The role of isoforms in the evolution of cryptic coloration in *Peromyscus* mice

Ricardo Mallarino[1]*, Tess A. Linden[1]*, Catherine R. Linnen[2], and Hopi E. Hoekstra[1§]

[1] Howard Hughes Medical Institute and Departments of Organismic & Evolutionary Biology and Molecular & Cellular Biology, Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge MA 02138 USA

[2] Department of Biology, University of Kentucky, 675 Rose Street, Lexington, KY 40506 USA

* Equal contribution

[§] Author for correspondence: hoekstra@oeb.harvard.edu

Abstract

A central goal of evolutionary biology is to understand the molecular mechanisms underlying phenotypic adaptation. While the contribution of protein-coding and *cis*-regulatory mutations to adaptive traits have been well documented, additional sources of variation—such as the production of alternative RNA transcripts from a single gene, or isoforms—have been understudied. Here, we focus on the pigmentation gene *Agouti*, known to express multiple alternative transcripts, to investigate the role of isoform usage in the evolution of cryptic color phenotypes in deer mice (genus *Peromyscus*). We first characterize the *Agouti* isoforms expressed in the *Peromyscus* skin and find two novel isoforms not previously identified in *Mus*. Next, we show that a locally adapted light-colored population of *P. maniculatus* living on the Nebraska Sand Hills shows an up-regulation of a single *Agouti* isoform, termed 1C, compared to their ancestral dark-colored conspecifics. Using *in vitro* assays, we show that this preference for isoform 1C may be driven by isoform-specific differences in translation. In addition, using an admixed population of wild-caught mice, we find that variation in overall *Agouti* expression maps to a region near exon 1C, which also has patterns of nucleotide variation consistent with strong positive selection. Finally, we show that the independent evolution of cryptic light pigmentation in a different species, *P. polionotus*, has been driven by a preference for the same *Agouti* isoform. Together, these findings present an example of the role of alternative transcript processing in adaptation and demonstrate molecular convergence at the level of isoform regulation.

Introduction

Understanding the molecular basis of adaptation is one of the principal goals of evolutionary biology, and considerable efforts have been focused on the contributions of protein-coding and *cis*-regulatory mutations, and their relative importance, to adaptive variation (Carroll 2005; Hoekstra & Coyne 2007; Stern & Orgogozo 2008). In contrast, the production of multiple isoforms through the inclusion of different exons in mRNA—known as alternative mRNA processing— is a major source of genetic variation whose role in phenotypic adaptation has been comparatively understudied. A single protein-coding gene can produce multiple isoforms either through alternative splicing, which results in transcripts with different combinations of exons, or through the use of alternative transcription initiation or termination sites, which generate mRNAs that differ at the 5' or 3' untranslated regions (UTRs). Alternative processing is ubiquitous throughout eukaryotic evolution but is more prevalent in higher eukaryotes than in lower ones, with mammals having the highest genome-wide rate of alternative splicing events (Barbosa-Morais *et al.* 2012; Merkin *et al.* 2012) and alternative promoter usage (Landry *et al.* 2003; Baek *et al.* 2007; Shabalina *et al.* 2014). In addition, recent studies of mammalian gene expression show that transcripts from the majority of protein-coding genes, especially in primates, undergo alternative processing and generate different isoforms (Blencowe 2006; Kim *et al.* 2008). Thus, by increasing transcriptomic and hence proteomic diversity, the production of multiple isoforms through alternative processing has been widely regarded as a key mechanism for generating phenotypic diversity and organismic complexity, particularly at large taxonomic scales (Barbosa-Morais *et al.* 2012; Merkin *et al.* 2012).

To investigate the role of alternative processing in adaptation between closely related species, we studied the pigmentation locus *Agouti*, which has been studied as a model for isoform

regulation in development (Vrieling *et al.* 1994) and has been repeatedly implicated in the evolution of adaptive pigmentation (Rieder *et al.* 2001; Nadeau *et al.* 2008). *Agouti* encodes a secreted paracrine factor that induces pigment-producing cells (melanocytes) in hair follicles to switch from the synthesis of black pigment (eumelanin) to yellow pigment (phaeomelanin) during hair growth (Jackson 1994). In the laboratory mouse, *Mus musculus*, the *Agouti* locus comprises three constitutively transcribed coding exons and four upstream, alternatively transcribed 5' UTRs, also called non-coding exons (Vrieling *et al.* 1994). Thus, in *M. musculus*, *Agouti* mRNAs are found as four different isoforms, each containing one to two non-coding exons upstream of the coding sequence. The "ventral-specific" non-coding exons 1A and 1A' are expressed in the ventral mesenchyme during embryonic development, whereas the "hair-cycle specific" non-coding exons 1B and 1C are expressed during hair growth across all regions of the body (Vrieling *et al.* 1994).

Deer mice (genus *Peromyscus*) populations vary tremendously in both color and pattern. Across populations, there is often a close correspondence between the color of mouse fur and the local soil, suggesting that color-matching is important for survival (Dice 1940; 1941; Haldane 1948). In both Nebraska and Florida, dark-colored mice have colonized extreme light substrate environments that appeared in the last 10,000 years — the Sand Hills in Nebraska (Ahlbrandt & Fryberger 1980; Loope & Swinehart 2000) and the coastal islands in Florida (Campbell 1985; Stapor & Mathews 1991). In both cases, mice have independently evolved significantly lighter coats than mice inhabiting darker surrounding substrates (Fig. 1A and 1B), and there is experimental evidence that visually hunting predators generate strong selection favoring substrate matching (Dice 1941; Vignieri *et al.* 2010; Linnen *et al.* 2013).

