this model, sex-specific effects of *Ghrd3* were observed at the organismal, transcriptomic, and lipidomic levels. In accordance with observations made in humans, *Ghrd3* led to smaller sizes in early life and faster weight-gain in later stages of development in males. Even earlier in life, it was found that 28 to 34-day-old *Ghrd3* mice exhibit sex-specific impacts on their liver transcriptome. Particularly, this deletion affected ribosome biogenesis and lipid metabolism in females, likely through its effect on the nutrient-dependent mTORC1 regulation pathway. These results support the notion that the effects of *Ghrd3* are pleiotropic and specific to the hormonal environment, which will differ between developmental stages, sexes, and organs.

We hypothesize that the effects of *GHRd3* have important ramifications for metabolic disorders, such as obesity and diabetes, but only within particular environmental contexts. For example, it was reported that *GHRd3* has a significant preventative impact on type 2 diabetes, but in the small number of diabetes patients who are homozygous for *GHRd3*, a significant increase in metabolic disorders was observed (*20*). This mouse model will make it possible to test this hypothesis in future studies. We expect the *Ghrd3* mice to display altered phenotypic and transcriptomic responses to low or high-calorie diets as compared to wild type mice.

Based on these findings, we hypothesize that growth hormone receptor variations may affect fitness within the context of the famine and abundance periods that had been a defining feature of human evolution. It was previously speculated that the *GHRd3* mediated upregulation of growth hormone pathway activity may be a response to nutritional deprivation (*21*). As compared to nonhuman great apes, humans generally cope with higher levels of seasonality and unpredictability of resource levels (*22*). This is due primarily to the transition of hominids from evergreen rainforests to savannas, which are affected more by seasonality. In this

ecological context, the fluctuation of resources may have conferred a major evolutionary stressor for humans, especially on traits pertaining to metabolism and reproduction. Thus, it is plausible that *GHRd3* has been maintained due to it granting a fitness advantage related to its effect on birth size, post-birth development, time-to-puberty, and metabolic activity. This scenario also fits with the observation that *GHRd3* is potentially fixed in ancient Eurasian hominins (i.e. Neanderthals and Denisovans), who were exposed to even greater seasonal fluctuations than African hominins. Furthermore, the relaxation of selective pressures and the marked reduction in *GHRd3* allele frequencies among Eurasian populations, which this study dated to approximately 30 thousand years before present, coincides with the emergence of multiple modern human behaviors and a concurrent population expansion (*23*, *24*). This prehistoric demographic signature could be explained by novel technological buffers against the effects of fluctuating resource levels.

Data Availability: All the fastq files from the RNAseq experiments are being uploaded to NCBI sequence read archive (https://www.ncbi.nlm.nih.gov/sra). All other data are available in the Supplementary Materials and Tables.

Acknowledgments: Funding**:** This work was financially supported by the National Science Foundation (no. 1714867). Work in the Mu lab was also supported by grants from the BrightFocus Foundation (G2016024) and the National Eye Institute (EY020545). S.R. received funding during his first two years of research from the Collaborative Learning and Integrated Mentoring in the Biosciences (CLIMB) program. Author contributions: O.G. and X.M. designed the experiments and oversaw the study, S.R. conducted the majority of the mouse work, genotyping, and transcriptomics analysis, M.S. conducted transcriptomics and population genetics analysis, L.R. and G.E.A-G conducted lipidomics analysis, S.N. and Y.S. conducted simulation-based selection analysis. F.M. helped in tissue collection in mice. O.G., S.R., and M.S. wrote the manuscript and prepared the figures. Competing interests: The authors declare no competing interests.

We are grateful to the Roswell Park Cancer Institute Gene Targeting and Transgenic Resource, especially Aiemee Stablewski, for their help in establishing the *Ghrd3* mouse line. We are grateful for the help of George (PJ) Perry, and Audrey Arner regarding the navigation of the UK Biobank dataset. We would like to acknowledge Kirsten Dean and Victoria Gellatly for their help in genotyping. We thank our departmental colleagues Drs. Xu-Friedman, Taylor, Lynch, and Albert for their constructive criticism throughout this study.

SUPPLEMENTARY MATERIALS

Supplementary figures

Figure S1. Linkage disequilibrium between SNPs near *GHRd3* and the deletion. This figure shows R^2 values calculated for the *GHR* deletion and its neighboring SNPs (black dots, left scale), along with recombination rates for the locus (the pink line, right scale). Recombination rates were retrieved from the 1000 genome selection browser (http://hsb.upf.edu/). Based on the linkage disequilibrium, we selected Hg19: chr5: 42624748-42628325 as the tag region.

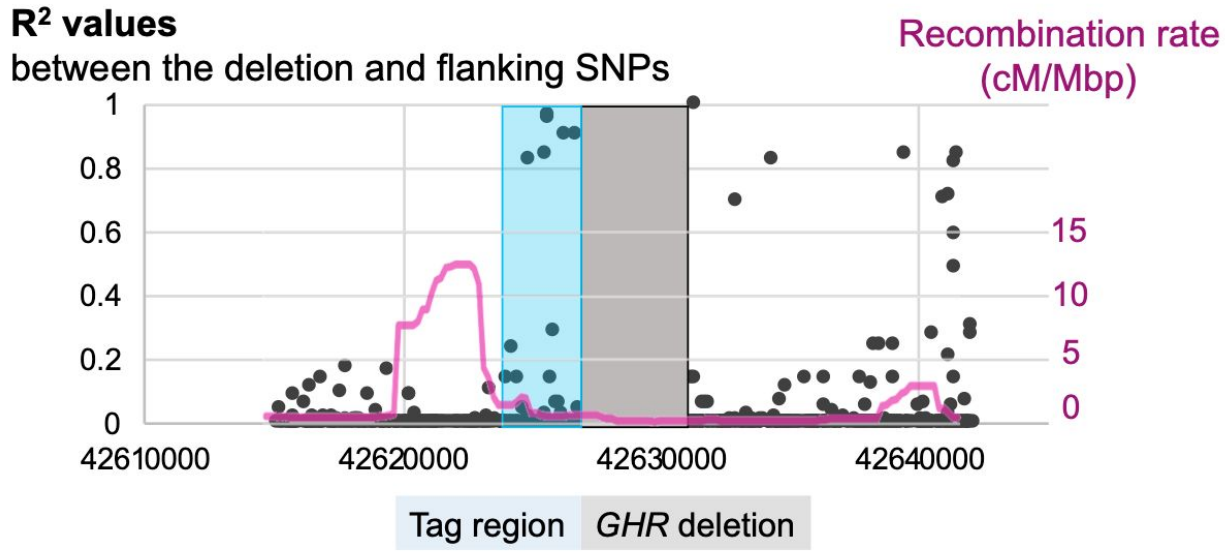


Figure S2. *GHRd3* tag SNP frequency map. Global frequencies of the *GHRd3* tag SNP from the Human Genome Diversity Project (https://www.hagsc.org/hgdp/); The C allele tags the derived deletion, despite being the ancestral allele.

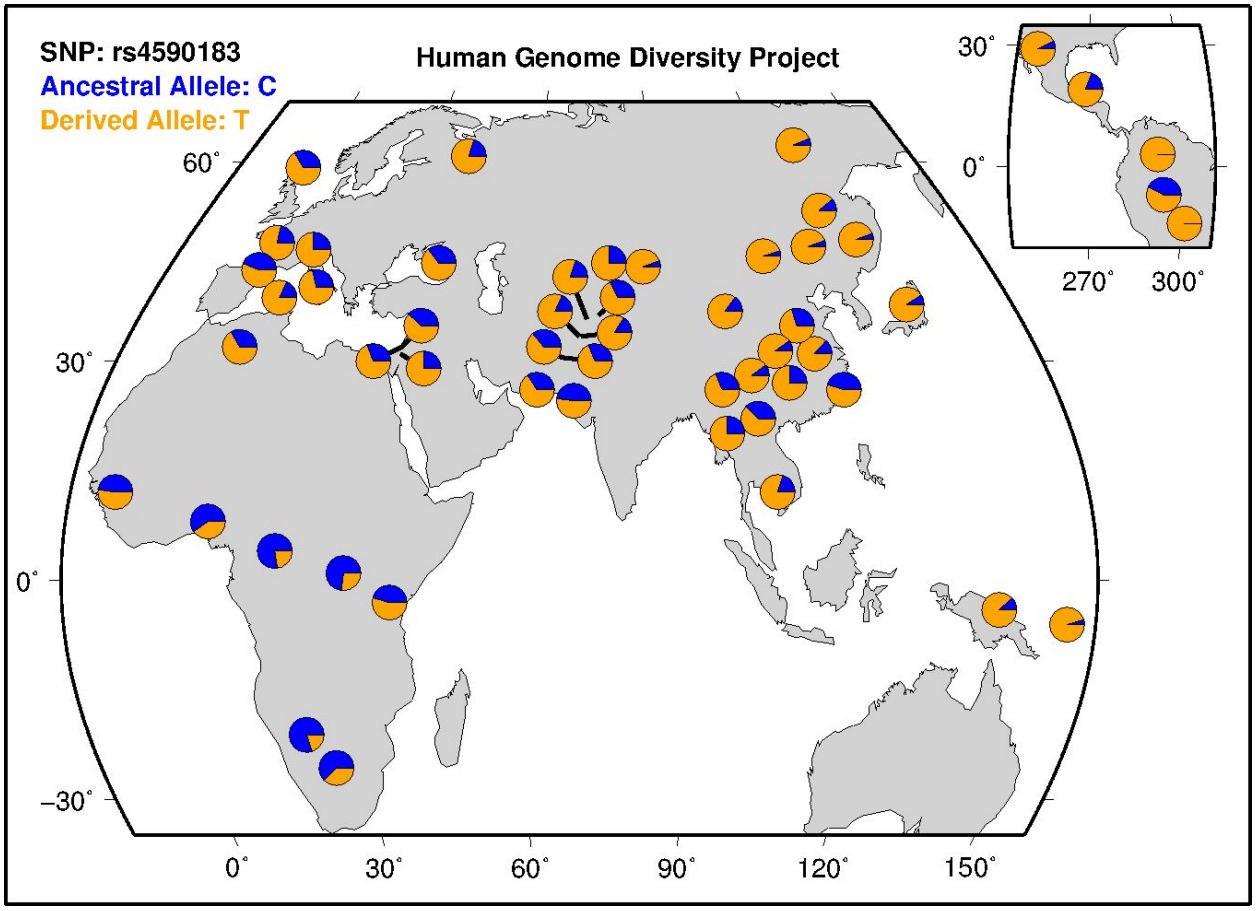


Figure S3. *GHR* exon 3 phylogenetic tree. This phylogenetic tree was constructed using randomly selected haplotypes that harbor the deleted and non-deleted *GHR* exon 3 alleles. It is clear from this phylogeny through robust branch supports that haplotypes harboring the deleted allele are more diverse and coalesce earlier than those that harbor the non-deleted allele. It is also noteworthy that both the Altai Neanderthal and Denisova genomes cluster with the haplotypes harboring the deleted allele.

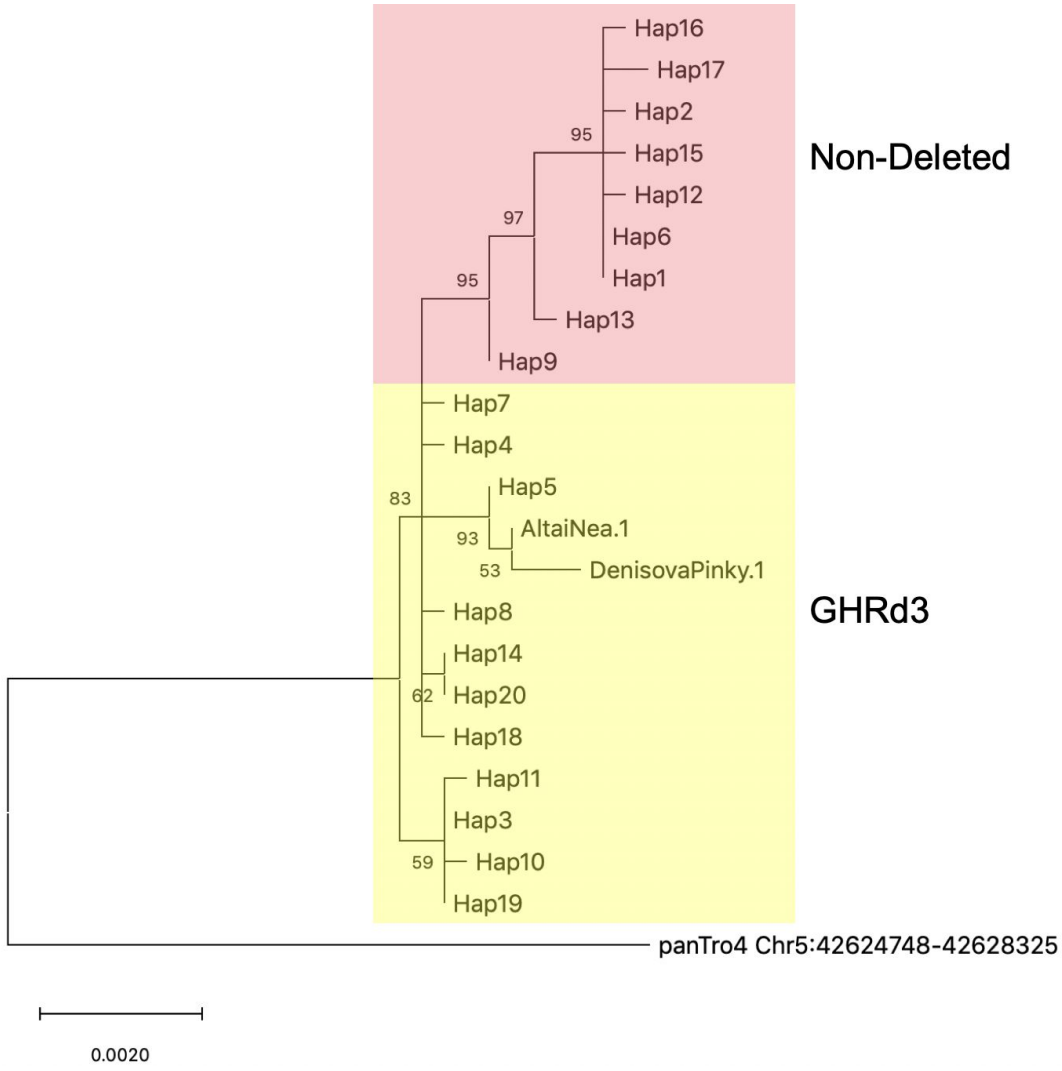


Figure S4. Tajima's D values for all haplotypes. Tajima's D values between the *GHRd3* tag region (Hg19: chr5: 42624748-42628325, blue) and 500 randomly selected regions across the genome (white) for three populations.

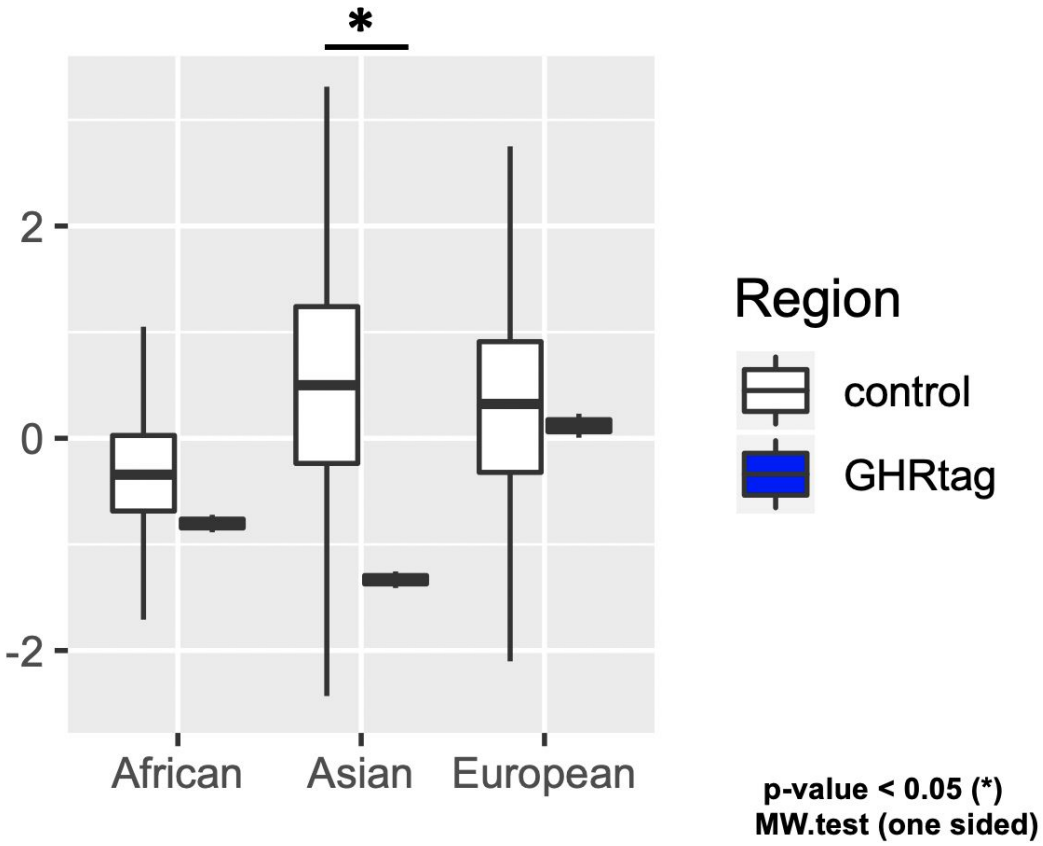


Figure S5. Approximate Bayesian computation testing different models. Decays of extended haplotype homozygosity (EHH) (left) and site frequency spectrum (SFS) (middle for the derived allele and right for the ancestral allele) in YRI (top) and CHB (bottom) populations. The EHH shows the probability that two randomly chosen haplotypes are homozygous at all SNPs within a given distance from a focal SNP site, which in this case is a tag variant for *GHRd3* (rs6873545). This measure is depicted as a value between 0 and 1. To fully capture the haplotype structure in this region, we recorded the physical positions of the variable sites on each side of the focal SNP where the EHH value decreased from 0.9 to 0.1, in steps of 0.1, for derived (red dots) and ancestral alleles (light blue dots) separately. The SFS covers the entire range of allele frequencies from singletons to mutations shared across all chromosomes, providing the local reduction in nucleotide diversity and the distortion of the SFS in a population. There were 113 chromosomes carrying the derived alleles and 103 carrying the ancestral alleles in the YRI samples; For the CHB samples, these numbers were 170 and 36, respectively. The bin size was set as "2". Therefore, the total number of bins for those chromosomes were 57 and 52 for YRI and 85 and 18 for CHB. We used all of the data shown in these plots as a set of summary statistics for the analysis of approximate Bayesian computation (ABC) (see Supplementary information).