In addition to phenotypic convergence, the same gene, *Agouti*, has been shown to play a significant role in mediating this color adaptation in both Nebraska and Florida populations (Hoekstra 2006; Hoekstra *et al.* 2006; Linnen *et al.* 2009; Manceau *et al.* 2011; Linnen *et al.* 2013). Specifically, in light-colored *P. maniculatus* inhabiting the Nebraska Sand Hills, referred to as "wideband" mice, *Agouti* is expressed at higher levels than in their ancestral dark-colored counterparts ("wild-type" mice). Moreover, measurements in neonatal mice showed that *Agouti* mRNA in wideband pups is expressed both at higher levels and for longer periods than in wild-type pups, causing an increase in the width of the pheomelanin band of individual hairs as the hairs grow, which is largely responsible for their overall lighter color (Linnen *et al.* 2009). Furthermore, wideband mice display marked differences in additional pigmentation traits (e.g., dorso-ventral boundary, ventral color, and tail stripe) that are each significantly associated with various regions in the *Agouti* locus (Linnen *et al.* 2013). In Florida, changes in *Agouti* and two other pigmentation loci are responsible for producing the lighter pigmentation phenotypes displayed by a derived population of beach mice (*P. p. leucocephalus)* inhabiting the coastal sand dunes, relative to their ancestral mainland conspecifics (*P. p. subgriseus*) (Steiner *et al.* 2007; Mullen & Hoekstra 2008). Like wideband mice from the Nebraska Sand Hills, Florida beach mice express *Agouti* at higher levels and in extended spatial domains compared to mainland mice (Manceau *et al.* 2011).

The implication of *Agouti* in the repeated evolution of cryptic coloration in *Peromyscus*, coupled with its regulatory architecture of alternative 5' non-coding exons expressed in a region- and temporal-specific pattern, make this an excellent system in which to study the role of isoform regulation in evolution and adaptation. Here, we first characterize the *Agouti* isoforms present in *Peromyscus* skin and examine their patterns of expression during hair growth. We then use *in*

114 *vitro* experiments to study transcript-specific functional differences. Next, we use a
115 phenotypically variable population in the Sand Hills to map variation in overall *Agouti* mRNA
116 expression to the *Agouti* locus and to examine patterns of nucleotide variation for signatures of
117 selection on the different isoforms. Finally, we study the expression of *Agouti* isoforms in Florida
118 *P. polionotus* to establish whether the changes in isoform regulation seen in Nebraska *P.*
119 *maniculatus* are shared with its sister species that independently evolved light coloration as an
120 adaptation to the local environment.
121
122 **Materials and Methods**
123 *Mice*
124 Lab mice strains: We originally purchased wild-derived strains from the *Peromyscus* Genetic
125 Stock Center (University of South Carolina) and now maintain them at Harvard University. As
126 representatives of the mice found in Nebraska, we used *P. maniculatus* wild-type (a^{+}/a^{+}) and *P.*
127 *maniculatus* heterozygous for the wideband allele and a non-agouti allele (a^{wb}/a^{-}) (Linnen *et al.*
128 2009; 2013). The non-agouti allele has a 125-kb deletion that removes the entire regulatory
129 region and first coding exon, resulting in a complete loss of expression (Kingsley *et al*. 2009);
130 thus, *P. maniculatus* a^{wb}/a^{-} are effectively hemizygous for the wideband allele. For many of the
131 experiments, however, we also generated wideband mice homozygous for the wideband allele
132 (a^{wb}/a^{wb}). As representatives of mice in Florida, we used *P. polionotus subgriseus* mainland mice
133 (homozygous for the *Agouti* Dark allele [*Agouti DD*]) and *P. p. subgriseus* mainland mice
134 homozygous for the *Agouti* Light beach allele (*Agouti LL*), hereafter referred to as *Agouti LL*
135 mice (Manceau *et al*. 2011).
136

137 Wild-caught mice: Mice used for the association and selection studies were collected as described
138 in (Linnen *et al.* 2009) from two Sand Hills sites located <15km apart in Cherry County,
139 Nebraska (site 1: Schlagel Creek Wildlife Management Area, N=62; site 2: Ballard's Marsh
140 Wildlife Management Area, N=29).
141
142 *Quantification of phenotype and soil coloration*
143 We prepared flat skins of mice from the strains mentioned above using standard museum
144 protocols. In addition, we included flat skins from *P. p. leucocephalus* beach mice deposited in
145 the Museum of Comparative Zoology at Harvard University. For all the specimens, we quantified
146 dorsal coloration using a USB4000 spectrophotometer and a PX-2 pulsed xenon light source, and
147 recorded readings using the program SpectraSuite (Ocean Optics). Using a reflectance probe with
148 a shield cut at a 45° angle (to minimize diffuse reflection), we took and averaged three
149 measurements of the dorsal skin of each mouse. Reflectance readings were limited to the 300-700
150 nm range, which represents the visible spectrum of most visual predators (Bennet & Lamoreux
151 2003; Mullen & Hoekstra 2008). We then generated color variables using the program CLRvars
152 (Montgomerie R, 2008, CLR, version 1.05; available at
153 http://post.queensu.ca/~mont/color/analyze.html) and performed principle component analysis in
154 R (R: A Language and Environment for Statistical Computing, R Core Team, 2014,
155 http://www.R-project.org/) using prcomp.
156 To quantify variation in soil color, we collected soil samples from four different localities
157 representing the natural habitats of the mice used in this study: Schlagel Creek Wildlife
158 Management Area, Cherry County, Nebraska (*P. maniculatus* wideband mice); Sparks, Cherry
159 County, Nebraska (*P. maniculatus* wild-type mice); Santa Rosa Island, Okaloosa County, Florida