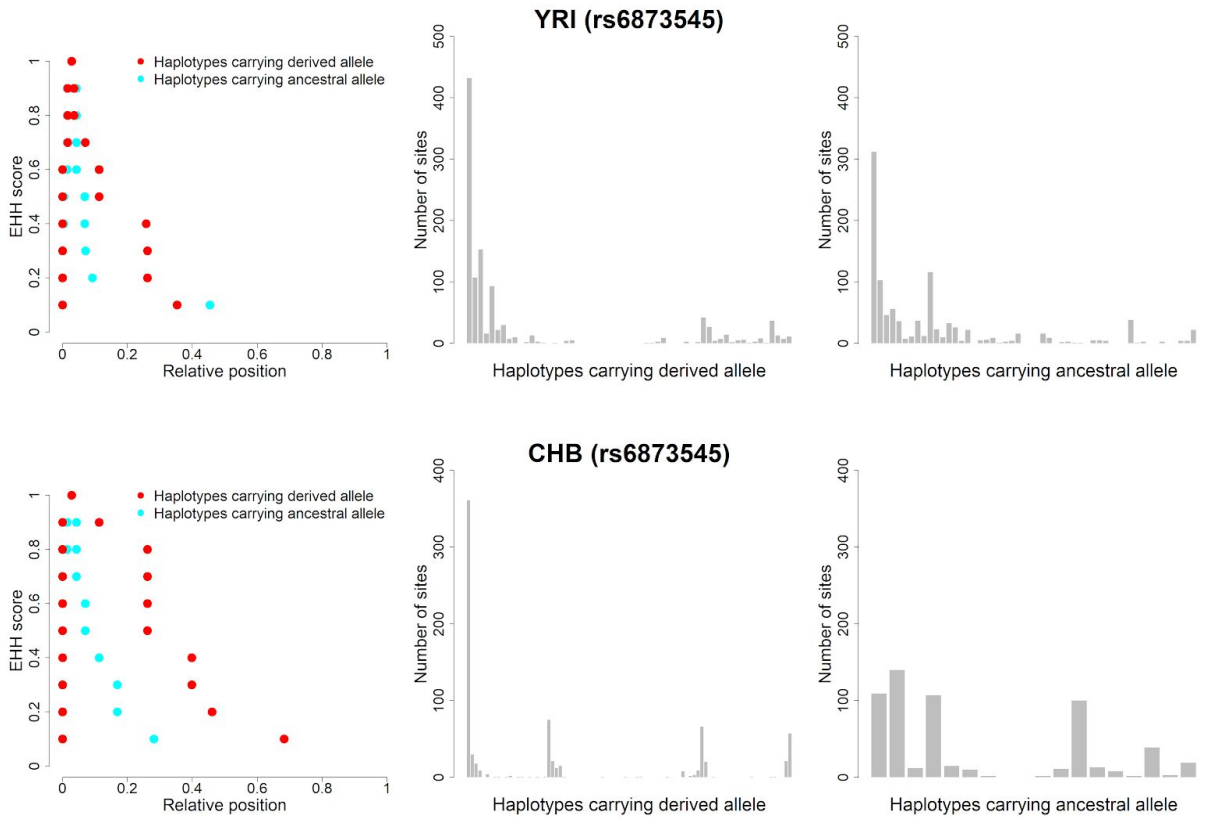


Figure S6. Estimation of allele frequency and the date for the onset of selection in the CHB population based on the results of the ABC simulations. The x-axis shows time, while the y-axis represents allele frequency. This plot shows means of T and f_t with 95% credible intervals. The dotted line indicates the difference in allele frequency between the present and t.

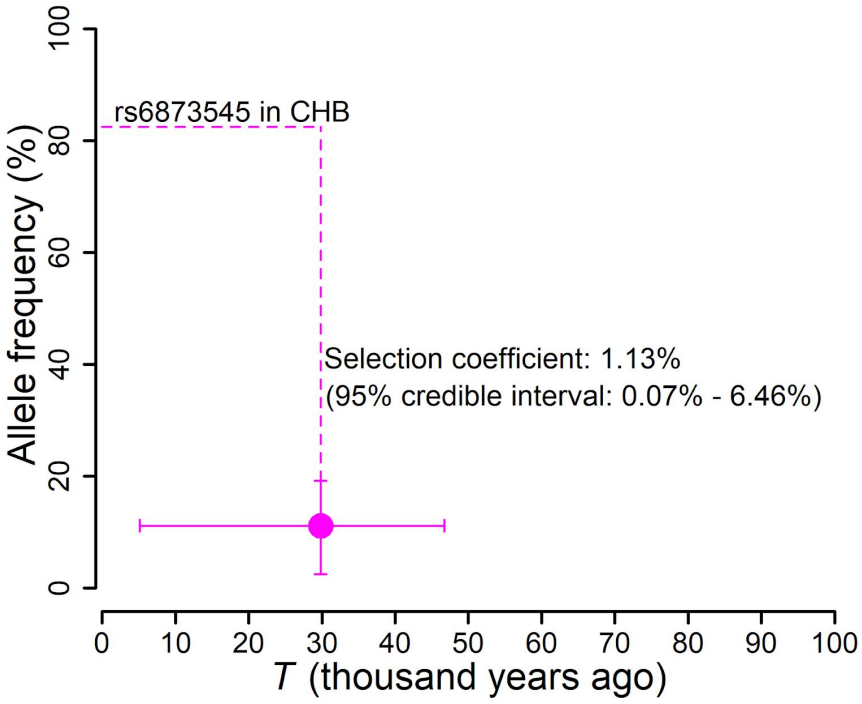


Figure S7. Phewas analysis on the *GHRd3* tag SNP. The UK biobank phewas (http://geneatlas.roslin.ed.ac.uk/phewas/) results using the tag SNP (rs4073476).

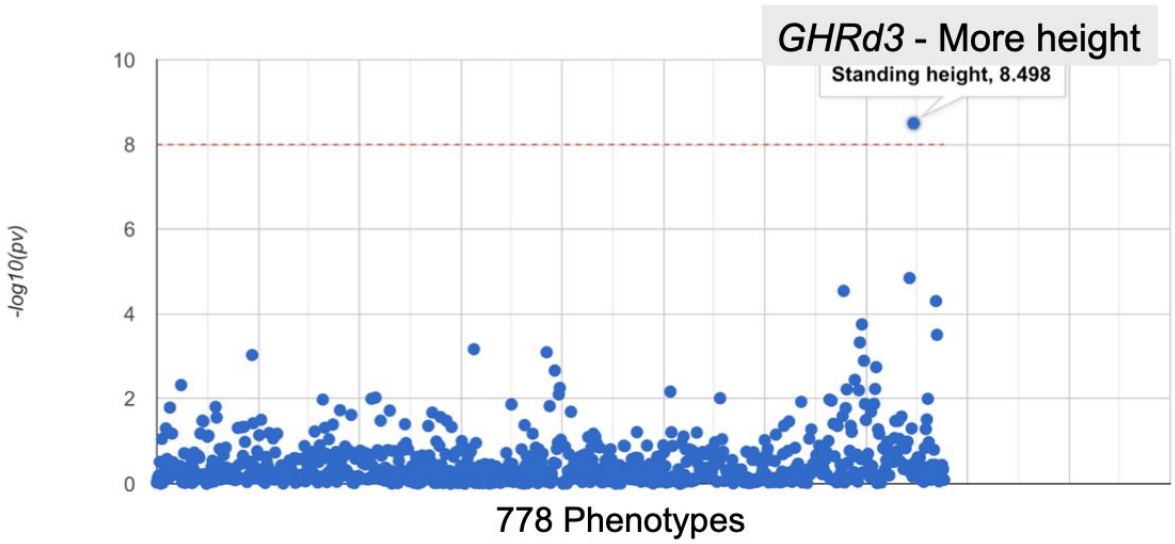


Figure S8. Weight in females. The weight measurements for females across a year period.

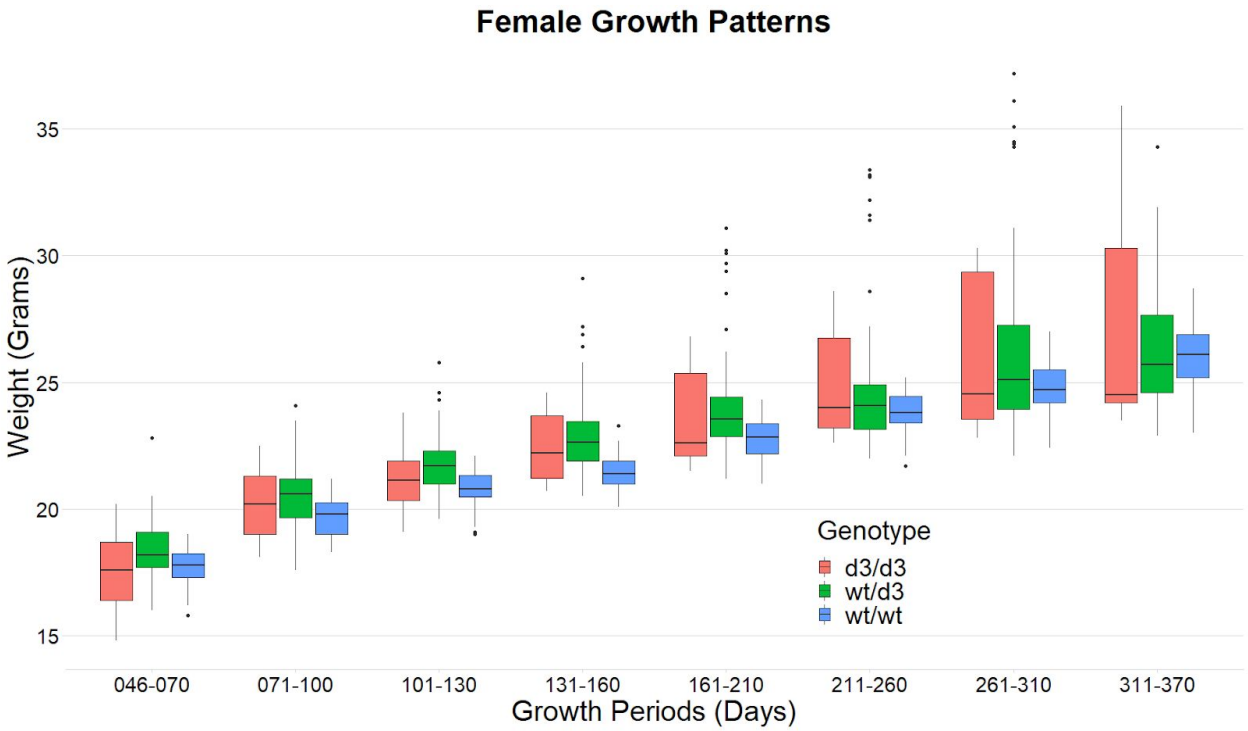


Figure S9. Volcano plots for the transcriptome analysis. The volcano plots depict the $\log_2$ fold change on the x-axis and the nominal -log p-value on the y-axis. The genes that show a nominal p-value < $5x10^{-7}$ were labeled on the graph.