160 (*P. p. leucocephalus* beach mice); and Graceville, Jackson County, Florida (*P. p. subgriseus*
161 mainland mice). We quantified soil reflectance with a spectrophotometer as described above.
162
163 *Rapid Amplification of cDNA Ends (RACE)*
164 We sampled dorsal and ventral skin of *P. maniculatus* wideband (a^{wb}/a^{-}) and wild-type
165 strains and used 5' rapid amplification of cDNA ends (RACE) to identify the *Agouti* isoforms
166 present in *Peromyscus*. To explore and characterize the full diversity of *Agouti* isoforms, we
167 sampled tissue from pups (postnatal day 4 [P4]), as previously done in *Mus* (Vrieling *et al.* 1994),
168 as well as from adults (> 60 days), since the pelage in adult *Peromyscus* differs from juveniles
169 (Fig. S1, supporting information; Golley *et al.* 1966) and therefore may contain additional
170 transcripts not found during the initial hair cycle.
171 For each sample, we extracted RNA from skin taken from 1-4 individuals. We dissected
172 skin immediately after sacrifice and stored it in RNAlater (Qiagen) at 4 ^{0}C. We extracted total
173 RNA using the Qiagen RNeasy Fibrous Tissue Kit. To maximize tissue disruption and RNA
174 yield, we performed two 2-min homogenizations at 50Hz using a 5mm steel ball and a
175 Tissuelyser LT (Qiagen). Following extraction, we quantified RNA using a Quant-it RNA kit and
176 a Qubit flourometer (Invitrogen). From total RNA, we then purified mRNA using NucleoTrap
177 mRNA kits (Clontech Laboratories, Inc) and quantified mRNA using fluorescence as described
178 above.
179 Following mRNA purification, we used the SMARTer RACE cDNA Amplification Kit
180 (Clontech) to prepare 5'-RACE ready cDNA and perform RACE PCRs. To increase the
181 specificity of the 5' RACE PCRs, we performed a nested PCR. For the first RACE PCR, we used
182 the kit-provided UPM primer in conjunction with a primer located in the exon 4 of *Agouti*

183 ("GSP1" sequence: GTTGAGTACGCGGCAGGAGCAGACG). For the nested PCR, we used
184 the kit-provided NUPA primer in conjunction with a primer located upstream of GSP1
185 ("NGSP1C", sequence: TCTTCTTCAGTGCCACAATAGAAACAG). Following the second
186 PCR, we gel-purified any obvious bands using a Qiaquick gel extraction kit (Qiagen). We then
187 ligated the resulting DNA to vectors and transformed competent cells using the pGEMt easy
188 Vector kit (Promega).
189 Following transformation, we performed colony PCRs using Qiagen TAQ and M13F and
190 M13R primers. PCR products were purified enzymatically using Exonuclease I and Shrimp
191 Alkaline Phosphatase (USB). Purified PCR products were sequenced on an ABI 3730xl Genetic
192 Analyzer at Harvard University's Genomics Core facility. In total, we obtained sequences from
193 24-89 *Agouti* clones from each of eight distinct genotype/stage/tissue combinations.
194
195 *Quantitative PCR (qPCR)*
196 Lab strains: We performed gene expression analyses in *P. maniculatus* wideband (a^{wb}/a^{wb}), *P.*
197 *maniculatus* wild-type, *P. polionotus subgriseus*, and *Agouti LL* mice. We used samples from
198 both pup and adult stages. We extracted RNA from dorsal and ventral tissue as indicated above,
199 isolated mRNA from total RNA using the NucleoTrap mRNA Kit (Clontech), and directly
200 synthesized cDNA using qScript cDNA SuperMix (Quanta BioSciences). We performed all
201 reactions in triplicate using PerfeCta SYBR Green FastMix (Quanta BioSciences) and calculated
202 relative expression between samples using the $2^{-\Delta\Delta C_T}$ method (Livak & Schmittgen 2001) using
203 β-actin as a reference gene.
204 To measure levels of total *Agouti*, we used the primers Agouti-Exon2-F and Agouti-
205 Exon3-R to amplify all transcripts containing the *Agouti* coding sequence. To measure levels of