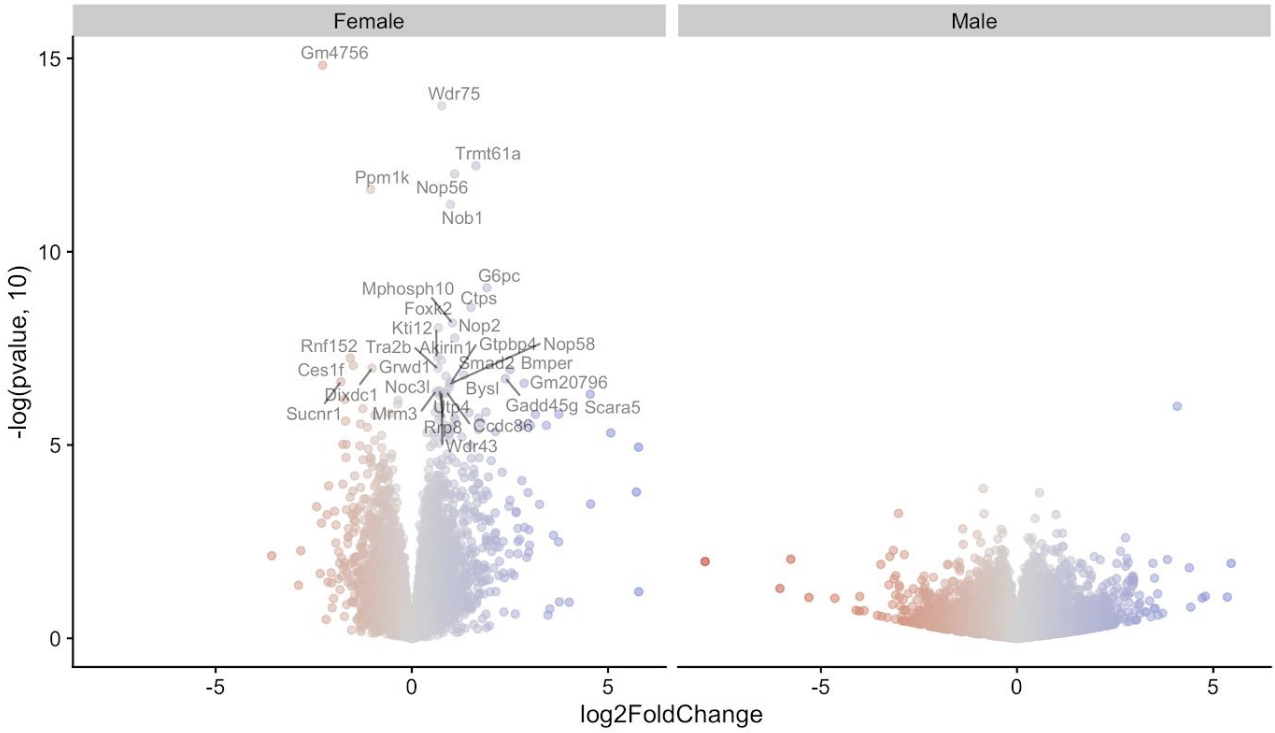


Figure S10. Variation in RNAseq. The analysis of variation in RNAseq data in two replicates. We found that the standard error of the fold-changes (as calculated by the DeSeq2 pipeline) are higher in males in both cases and in the merged dataset.

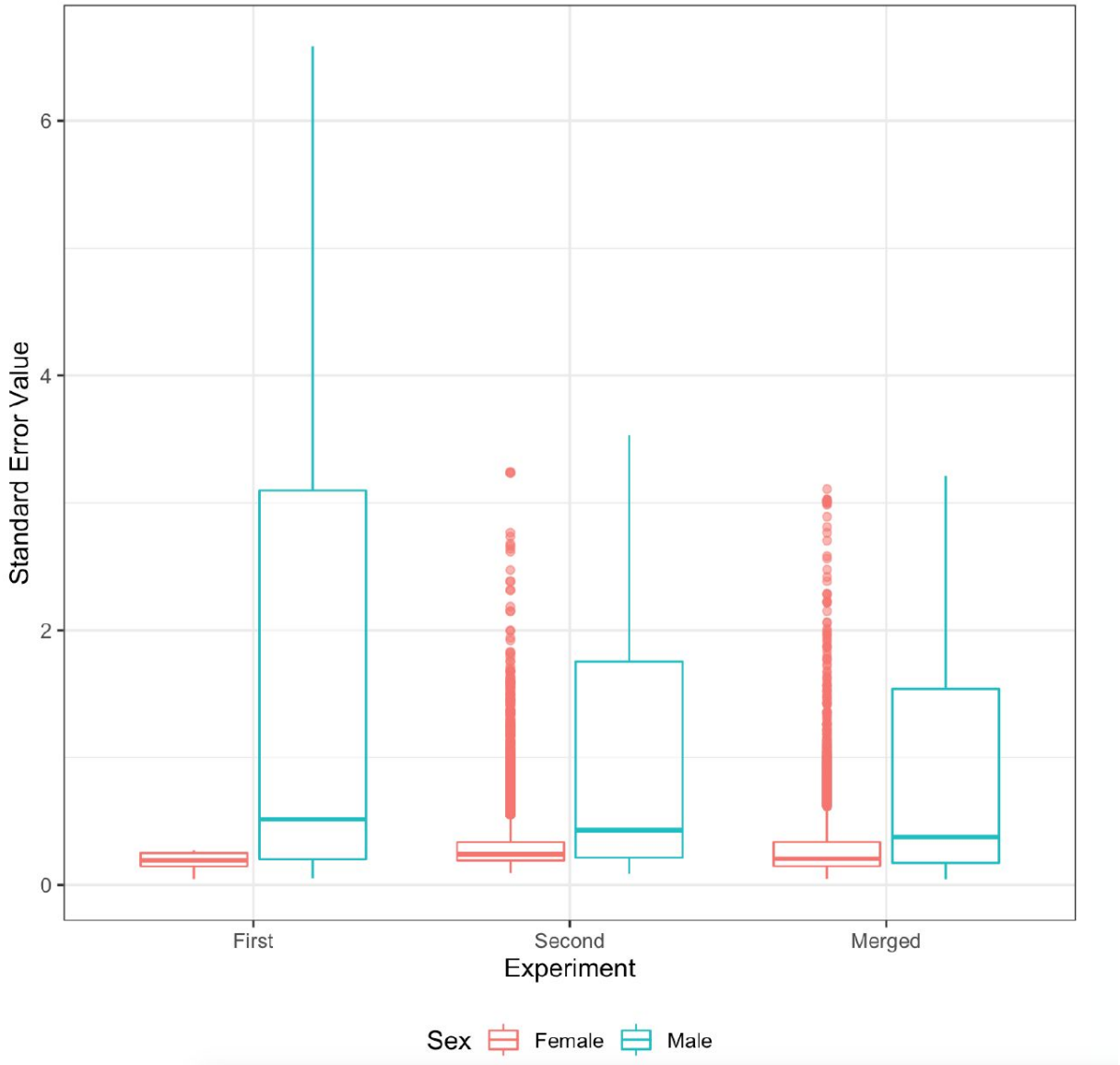


Figure S11. The standard error values among male livers for the genes that are significant or nonsignificant in female livers. We compared the standard error of the fold-changes (as calculated by the DeSeq2 pipeline) in males for genes that have different base expression levels (x-axis). We found that the genes that show significant differentiation between female *GHRd3* and wt mice also have significantly more inter-individual variation in males independent of their base expression. All pairwise comparisons are significant (**p<0.01, MW test**).

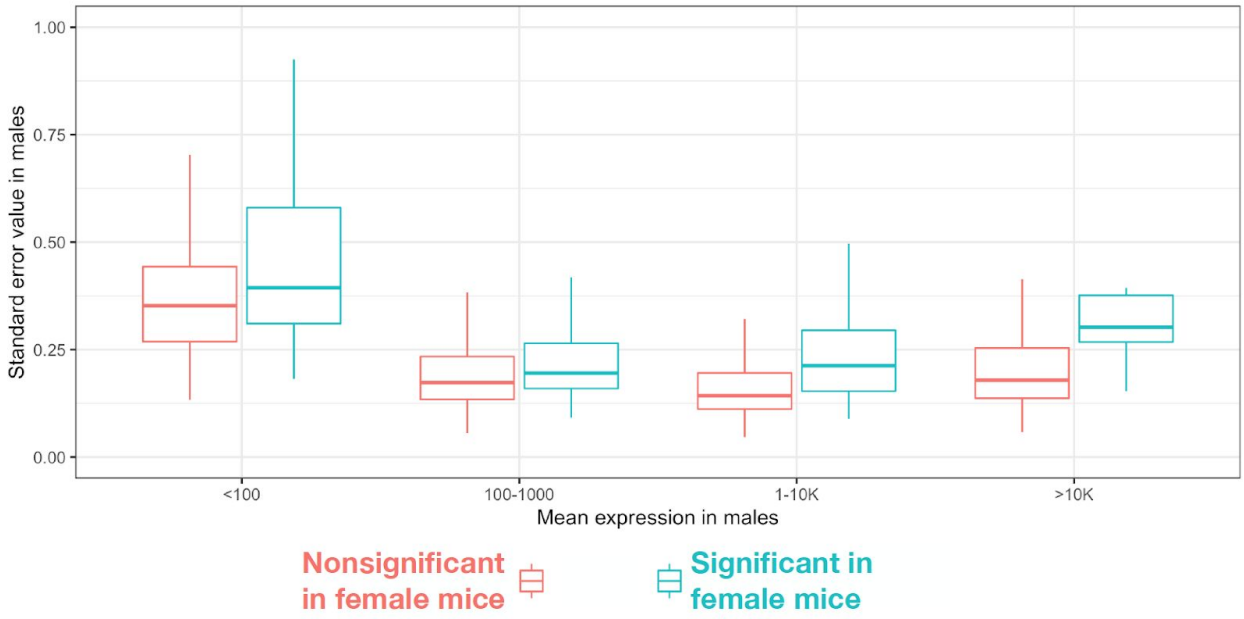
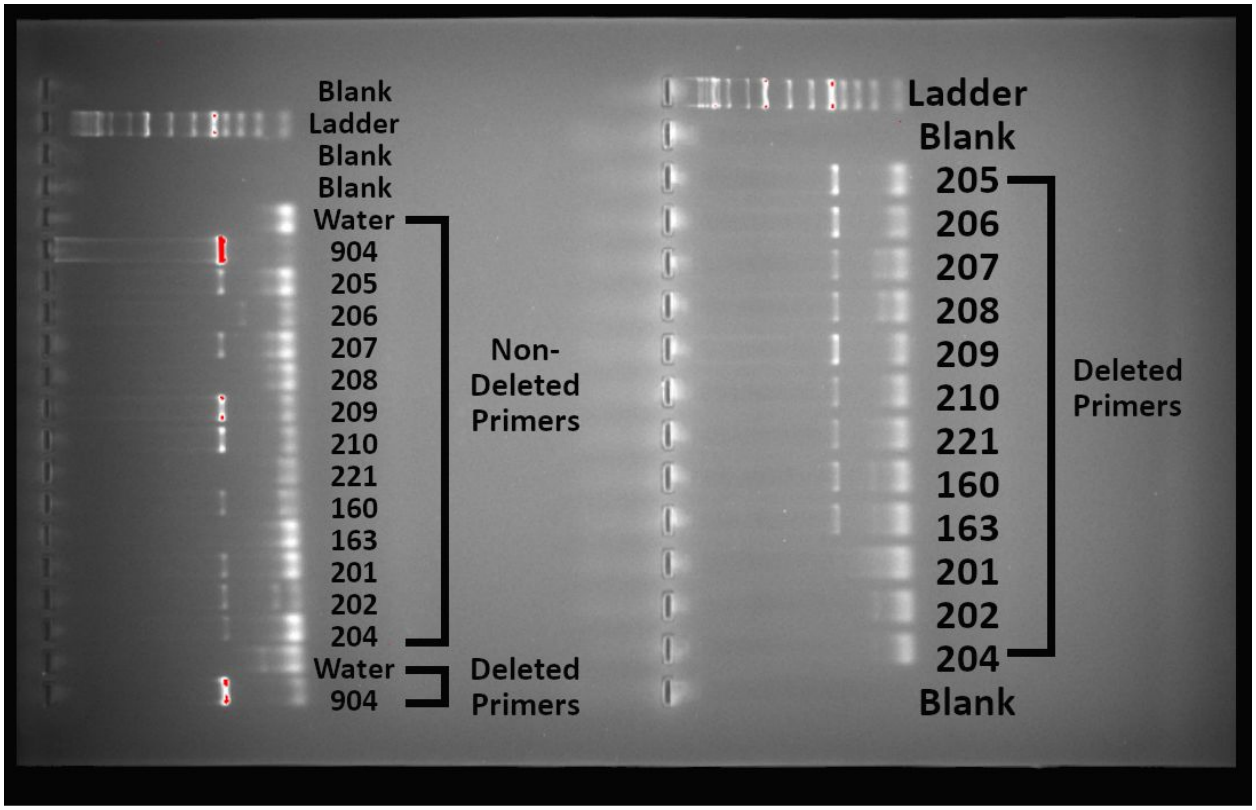


Figure S12. A gel exemplifying our genotyping approach. A two percent agarose gel showing polymerase chain reaction products. Two sets of primers were used. A sample that showed amplification with the non-deleted primers, but not the deleted primers would be deemed wt/wt; A sample that showed amplification with the deleted primers, but not the non-deleted primers would be deemed d3/d3; A sample that showed amplification with both sets of primers would be deemed wt/d3. Water was used as a negative control and a known heterozygote (904) was used as a positive control.



Conclusions

Water - No amplification (Negative Control)
904 - wt/d3 (Positive Control)
205 - wt/d3
206 - d3/d3
207 - wt/d3
208 - d3/d3
209 - wt/d3
210 - wt/d3
221 - d3/d3
160 - wt/d3
163 - d3/d3
201 - wt/wt
202 - wt/wt
204 - wt/wt

SUPPLEMENTARY METHODS

Haplotype network and read-depth analysis

We used the program vcftools (0.1.16) (*25*) to calculate the R^2 values between *GHRd3* (esv3604875) and flanking SNPs to set a target region (Hg19: chr5: 42624748-42628325). A vcf file of the target region obtained from the 1000 Genome Project phase 3 dataset (*26*), the hg19 reference genome, the chimpanzee reference genome (*27*), the Altai Neanderthal genome, and the Denisovan genome (*10*, *28*) were all used by the program VCTtoTree (V3.0.0) (*29*) to draw the haplotype networks. PopART (Version 1.7) (*30*) was used for the visualization of the network (**Fig. 2B**). The program rworldmap (*31*) was used to visualize the global allele frequency data from the 1000 Genome Project phase 3 dataset (*26*).

Conservation Analysis

We obtained the *GHR* coding sequence alignment from **Fig. 1B** through the UCSC genome browser by utilizing its "Other Species Alignments" function (MAF table: multiz100way) (*32*). This alignment was then viewed and exon 3 was highlighted using the program Molecular Evolutionary Genetics Analysis (MEGA v 10.0.5) (*33*).

The software program bedtools (v2.27.1) (*34*) was used to obtain 20,000 random sites from chromosome 5, where the *GHR* gene is located, for the conservation analysis. The phastCons46way.placental datasets (*35*) from the UCSC Genome Browser were used to compare the conservation scores of *GHR* exon 3 and the randomly selected regions (**Fig. 1C**).

Neutrality Tests

To obtain random single nucleotide variants, we used bedtools (v2.27.1) (*34*) to construct random chromosomal coordinates on chromosome 5. We used 500 random regions for the *GHRd3* locus comparison (**Fig. 2C**). Tajima's D (*36*) values for merged haplotypes and XP-EHH (*37*) values were downloaded from the 1000 Genomes selection browser (*38*) and used for the bins containing the target region (Hg19: chr5: 42624748-42628325) and control region. To increase sample numbers for our Tajima's D calculations (**Fig. 2C**) we extended the analysis beyond the CEU, YRI, and CHB populations. The European category was expanded to include the following populations: Utah residents with Northern and Western European ancestry (CEU), Toscani in Italy (TSI), Finnish in Finland (FIN), British in England and Scotland (GBR), and Iberian populations in Spain (IBS); The African category was expanded to include the following populations: Gambian in Western Division in the Gambia (GWD), Mende in Sierra Leone (MSL), Esan in Nigeria (ESN), Yoruba in Ibadan, Nigeria (YRI), and Luhya in Webuye, Kenya (LWK); The east Asian category was expanded to include the following populations: Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), Southern Han Chinese, China (CHS), Chinese Dai in Xishuangbanna, China (CDX), and Kinh in Ho Chi Minh City, Vietnam (KHV). The visualization was constructed through ggplot2 (*39*).

For the analysis of LD patterns in different site frequency spectra, we followed the approach outlined by Fujito et al. (*13*). We used the two-dimensional site frequency spectrum (*13*, *40*) to detect the signature of selection on *GHRd3*. This method can eliminate the effect of recombinations, which affect haplotype structures. We used the 1000 Genome meta-populations (Africa, East Asia, and Europe) for this analysis. By combining the site

frequency spectrum method and coalescence simulations, we detected the signature of selection on the *GHR* non-deletion allele ($p < 0.05$) with the estimated date of selection onset being 27.5ky. The results of this analysis are summarized in the table below.