206 individual isoforms, we paired a forward primer within the noncoding exon (1C-F, 1D-F, or 1E-
207 F, respectively) with a reverse primer in the coding region (Agouti-Exon2-R). We amplified *β*-
208 actin using the primers *β*-actin-Pero-F and *β*-actin-Pero-R. Primer sequences can be found in
209 Supplementary Table 1.
210
211 Wild-caught mice: We removed a 5mm skin biopsy from the dorsum, preserved tissue in
212 RNAlater, and extracted RNA as indicated above. We synthesized cDNA and performed qPCR
213 reactions to measure total *Agouti* and *β*-actin as described for lab strains.
214
215 *Dorsal depilation*
216 It is well established that *Agouti* expression is tightly linked to hair growth (Vrieling *et al.*
217 1994). Contrary to newborn pups, in which hair emerges simultaneously across the body, hair
218 follicles of adult mice are desynchronized with respect to the hair cycle (Muller-Rover & Paus
219 2001). Thus, the quantification of *Agouti* expression in adult dorsal skin represents an average
220 expression level across many follicles that are each at different stages in the cycle. To understand
221 the detailed dynamics of *Agouti* isoform expression in adults, we depilated the backs of adult
222 wild-type and wideband mice, a procedure that resets and synchronizes the hair follicle program
223 (Paus & Cotsarelis 1999) and examined isoform expression at different time points. We
224 anesthetized adult *P. maniculatus* wideband and wild-type mice by performing intraperitoneal
225 injections with a cocktail of Ketamin/Xylazine (0.1mL/20g mouse wt) and depilated a small
226 patch (~1cm^2) in the dorsum using a melted beeswax/resin mixture. After the procedure, we
227 applied topical Lidocaine for pain relief every hour for the next eight hours.
228

229 *mRNA stability assays*

230 Our RACE experiment revealed that the dorsal skin in *Peromyscus* expresses three *Agouti*
231 isoforms (1C, 1D, and 1E) simultaneously, differing only in the first, non-coding exon. To
232 determine whether there were differences in mRNA half-life, we cloned each of the three
233 isoforms from *P. maniculatus* wild-type into a pHAGE-CMV-eGFP-W expression vector such
234 that the noncoding exon and *Agouti* coding sequence replaced the eGFP coding sequence. We
235 used Lipofectamine (Life Technologies) to transfect plasmids into human embryonic kidney
236 (HEK293) cells, which lack endogenous *Agouti* expression, and forty hours after transfection
237 halted transcription by treating cells with 10 μg/ml actinomycin D (Sigma) in DMSO. As a
238 control, we treated cells with an equal volume of 100% DMSO. At zero, two, or four hours after
239 actinomycin treatment, we collected cells in TRIzol reagent (Invitrogen) and extracted RNA
240 using the Direct-zol RNA Mini Kit (Zymo Research). We then performed qPCR as described
241 above, using primers common to all constructs (i.e., located in exons 2 and 3): Agouti-Exon2-F
242 and Agouti-Exon3-R to amplify *Agouti* and *β*-actin-Human-F and -R to amplify endogenous *β*-
243 actin. Primer sequences can be found in Supplementary Table 1. Three replicates were used for
244 each condition and time point. *Agouti* expression in actinomycin-treated cells at each time point
245 was calculated as a percentage of *Agouti* expression in DMSO-treated cells at the same time point
246 using the $2^{-\Delta\Delta C}{}_T$ method.
247
248 *Luciferase assays*
249 To establish whether the non-coding exons of the isoforms expressed in the dorsal skin of
250 *Peromyscus* showed differences in translation, we generated luciferase reporter plasmids by
251 cloning each of the three noncoding exons (1C, 1D, or 1E) from both *P. maniculatus* wideband

and wild-type strains into a 5' UTR reporter vector (pLightSwitch_5UTR; Switchgear Genomics) upstream of the luciferase coding sequence and downstream of the ACTB promoter. We generated mutated 1D reporter plasmids from the wild-type 1D plasmid using the Q5 Site-Directed Mutagenesis Kit (New England BioLabs). We transfected plasmids into human embryonic kidney (HEK293) cells using Lipofectamine (Invitrogen), with six replicate transfections per construct. Forty-eight hours after transfection, we measured luminescence as a readout of protein production using a microplate reader (SpectraMax L). We quantified transcription from each plasmid as follows: we collected six replicates of transfected cells in TRIzol reagent (Invitrogen) and extracted RNA using the Direct-zol RNA Mini Kit (Zymo Research). We then carried out qPCR as described above using the primer pairs Luciferase-F and -R, and β-actin-Human-F and –R. Primer sequences can be found in Supplementary Table 1.

Association tests

Previous work has shown that different color traits, all of which are perfectly correlated with *Agouti* genotype in the lab, map to distinct regions in the *Agouti* locus (Linnen *et al*. 2013). Similarly, any associations between *Agouti* genotype, *Agouti* expression level, and color phenotype in our laboratory mice could either be due to direct functional connections (i.e., a casual variant leads to a lighter coat via increased expression) or to genetic linkage. To determine whether color phenotypes and *Agouti* expression co-localize to the same *Agouti* regions, we measured *Agouti* expression (overall *Agouti* and isoform 1C) as described above in 91 mice that had been genotyped and phenotyped for color traits as described in Linnen *et al.* 2013. After discarding three samples with degraded RNA, our final sample size for the expression association analysis was 88 mice. To test for an association between *Agouti* genotypes and *Agouti*

275 expression, we performed single-SNP linear regressions in PLINK v1.07 (Purcell *et al.* 2007). To
276 control for population structure, we included four significant genetic principal components,
277 estimated from genome-wide markers, as covariates in the analysis. Prior to analysis, we
278 performed a normal-quantile transformation on the expression data in R (Guan & Stephens
279 2011). To correct for multiple testing, we used the step-up method for controlling the False
280 Discovery Rate, which we set to a threshold of 10% (Benjamini & Hochberg 1995).
281
282 **Results**
283 ***Peromyscus* expresses two *Agouti* isoforms not previously identified in *Mus***
284 RACE from *P. maniculatus* ventral samples revealed the presence of the ventral specific
285 isoforms 1A, 1A', and 1A1A' (an isoform containing both the 1A and 1A' exons), and the hair-
286 cycle specific 1C, similar to what is seen in *M. musculus* (Vrieling *et al.* 1994). However, none of
287 our clones contained sequences corresponding to exon 1B found in *M. musculus* (Vrieling *et al.*
288 1994) (Fig. 2A and Fig. S2, supporting information). Our RACE experiment in dorsal skin
289 revealed that, in addition to hair-cycle specific isoform 1C, *P. maniculatus* expressed two
290 additional isoforms of *Agouti*, hereafter referred to as 1D and 1E, which had not been previously
291 reported in *M. musculus* (Fig. 2A,B and Fig. S2, supporting information). Like the previously
292 described *Agouti* isoforms, 1D and 1E each consist of an alternate non-coding exon (exon 1D or
293 1E), followed by three coding exons (exons 2, 3, and 4), which are shared by all *Agouti* isoforms.
294 Both exons 1D and 1E are located downstream of exons 1A and 1C in the *Agouti* locus (Fig. 2B)
295 and show polymorphisms between wideband and wild-type *P. maniculatus* (Fig. S2, supporting
296 information). To confirm the absence of isoform 1B in ventral/dorsal tissue and of isoforms 1D
297 and 1E in ventral tissue, as indicated by our RACE experiments, we carried out qPCR using

298 isoform-specific primers in the respective tissues and did not detect any expression of these
299 transcripts. Together, our results indicate that *P. maniculatus* expresses some, but not all, of the
300 *Agouti* isoforms previously described in *M. musculus* (the ventral-specific 1A, 1A', 1A1A' and
301 the hair cycle specific 1C), and also contains two novel dorsal-specific isoforms that had not been
302 described previously (1D and 1E).
303
304 **Differences in *Agouti* mRNA levels between wideband and wild-type mice are driven by**
305 **upregulation of isoform 1C**
306 As mentioned above, wideband and wild-type mice differ markedly in their dorsal
307 coloration; therefore, we focused here on analyzing the isoforms present in the dorsum. Our
308 RACE experiments demonstrated that the dorsal skin of *P. maniculatus* expresses at least three
309 different *Agouti* transcripts (1C, 1D, and 1E) simultaneously, but it remained unknown whether
310 all three contribute to the increase in *Agouti* mRNA levels seen in wideband mice relative to
311 wild-type ones (Linnen *et al.* 2009) or whether this difference is driven only by a subset. To
312 answer this question, we used qPCR to measure the relative expression of total *Agouti* mRNA
313 and each of its isoforms in dorsal skin of wideband and wild-type *P. maniculatus*. The expression
314 of overall *Agouti* was 18-fold higher in wideband mice than in wild-type mice ($P = 0.031$, two-
315 tailed *t*-test) (Fig. 3A). We then measured isoform-specific expression and found that isoform 1C
316 was approximately 17-fold higher in wideband than in wild-type mice (Fig. 3B; $P = 0.049$, two-
317 tailed *t*-test). In contrast, we did not observe significant differences in mRNA levels between the
318 two strains in expression of isoforms 1D or 1E ($P = 0.075$ and $P = 0.257$, respectively; two-tailed
319 *t*-tests) (Fig. 3B).

We next quantified total *Agouti* and isoform-specific expression at different time points following dorsal depilation in order to examine the dynamics of *Agouti* isoform expression during hair growth. In both strains, total *Agouti* is expressed initially at low levels, peaks at day 7 after depilation, and then decreases (Fig. 3C), a pattern that mirrors the expression of *Agouti* during the first hair cycle in pups (Linnen *et al.* 2009). Total *Agouti* levels in wideband mice were higher than in wild-type mice at days 3, 7, and 9 after depilation ($P = 0.0011$, $P = 0.0228$, and $P = 0.0006$, respectively, two-tailed *t*-tests) (Fig. 3C). When we measured transcript-specific mRNA levels in wideband and wild-type mice, we found that the expression of isoform 1C closely matched the expression of overall *Agouti*, peaking at day 7, and differing from wild-type mice also at days 3, 7, and 9 ($P = 0.0012$, $P = 0.0109$, and $P = 0.0003$, respectively, two-tailed *t*-tests) (Fig. 3D). In contrast, isoform 1D and 1E did not differ between wideband and wild-type mice at any time points, with one exception: expression of 1E differed at day 5 ($P = 0.0166$, two-tailed *t*-test) (Fig. 3D).