	Africa:n=1322				East Asian:n=1008				European:n=1006			
ID	esv3604875				esv3604875				esv3604875			
position	42628418				42628418				42628418			
	der.		anc.		der.		anc.		der.		anc.	
m (l)	600 (17.65kb)		722 (17.6 kb)		205 (5.7kb)		803 (5.7 kb)		307 (155kb)		701 (155kb)	
S	234		234		49		49		1032		1032	
Fc	0.450	0.793	0.448	0.105	0.927	0.999	0.0588	0.001	0.630	0.972	0.632	0.17
Gc0	23.5	0.196	46.3	0.373	73.1	0.993	2.52	0.011	14.1	0.539	28.4	0.496
Gc0*	39.1	0.25	62.7	0.352	100.1	0.952	3.38	0.002	19.5	0.404	65.3	0.719
Lc0	0.0855	0.211	0.183	0.123	0.0713	0.901	0.00558	0.001	0.0283	0.452	0.0896	0.127
imax	233	0.095	555	0.911	148	0.91	9	0.003	197	0.726	699	0.999
gamma10	0.260	0.109	0.352	0.075	0.545	0.973	0	0.003	0.204	0.562	0.144	0.021
gamma10*	0.439	0.138	0.481	0.037	0.750	0.873	0	0.003	0.289	0.309	0.337	0.047
C(s.e.)	3.78 (0.932)		6.73 (1.85)		3.92 (1.63)		0.0785		12.6 (1.79)		17.5 (3.71)	
ul (x10-6/year)	8.825		8.825		2.85		2.85		77.5		77.5	
tD(my)	0.428329		0.762606		1.375439		0.02754386		0.162581		0.225806	
(ky)	428		763		1375		27.5		163		225	

To further clarify the population dynamics of *GHRd3* and the non-deleted alleles in YRI and CHB, we applied an approximate Bayesian computation framework developed in a previous study (*14*) to infer the mode and tempo of natural selection. We first simulated data from three different models, selection on a new mutation, selection on a standing variant, and neutrality, by using four parameters: allele frequency at present (f_0), the onset of natural selection (t), selection coefficient (s), and allele frequency at t (f_t). The details of our simulation conditions are shown in the table below. We then calculated the goodness-of-fit of each model to the observed summaries (**Fig. S5**) as an approximate marginal likelihood (aML) using a kernel density estimate (*41*). Once we identified the best fitting model in terms of approximate Bayes factors (aBFs), we estimated the parameters under the model by kernel ABC (*42*). All programs used in this analysis will be made available from https://github.com/shigekinakagome/.

Population	Demography	Prior conditions		
		s	t	f_t
YRI	Constant (N = 12,000)	$\log_{10}U(-3, -0.5)$	*Log-normal*$(1600, 1600^2)$	$U(0.0, 0.2)$
CHB	Population bottleneck and expansion ～1600 generations ago: N = 12,000 1100～1600: N = 1,200 500～1100: N = 12,000 500～Present: N = 120,000			

In the YRI population, the standing variation model showed the highest aML. However, this aML was almost comparable to the one under the neutral model (see the table below). The aML calculated for the CHB population was significantly higher in the standing variation model than in the other models, suggesting that the non-deleted allele became advantageous during the expansion into East Eurasia.

Population		Models		
		Neutral	Selection on new mutation	Selection on standing variant
YRI	Log(aML[1])	-214.687	-249.968	-212.995
	Log_{10}(aBF[2]) between a given model and the model with the highest aML	0.735	16.057	-
CHB	Log(aML)	-206.437	-224.875	-194.349
	Log_{10}(aBF) between a given model and the model with the highest aML	5.250	13.257	-

[1]Approximated marginal likelihood under a given model.
[2]Approximated Bayes factor shows how much one model is better fit to the observed data than the other.

To test this hypothesis, we further estimated the onset of natural selection under the standing variation model (see **Fig. S6**). The estimates showed that selection on the non-deleted allele occurred around 30,000 years ago when this allele was present at 11%. The frequency increased over time with a selection coefficient of 1.13%.

Association analysis

One of the single nucleotide variants (rs4073476) upstream of the exon 3 deletion almost perfectly tags *GHRd3* in European populations (R^2 >0.95). This variant was entered into the GeneATLAS PheWAS database (http://geneatlas.roslin.ed.ac.uk/phewas/) for an association analysis (*16*). This database identifies associations between SNVs and 778 phenotypic traits using the UK Biobank cohort (https://www.ukbiobank.ac.uk/). The nominal p-value 10^{-8} was used as a conservative threshold of significant association to find a link between rs4073476 and standing height in the UK population.

To further the association analysis, another database was utilized (http://www.nealelab.is/uk-biobank/). This database, created by the Neale lab, provides GWAS summary statistics for 4,203 phenotypes from the UK Biobank cohort. Summary statistics were downloaded for males, females, and both datasets combined for the top ten associations identified through the previous GeneATLAS analysis. These summary statistics allowed for gender specific association analyses.

Genome-edited mouse

To model *GHRd3*, we designed sgRNAs that flank both sides (5' and 3') of the mouse *Ghr* exon orthologous to the human *GHR* exon 3 with the help of the Gene Targeting and Transgenic

Resource at Roswell Park Comprehensive Cancer Center. Specifically, 18 5' and 11 3' primers were designed. These primers were screened for common single nucleotide polymorphisms in dbSNP (https://www.ncbi.nlm.nih.gov/snp/) and tested for off target cuts using a standard T7 endonuclease 1 (T7E1) mismatch detection assay (*43*). Based on the results of this assay, the following 4 sgRNA sequences were used for the downstream analysis:

mGHR5'.g5 - AATACAATTGGCTAATACCGNGG
mGHR5'.g6 - TTGGCTAATACCGAGGTGAGNGG
mGHR3'.g10 - TAAGATTTTTAGTGATGTAANGG
mGHR3'.g11 - AAACATGACCATTGAATTAANGG

These primers were further validated using Bioanalyzer and Qubit for adequate concentrations (~300ng/uL). C57BL/6N fertilized embryos were harvested at day 0.5 dpc. Fertilized embryos were then injected with guide RNAs and Cas9 into the pronucleus to perform genome editing.

Genotyping

The Vindija (*10*) and Chagyrskaya Neanderthal bam files were analyzed using the Integrative Genomics Viewer software (v2.3.75) (*44*) to determine if these genomes contained *GHRd3* (https://www.eva.mpg.de/genetics/genome-projects/chagyrskaya-neandertal/home.html?Fsize=0Svea-Developmental). These results were then added to previous findings regarding the Altai Neanderthal and Denisovan genomes (*9*) **(Fig. 1A)**.

The mice were screened for *GHRd3* using the Kapa Biosystems KAPA2G Fast HS Genotyping kit alongside two sets of primers (**Fig. S11**), following the protocol that can be found on their website (www.kapabiosystems.com). One set of primers leads to amplification if the deletion is present; the other set of primers leads to amplification if the non-deleted, ancestral haplotype is present. For heterozygous individuals, amplification occurs with both sets of primers. For each experiment, water was used as a negative control and a DNA sample from a known heterozygote was used as a positive control.

Deletion Primers: Forward - SM444CEL F (AGAGTACCCAGTGTATGGCCT). Reverse - SM445CEL R (TGCTGTCTGGCACACATGAT)
Non-Deletion Primers: Forward SM444CEL F (AGAGTACCCAGTGTATGGCCT). Reverse - SM444CEL R (AGTTCTGTGAGCTGGTGTAGC)

In parallel, to validate the successful genome-editing of the mice at the transcriptome level, we conducted a read-depth analysis on bam files made from the transcriptomes. To elaborate, the processed fastQC files were converted into bam files using the GRCm38 mouse reference genome and the programs tophat2 (*45*) and bowtie2 (*46*). We then observed the read depths of transcripts from exon 3 and exon 4 (the latter for comparison) of the mouse *Ghr* by using the following Samtools (*47*).

```
samtools depth -q 0 -Q 0 -b mouseGHRex3.bed M3.bam  | awk '{sum+=$3;cnt++}END{print sum/cnt" "sum" "cnt}'>  M3.depth
```

Based on the read-depth information, the presence or absence of exon 3 expression was readily detectable (**Fig. S5**).

Weight measurement

Weight measurements were taken for 41 mice born of heterozygous parents until each of them turned a year old (**Table S1**). Measurements were taken anywhere between twice a week and once every two weeks depending on the developmental period the mice were in. Food and water supplies were kept constant and consistent across cages (standard chow diet). The mice were separated by sex, but kept in cages alongside their siblings regardless of genotype. At most, a cage would contain 4 male mice or 5 female mice.

Transcriptomics data and analysis

28 to 34-day-old mice were sacrificed (**Table S2**) and liver samples were taken and directly put into RNAlater (Thermo Fisher). They were then sent for RNA extraction and sequencing by GENEWIZ. RNA sequencing was performed via Illumina HiSeq, 2x150bp configuration. Quality control of the obtained sequences was performed using FastQC (Andrews S. FASTQC. A quality control tool for high throughput sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 7/2019), and the results of all the samples were reviewed by MultiQC (*48*). Adaptor sequences and low quality bases were discarded by Trimmomatic (*49*). Filtered reads were mapped to mouse transcriptome reference (GRCm38) from Ensembl (*50*) and quantified using Kallisto (*51*). The transcriptswere merged into genes with tximport and ensembl biomaRt (*52*) Differential expression analyses were performed by DESeq2 (*53*), which calculates the fold change of each gene using the Wald test and a correction for multiple hypotheses. The gene expressions of the samples are also provided in **Table S2**. We then defined genes that were upregulated or downregulated using the adjusted (i.e., multiple hypothesis corrected) *p*-value threshold of 0.05 (Wald test in DeSeq2).

The genes that were upregulated or downregulated in *Ghrd3* mice (**Table S3**) were used against a whole genome background to investigate gene ontology enrichment using ShinyGo (*54*).

Lipidomics Analysis

An internal standard mix was prepared in $CHCl_3$ containing exogenous lipids d_9 oleic acid (2 μM), d_{11} arachidonic acid (2 μM), *d*70 distearoylphosphatidylcholine (2 μM), d_{31} C16 sphingomyelin (2 μM), C17 glucosylceramide (2 μM), C57 triacylglycerol (1 μM), C28 diacylglycerol (0.5 μM), dihydrolanosterol (10 μM), and C17 ceramide (0.5 μM).

Lipid extraction and analysis were carried out as described in previous studies (*55*, *56*). 28 to 34-day-old mice were sacrificed (**Table S4**) and serum samples were syphoned. The serum was collected using standard centrifugation methods (*57*) and later stored at -20 degrees. For the LC-MS analysis, serum samples were thawed on ice, gently mixed by pipetting, then 30 μL samples (or water for blank extractions) were transferred to 1 mL cold PBS contained in a 2-dram glass vial on ice. Samples were then vortexed three times for 3 seconds to mix and a 30 μL aliquot was taken for protein measurement using a Coomassie assay. The relative standard deviation in protein concentration among the samples was 5%. 1 mL of methanol, 1.8 mL of $CHCl_3$, and 200 μL of internal standard mix were added to the remainder. All solutions were kept cold on ice. Samples were vortexed for 10 seconds, followed by 1 min on ice, three times.

Samples were then centrifuged for 30 minutes at 3000 rcf at 4°C. The bottom ($CHCl_3$) layer was transferred to a 1 dram glass vial and kept on ice. The upper layer was re-extracted with 2 mL of additional $CHCl_3$ and the two $CHCl_3$ layers were combined. 3.75 mL of the combined $CHCl_3$ layers were transferred to a new 1 dram vial and the solvent was removed by rotary evaporation.

Samples were resuspended in 150 µL of $CHCl_3$ spiked with $^{13}C_{18}$ oleic acid (2 µM), C39 triacylglycerol (1 µM), and C6 ceramide (0.5 µM). Samples were analyzed by LC-MS using an Agilent Infinity 1260 HPLC/ Agilent 6530 Jet Stream ESI-QToF-MS system. Mobile phase A was composed with 95% H_2O and 5% methanol. Mobile phase B was composed with 60% isopropanol, 35% methanol, and 5% water. Samples were analyzed in positive ionization mode using 0.1% (v/v) formic acid and 5 mM ammonium formate as additives in the mobile phases with separation performed using a Luna C5 column (5 µm C5 100 Å, 50 x 4.6 mm) with a C5 guard cartridge. Samples were analyzed in negative ionization mode using and 0.1% ammonium hydroxide as an additive in the mobile phases with separation performed using a Gemini C18 column (5 µm C18 110 Å, 50 x 4.6 mm) with a C18 guard cartridge. 5 µL of the resuspended sample was injected for analysis. The LC method began with 5 min of 0% B at 0.1 mL/min, then the flow rate was increased to 0.5 mL/min and mobile phase gradient began with 0% B to 100% B over 60 min. 100% B was maintained for 7 min, then switched to 0% B for 8 min. LC-MS data analysis was performed using the Agilent MassHunter Qualitative Analysis software (v. B.06.00). LC-MS grade methanol and isopropanol were obtained from Millipore Sigma; LC-MS grade chloroform was obtained from Honeywell. Internal standards for LC-MS were obtained from Avanti Polar lipids, with the exceptions of $^{13}C_{18}$ oleic acid (Cambridge Isotope laboratories) and C39, C57 TAGs (Millipore Sigma). LC columns were obtained from Phenomenex. Coomassie protein assay kits were obtained from Thermo Scientific.

The lipid levels of the samples are provided in **Table S4**.

Supplementary tables:

Table S1. Weight data

Table S2. Transcriptome (tab 1. Sample list, Tab2. Raw data, Tab3. pvalues, fold change females, Tab4. Males

Table S3. GO analysis results

Table S4. Lipid analysis

REFERENCES

1. A. J. Brooks, M. J. Waters, The growth hormone receptor: mechanism of activation and clinical implications. *Nat. Rev. Endocrinol.* **6**, 515–525 (2010).

2. M. Zoledziewska, C. Sidore, C. W. K. Chiang, S. Sanna, A. Mulas, M. Steri, F. Busonero, J. H. Marcus, M. Marongiu, A. Maschio, D. Ortega Del Vecchyo, M. Floris, A. Meloni, A. Delitala, M. P. Concas, F. Murgia, G. Biino, S. Vaccargiu, R. Nagaraja, K. E. Lohmueller, UK10K consortium, N. J. Timpson, N. Soranzo, I. Tachmazidou, G. Dedoussis, E. Zeggini, Understanding Society Scientific Group, S. Uzzau, C. Jones, R. Lyons, A. Angius, G. R. Abecasis, J. Novembre, D. Schlessinger, F. Cucca, Height-reducing variants and selection for short stature in Sardinia. *Nat. Genet.* **47**, 1352–1356 (2015).

3. S. Amselem, P. Duquesnoy, B. Duriez, F. Dastot, M. L. Sobrier, S. Valleix, M. Goossens, Spectrum of growth hormone receptor mutations and associated haplotypes in Laron syndrome. *Hum. Mol. Genet.* **2**, 355–359 (1993).

4. R. Padidela, S. M. Bryan, S. Abu-Amero, R. E. Hudson-Davies, J. C. Achermann, G. E. Moore, P. C. Hindmarsh, The growth hormone receptor gene deleted for exon three (GHRd3) polymorphism is associated with birth and placental weight. *Clin. Endocrinol.* . **76**, 236–240 (2012).

5. K. Sørensen, L. Aksglaede, J. H. Petersen, H. Leffers, A. Juul, The exon 3 deleted growth hormone receptor gene is associated with small birth size and early pubertal onset in healthy boys. *J. Clin. Endocrinol. Metab.* **95**, 2819–2826 (2010).

6. D. Ben-Avraham, D. R. Govindaraju, T. Budagov, D. Fradin, P. Durda, B. Liu, S. Ott, D. Gutman, L. Sharvit, R. Kaplan, P. Bougnères, A. Reiner, A. R. Shuldiner, P. Cohen, N. Barzilai, G. Atzmon, The GH receptor exon 3 deletion is a marker of male-specific exceptional longevity associated with increased GH sensitivity and taller stature. *Sci Adv*. **3**, e1602025 (2017).

7. K. Sørensen, L. Aksglaede, T. Munch-Andersen, N. J. Aachmann-Andersen, H. Leffers, J. W. Helge, L. Hilsted, A. Juul, Impact of the growth hormone receptor exon 3 deletion gene polymorphism on glucose metabolism, lipids, and insulin-like growth factor-I levels during puberty. *J. Clin. Endocrinol. Metab.* **94**, 2966–2969 (2009).

8. C. Dos Santos, L. Essioux, C. Teinturier, M. Tauber, V. Goffin, P. Bougnères, A common polymorphism of the growth hormone receptor is associated with increased responsiveness to growth hormone. *Nat. Genet.* **36**, 720–724 (2004).

9. Y.-L. Lin, P. Pavlidis, E. Karakoc, J. Ajay, O. Gokcumen, The evolution and functional impact of human deletion variants shared with archaic hominin genomes. *Mol. Biol. Evol.* **32**, 1008–1019 (2015).

10. K. Prüfer, F. Racimo, N. Patterson, F. Jay, S. Sankararaman, S. Sawyer, A. Heinze, G. Renaud, P. H. Sudmant, C. de Filippo, H. Li, S. Mallick, M. Dannemann, Q. Fu, M. Kircher, M. Kuhlwilm, M. Lachmann, M. Meyer, M. Ongyerth, M. Siebauer, C. Theunert, A. Tandon, P. Moorjani, J. Pickrell, J. C. Mullikin, S. H. Vohr, R. E. Green, I. Hellmann, P. L. F. Johnson, H. Blanche, H. Cann, J. O. Kitzman, J. Shendure, E. E. Eichler, E. S. Lein, T. E.

Bakken, L. V. Golovanova, V. B. Doronichev, M. V. Shunkov, A. P. Derevianko, B. Viola, M. Slatkin, D. Reich, J. Kelso, S. Pääbo, The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*. **505**, 43–49 (2014).

11. F. Tajima, Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. **123**, 585–595 (1989).

12. P. C. Sabeti, D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson, G. J. McDonald, H. C. Ackerman, S. J. Campbell, D. Altshuler, R. Cooper, D. Kwiatkowski, R. Ward, E. S. Lander, Detecting recent positive selection in the human genome from haplotype structure. *Nature*. **419**, 832–837 (2002).

13. N. T. Fujito, Y. Satta, T. Hayakawa, N. Takahata, A new inference method for detecting an ongoing selective sweep. *Genes Genet. Syst.* **93**, 149–161 (2018).

14. S. Nakagome, R. R. Hudson, A. Di Rienzo, Inferring the model and onset of natural selection under varying population size from the site frequency spectrum and haplotype structure. *Proc. Biol. Sci.* **286**, 20182541 (2019).

15. C. Sudlow, J. Gallacher, N. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. Landray, B. Liu, P. Matthews, G. Ong, J. Pell, A. Silman, A. Young, T. Sprosen, T. Peakman, R. Collins, UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

16. O. Canela-Xandri, K. Rawlik, A. Tenesa, An atlas of genetic associations in UK Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).

17. P. J. Danneman, M. A. Suckow, C. Brayton, *The Laboratory Mouse* (CRC Press, 2012).

18. G. Norstedt, R. Palmiter, Secretory rhythm of growth hormone regulates sexual differentiation of mouse liver. *Cell*. **36**, 805–812 (1984).

19. R. A. Saxton, D. M. Sabatini, mTOR Signaling in Growth, Metabolism, and Disease. *Cell*. **169**, 361–371 (2017).

20. R. J. Strawbridge, L. Kärvestedt, C. Li, S. Efendic, C. G. Ostenson, H. F. Gu, K. Brismar, GHR exon 3 polymorphism: association with type 2 diabetes mellitus and metabolic disorder. *Growth Horm. IGF Res.* **17**, 392–398 (2007).