To determine whether the same patterns occur in neonates, we measured isoform-specific mRNA levels at postnatal day 4, which corresponds to the stage in the first hair cycle when *Agouti* expression is highest (Linnen *et al.* 2009). Quantitative PCR confirmed that wideband mice express higher levels of total *Agouti* mRNA relative to wild-type mice ($P = 0.00097$, two-tailed *t*-test) (Fig. S3). In addition, like in adults, isoform 1C expression was significantly higher in wideband mice than in wild-type ($P = 0.00031$, two-tailed *t*-test), whereas there were no significant differences between the two strains in the expression of isoforms 1D or 1E ($P = 0.1575$ and $P = 0.3231$, respectively, two-tailed *t*-tests). Together, the results of our measurements of isoform-specific *Agouti* mRNA levels in adults and pups, both in skin containing hair follicles at multiple stages of the hair cycle and skin with hair follicles that are

343 synchronized, indicate that the marked increase in *Agouti* expression seen in wideband mice,
344 relative to wild-type mice, is primarily driven by upregulation of isoform 1C.
345
346 **_Agouti_ isoforms differ in luciferase production**
347 To investigate functional variation associated with different *Agouti* isoforms, we
348 measured their half-lives *in vitro*. After halting transcription from a vector expressing each
349 isoform, we used qPCR to measure mRNA levels at different time points during the transcript
350 decay that followed and did not detect any statistically significant differences between the half-
351 lives of the three isoforms ($P = 0.123$, one-way ANCOVA) (Fig. S4). Thus, our experiment
352 indicates that the three alternative exons simultaneously expressed in *Peromyscus* dorsal skin do
353 not have measurable effects on the stability of the *Agouti* transcripts.
354 To determine if the isoforms differed in their regulation of translation, we next examined
355 whether each of the alternative exons affects the amount of protein produced from a luciferase
356 transcript. Relative to a control vector expressing a luciferase coding sequence without a 5'UTR,
357 a vector carrying exon 1C showed a marked increase in luciferase activity ($P = 0.039$ [wideband
358 sequence] and $P = 0.00013$ [wild-type sequence], two-tailed *t*-tests), whereas 1D showed a
359 marked decrease ($P = 4.4 \times 10^{-5}$ [wideband sequence] and $P = 0.00022$ [wild-type sequence],
360 two-tailed *t*-tests). Exon 1E from wideband mice was not significantly different from the control
361 ($P = 0.2577$ [wideband sequence], two-tailed *t*-test) whereas exon 1E from wild-type mice
362 showed increased luciferase activity compared to the control ($P = 0.0082$, two-tailed *t*-test) (Fig.
363 4A). Importantly, we found that mRNA levels did not differ between any of the vectors ($P =$
364 0.379, ANOVA), demonstrating that differences in translation, not transcription, are exclusively
365 responsible for the differences in luminescence (Fig. 4B).

366 To further dissect the mechanisms underlying the differences in protein translation
367 observed between the isoforms, we examined their sequences in more detail and found that exon
368 1D contains start codons (ATGs) upstream of the *Agouti* start codon (three in the wild-type 1D
369 sequence and two in the wideband sequence) (Fig. 4C and Fig. S2). Upstream start codons have
370 been shown to decrease translation efficiency by recruiting ribosomes away from the start codon
371 (Kozak 2002; Rosenstiel *et al.* 2007; Song *et al.* 2007; Medenbach *et al.* 2011). To determine
372 whether this can explain the reduced translation of isoform 1D, we mutated the ATG sites to
373 ACG in the *P. maniculatus* wild-type sequence and quantified the relative amount of luciferase
374 produced (Fig. 4C). When we mutated and tested each ATG site individually, we found that
375 mutating site 1 led to a significant increase in luciferase production relative to the wild-type 1D
376 sequence (P=1.268 x 10^{-8}, two-tailed *t*-test), mutating site 2 resulted in no significant difference
377 (P = 0.5966, two-tailed *t*-test), and mutating site 3 led to a significant reduction (P=1.6739 x 10^{-6},
378 two-tailed *t*-test) (Fig. 4C). Thus, the different ATG sites in exon 1D differ in their ability to
379 modulate luciferase production, possibly due to differences in their surrounding sequences
380 (Kozak 1999; 2002). When we mutated the three ATG sites simultaneously, luciferase production
381 from the mutant transcript was significantly higher than that from the wild-type 1D transcript (P
382 = 0.000015, two-tailed *t*-test) and did not differ from the control (P = 0.092, two-tailed *t*-test),
383 indicating that the upstream start codons contained within exon 1D are responsible for the
384 marked decrease in translation from this transcript. Together, these experiments demonstrate that
385 the non-coding exons of the *Agouti* isoforms expressed in the dorsal skin of *P. maniculatus* differ
386 in their regulation of protein translation, with non-coding exon 1C generating the highest level of
387 luciferase, relative to the control, and non-coding exon 1D with its multiple ATG sites generating
388 the lowest.