21. D. S. Millar, M. D. Lewis, M. Horan, V. Newsway, D. A. Rees, T. E. Easter, G. Pepe, O. Rickards, M. Norin, M. F. Scanlon, M. Krawczak, D. N. Cooper, Growth hormone (GH1) gene variation and the growth hormone receptor (GHR) exon 3 deletion polymorphism in a West-African population. *Mol. Cell. Endocrinol.* **296**, 18–25 (2008).

22. J. C. K. Wells, The evolution of human fatness and susceptibility to obesity: an ethological approach. *Biol. Rev. Camb. Philos. Soc.* **81**, 183–205 (2006).

23. A. Powell, S. Shennan, M. G. Thomas, Late Pleistocene demography and the appearance of modern human behavior. *Science*. **324**, 1298–1301 (2009).

24. R. G. Klein, Archeology and the evolution of human behavior. *Evolutionary Anthropology:*

Issues, News, and Reviews. **9**, 17–36 (2000).

25. 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**, 2156–2158 (2011).

26. P. H. Sudmant, T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M. Hsi-Yang Fritz, M. K. Konkel, A. Malhotra, A. M. Stütz, X. Shi, F. Paolo Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M. J. P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H. Y. K. Lam, X. Jasmine Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J. M. Kidd, Y. Kong, E.-W. Lameijer, S. McCarthy, P. Flicek, R. A. Gibbs, G. Marth, C. E. Mason, A. Menelaou, D. M. Muzny, B. J. Nelson, A. Noor, N. F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E. E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A. A. Shabalin, A. Untergasser, J. A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M. A. Batzer, S. A. McCarroll, 1000 Genomes Project Consortium, R. E. Mills, M. B. Gerstein, A. Bashir, O. Stegle, S. E. Devine, C. Lee, E. E. Eichler, J. O. Korbel, An integrated map of structural variation in 2,504 human genomes. *Nature*. **526**, 75–81 (2015).

27. The Chimpanzee Sequencing Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*. **437**, 69–87 (2005).

28. D. Reich, R. E. Green, M. Kircher, J. Krause, N. Patterson, E. Y. Durand, B. Viola, A. W. Briggs, U. Stenzel, P. L. F. Johnson, T. Maricic, J. M. Good, T. Marques-Bonet, C. Alkan, Q. Fu, S. Mallick, H. Li, M. Meyer, E. E. Eichler, M. Stoneking, M. Richards, S. Talamo, M. V. Shunkov, A. P. Derevianko, J.-J. Hublin, J. Kelso, M. Slatkin, S. Pääbo, Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature*. **468**, 1053–1060 (2010).

29. D. Xu, Y. Jaber, P. Pavlidis, O. Gokcumen, VCFtoTree: a user-friendly tool to construct locus-specific alignments and phylogenies from thousands of anthropologically relevant genome sequences. *BMC Bioinformatics*. **18**, 426 (2017).

30. J. W. Leigh, D. Bryant, popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**, 1110–1116 (2015).

31. A. South, rworldmap: A New R package for Mapping Global Data. *R J.* **3**, 35–43 (2011).

32. R. M. Kuhn, D. Haussler, W. J. Kent, The UCSC genome browser and associated tools. *Brief. Bioinform.* **14**, 144–161 (2013).

33. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

34. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. **26**, 841–842 (2010).

35. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral

substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).

36. F. Tajima, Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics*. **135**, 599–607 (1993).

37. P. C. Sabeti, P. Varilly, B. Fry, J. Lohmueller, E. Hostetter, C. Cotsapas, X. Xie, E. H. Byrne, S. A. McCarroll, R. Gaudet, S. F. Schaffner, E. S. Lander, International HapMap Consortium, K. A. Frazer, D. G. Ballinger, D. R. Cox, D. A. Hinds, L. L. Stuve, R. A. Gibbs, J. W. Belmont, A. Boudreau, P. Hardenbol, S. M. Leal, S. Pasternak, D. A. Wheeler, T. D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, H. Zhao, J. Zhou, S. B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R. C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, Y. Shen, W. Sun, H. Wang, Y. Wang, Y. Wang, X. Xiong, L. Xu, M. M. Y. Waye, S. K. W. Tsui, H. Xue, J. T.-F. Wong, L. M. Galver, J.-B. Fan, K. Gunderson, S. S. Murray, A. R. Oliphant, M. S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J.-F. Olivier, M. S. Phillips, S. Roumy, C. Sallée, A. Verner, T. J. Hudson, P.-Y. Kwok, D. Cai, D. C. Koboldt, R. D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L.-C. Tsui, W. Mak, Y. Q. Song, P. K. H. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C. P. Bird, M. Delgado, E. T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B. E. Stranger, P. Whittaker, D. R. Bentley, M. J. Daly, P. I. W. de Bakker, J. Barrett, Y. R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D. J. Richter, P. Sabeti, R. Saxena, S. F. Schaffner, P. C. Sham, P. Varilly, D. Altshuler, L. D. Stein, L. Krishnan, A. V. Smith, M. K. Tello-Ruiz, G. A. Thorisson, A. Chakravarti, P. E. Chen, D. J. Cutler, C. S. Kashuk, S. Lin, G. R. Abecasis, W. Guan, Y. Li, H. M. Munro, Z. S. Qin, D. J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L. R. Cardon, G. Clarke, D. M. Evans, A. P. Morris, B. S. Weir, T. Tsunoda, T. A. Johnson, J. C. Mullikin, S. T. Sherry, M. Feolo, A. Skol, H. Zhang, C. Zeng, H. Zhao, I. Matsuda, Y. Fukushima, D. R. Macer, E. Suda, C. N. Rotimi, C. A. Adebamowo, I. Ajayi, T. Aniagwu, P. A. Marshall, C. Nkwodimmah, C. D. M. Royal, M. F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I. F. Adewole, B. M. Knoppers, M. W. Foster, E. W. Clayton, J. Watkin, R. A. Gibbs, J. W. Belmont, D. Muzny, L. Nazareth, E. Sodergren, G. M. Weinstock, D. A. Wheeler, I. Yakub, S. B. Gabriel, R. C. Onofrio, D. J. Richter, L. Ziaugra, B. W. Birren, M. J. Daly, D. Altshuler, R. K. Wilson, L. L. Fulton, J. Rogers, J. Burton, N. P. Carter, C. M. Clee, M. Griffiths, M. C. Jones, K. McLay, R. W. Plumb, M. T. Ross, S. K. Sims, D. L. Willey, Z. Chen, H. Han, L. Kang, M. Godbout, J. C. Wallenburg, P. L'Archevêque, G. Bellemare, K. Saeki, H. Wang, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A. L. Holden, L. D. Brooks, J. E. McEwen, M. S. Guyer, V. O. Wang, J. L. Peterson, M. Shi, J. Spiegel, L. M. Sung, L. F. Zacharia, F. S. Collins, K. Kennedy, R. Jamieson, J. Stewart, Genome-wide detection and characterization of positive selection in human populations. *Nature*. **449**, 913–918 (2007).

38. M. Pybus, G. M. Dall'Olio, P. Luisi, M. Uzkudun, A. Carreño-Torres, P. Pavlidis, H. Laayouni, J. Bertranpetit, J. Engelken, 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic Acids Res.*

42, D903–9 (2014).

39. H. Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (Springer Publishing Company, Incorporated, ed. 2nd, 2009).

40. Y. Satta, W. Zheng, K. Nishiyama, R. Iwasaki, T. Hayakawa, F. T. Naoko, Two-dimensional site frequency spectrum for detecting classifying, and dating incomplete selective sweeps. *In Review*.

41. N. Osada, S. Nakagome, S. Mano, Y. Kameoka, I. Takahashi, K. Terao, Finding the factors of reduced genetic diversity on X chromosomes of Macaca fascicularis: male-driven evolution, demography, and natural selection. *Genetics*. **195**, 1027–1035 (2013).

42. S. Nakagome, K. Fukumizu, S. Mano, Kernel approximate Bayesian computation in population genetic inferences. *Stat. Appl. Genet. Mol. Biol.* **12**, 667–678 (2013).

43. R. D. Mashal, J. Koontz, J. Sklar, Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* **9**, 177–183 (1995).

44. H. Thorvaldsdóttir, J. T. Robinson, J. P. Mesirov, Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

45. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).

46. B. Langmead, *Curr. Protoc. Bioinformatics*, in press.

47. 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**, 2078–2079 (2009).

48. P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. **32**, 3047–3048 (2016).

49. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

50. D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018. *Nucleic Acids Res.* (2017), doi:10.1093/nar/gkx1098.

51. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq

quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

52. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, W. Huber, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*. **21**, 3439–3440 (2005).

53. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

54. S. Ge, D. Jung, ShinyGO: a graphical enrichment tool for animals and plants. *bioRxiv*, 315150 (2018).

55. V. del Solar, D. Y. Lizardo, N. Li, J. J. Hurst, C. J. Brais, G. E. Atilla-Gokcumen, Differential Regulation of Specific Sphingolipids in Colon Cancer Cells during Staurosporine-Induced Apoptosis. *Chem. Biol.* **22**, 1662–1670 (2015).

56. A. Saghatelian, S. A. Trauger, E. J. Want, E. G. Hawkins, G. Siuzdak, B. F. Cravatt, Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry*. **43**, 14332–14339 (2004).

57. M. K. Tuck, D. W. Chan, D. Chia, A. K. Godwin, W. E. Grizzle, K. E. Krueger, W. Rom, M. Sanda, L. Sorbara, S. Stass, W. Wang, D. E. Brenner, Standard operating procedures for serum and plasma collection: early detection research network consensus statement standard operating procedure integration working group. *J. Proteome Res.* **8**, 113–117 (2009).