389

390 **Genetic variation near exon 1C is associated with dorsal color and *Agouti* expression in**
391 **wild-caught mice**

392 To evaluate evidence for exon 1C's contribution to adaptive coat color variation in natural
393 populations, we examined patterns of genotype-color associations (from Linnen *et al.* 2013) and
394 genotype-expression associations (this study) across the 180-kb *Agouti* locus in a phenotypically
395 variable population of *P. maniculatus* (Linnen *et al.* 2009). Here, we focus here on dorsal
396 brightness because we expect that this trait (1) has a large impact on substrate matching, and (2)
397 has the potential to be strongly influenced by the hair-cycle isoform 1C due to *Agouti*'s direct
398 impact on pigment deposition in hairs. For dorsal brightness, we previously identified a peak in
399 association centered directly on exon 1C (Fig. 5A). Intriguingly, we also observed a peak in
400 association with expression at the same location (Fig. 5B). We note, however, that these
401 expression data were generated using an assay that detects all *Agouti* isoforms. Nevertheless,
402 when we evaluated the correlation between genotype and isoform 1C expression using 1C-
403 specific probes, the exon 1C SNPs remained significantly associated (SNP 109,902, $P = 0.0026$;
404 SNP 109,882, $P = 0.011$). Although the lack of polymorphic positions in exon 1C that are derived
405 in wideband mice suggests that the causal mutation is not in exon 1C itself, these association
406 mapping data strongly suggest that there is a causal mutation somewhere in its immediate vicinity
407 that simultaneously increases both expression of isoform 1C and, as a consequence, dorsal
408 brightness.

409

410 **Exon 1C has undergone strong positive selection**

Although we have not yet identified the causal mutation, the results of our association mapping indicate that it should be in strong linkage disequilibrium with variants located near exon 1C (Fig. 5A, 5B). In this way, we can use the association mapping results to define light and dark haplotypes (i.e., those that contain the causal mutation and those that do not). If the light mutation has undergone positive selection in the light Sand Hills habitat, we expect to see signatures of selection on the light, but not dark, haplotypes. To evaluate evidence of selection on light and dark haplotypes, we previously (Linnen et al. 2013) used the composite-likelihood method implemented in Sweepfinder (Nielsen *et al*. 2005), a method that compares, for each location, the likelihood of the data under a selective sweep to the likelihood under no sweep. The significance of the CLR test statistic is then determined via neutral simulations. In our case, neutral simulations were conducted under a demographic model estimated from genome-wide SNPs (as described in Linnen *et al.* 2013). Figure 5C depicts the resulting likelihood surfaces and significance thresholds for light and dark haplotypes across a ~20-kb window centered on exon 1C (Fig. 5A, 5B). This analysis indicates strong evidence of selection on the light, but not dark, haplotypes in this region of the *Agouti* locus. Together with the results of our association mapping, these analyses indicate that a mutation(s) in the immediate vicinity of exon 1C contributes to dorsal color and is currently undergoing strong positive selection (s = 0.14; Linnen *et al.* 2013).

Evolutionary convergence of isoform regulation in *Peromyscus*

Our results show that *Agouti* isoform 1C is specifically upregulated in the light-colored *P. maniculatus* wideband mice from the Nebraska Sand Hills relative to the dark-colored ancestral population. We next investigated whether similar regulatory mechanisms are observed in another

434 population of *Peromyscus* that independently underwent selection for light pigmentation. We
435 examined patterns of isoform expression in *Agouti LL* mice and compared them to mainland *P. p.*
436 *subgriseus* (Manceau *et al.* 2011). *Agouti LL* mice are significantly lighter than mainland mice
437 but not as light as beach mice, which also differ from mainland mice at two additional
438 pigmentation loci (Fig. S5). Quantitative PCR revealed that expression of *Agouti* was
439 approximately two-fold higher in *Agouti LL* mice compared to mainland mice ($P = 0.0184$, two-
440 tailed *t*-test; Fig. 6A). Measurements of *Agouti* isoform-specific mRNA levels revealed that there
441 were no significant differences in the expression of isoforms 1D or 1E between *Agouti LL* and
442 mainland mice ($P = 0.259$ and $P = 0.111$, respectively; two-tailed *t*-tests). In contrast, we found
443 that isoform 1C was significantly upregulated in *Agouti LL* mice compared to mainland mice (P
444 $= 0.0279$, two-tailed *t*-test; Fig. 6A). Together, our measurements of mRNA levels demonstrate
445 that the increase in *Agouti* expression seen in *P. polionotus* homozygous for the *Agouti* allele
446 found in beach mice, relative to those carrying the ancestral allele, is produced primarily by a
447 specific upregulation of isoform 1C, a pattern that matches what we observed in *P. maniculatus*
448 (Fig. 3A).
449
450 **Discussion**
451 The *Agouti* locus, which contains multiple independently regulated transcription start sites
452 and has been linked to pigment variation in *Peromycus* (Steiner *et al.* 2007; Mullen & Hoekstra
453 2008; Linnen *et al.* 2009; 2013), other mammals (e.g., Rieder *et al.* 2001; Voisey *et al.* 2001;
454 Schmutz & Berryere 2007; Seo *et al.* 2007, and other vertebrates (e.g., Nadeau *et al.* 2008),
455 represents an ideal study system to understand the importance of alternative transcript processing
456 in adaptation to new environments. While different isoforms have been well studied in *Mus*

457 *musculus*, in *Peromyscus* we both identify new dorsally expressed isoforms (1D and 1E) as well
458 as the lack of expression of 1B across the body. These differences are consistent with genome-
459 wide surveys of isoform variation, which find rapid evolution of isoform usage between species
460 (Barbosa-Morais *et al.* 2012; Merkin *et al.* 2012).

461 In this study, we also find that although the dorsal skin of *Peromyscus* mice expresses
462 three different *Agouti* isoforms simultaneously, differing only in their first non-coding exon (1C,
463 1D, and 1E), the marked differences in overall *Agouti* expression seen between *P. maniculatus*
464 strains (wideband vs. wild-type) and between *P. polionotus* subspecies (*P. p. leucocephalus* vs.
465 *P. p. subgriseus*) are exclusively driven by one of the isoforms (1C). Thus, populations of *P.*
466 *maniculatus* and *P. polionotus* experiencing selection pressures for light dorsal pigmentation
467 have independently converged not only on the same gene, but also on the specific upregulation of
468 the same isoform. One explanation for this result is that exon 1C has inherent sequence properties
469 that result in large amount of protein (relative to other *Agouti* isoforms), indicating that such
470 convergence in isoform upregulation may be driven by selection for the molecular mechanism
471 promoting the highest amount of *Agouti* protein production. In support of this, we find that in a
472 admixed population in the Sand Hills, dorsal color and *Agouti* gene expression is significantly
473 associated with genetic variation around exon 1C and that this region shows a pattern of strong
474 selection.

475 Given the patterns of phenotypic and gene expression association as well as the signatures
476 of selection we have reported here and elsewhere (Linnen *et al.* 2013), it is likely that at least one
477 causal polymorphism is located somewhere in the vicinity of exon 1C, with the most promising
478 locations within 2.5kb upstream or downstream of the exon (Fig. 5). An important goal of future
479 work is to characterize and functionally test all variants in this region, including indels and low-

480 coverage SNPs that may have been absent in the capture-based genotype data. In the case of *P.*
481 *polionotus* beach mice, despite the fact that selection on pigmentation is strong (*s* = 0.5; Vignieri
482 *et al.* 2010) and *Agouti* is known to be a major contributor to pigment differences (Steiner *et al.*
483 2007; Mullen & Hoekstra 2008), any inferences of positive selection for regions in or around
484 exon 1C are confounded by the unique demographic history of this species, which has
485 experienced severe population bottlenecks associated with colonization events from the mainland
486 to novel habitats in the Gulf Coast (Thornton *et al.* 2007; Domingues *et al.* 2012; Poh *et al.*
487 2014). Thus, it is not possible to evaluate with certainty whether specific regions in the *Agouti*
488 locus show signatures of similar selective pressures between *P. maniculatus* and *P. polionotus.*
489 We find that the dorsal skin of *Peromyscus* expresses two isoforms that are not found in
490 *Mus* or in other species (1D and 1E). However, it is unlikely that they have played a major role in
491 the evolution of light coloration in *Peromyscus* because their expression patterns do not differ
492 between the light and dark-colored strains examined here and do not follow the pulse of *Agouti*
493 expression that occurs during hair growth (Vrieling *et al.* 1994). From a functional perspective,
494 isoform 1D contains sequences that cause a marked repression of protein translation, so it is not
495 surprising that this particular exon does not constitute a target for selection on lighter phenotypes,
496 and does not show differences in expression between *P. maniculatus* strains or *P. polionotus*
497 subspecies. In the case of isoform 1E, our functional experiments suggest that the sequence of
498 exon 1E found in wild-type *P. maniculatus* increases protein translation, whereas the exon 1E
499 sequence found in wideband *P. maniculatus* does not impact translation. The two strains'
500 sequences differ only at the end of the exon, where wild-type *P. maniculatus* has a six base pair
501 deletion (Fig S2, supporting information). Importantly, however, the exon 1E sequence found in
502 an outgroup to the two strains, *P. maniculatus rufinus*, is identical to that found in wideband *P.*

503 *maniculatus*, indicating that the deletion found in the wild-type 1E sequence is in fact derived and
504 arose after the split between wild-type and wideband populations. Thus, in the common ancestor
505 of wild-type and wideband mice, exon 1C—and not exon 1E—would have been the only non-
506 coding exon that promoted increased protein production. The driving forces underlying the
507 evolution and maintenance of isoforms 1D and 1E in *Peromyscus* populations are yet unknown.
508 Since *Agouti's* function is primarily linked to regulating pigment-type switching in
509 melanocytes, changes affecting this gene are less likely to have negative pleiotropic
510 consequences and thus, selection pressures for eliminating particular transcripts from populations
511 may be relaxed. Alternatively, isoforms 1D and 1E could be playing a role in other aspects of
512 *Agouti's* function independent of its interaction with melanocytes that was not uncovered by our
513 experiments, such as secretion and/or transport from the dermal papillae. Examining
514 presence/absence and expression patterns of these isoforms in additional *Peromyscus* species and
515 populations may shed light on some of these possibilities.
516 The findings presented here also bear on the molecular basis of convergent evolution. It
517 has long been a topic of contention in evolutionary biology whether similar phenotypes that
518 evolve independently tend to be generated by similar or different molecular changes (e.g.,
519 Manceau et al. 2010; Stern 2013; Rosenblum *et al*. 2014). Some studies of convergent
520 phenotypes have found that they are controlled by independent mutations at different loci (e.g.,
521 Steiner *et al*. 2009; Weng *et al.* 2010; Kowalko *et al.* 2013) while others have found that similar
522 evolutionary pressures on two populations can result in changes at the same gene (e.g., Protas *et*
523 *al.* 2006; Woods *et al.* 2006; Chan *et al*. 2010; Reed *et al*. 2011), and in some cases, even the
524 same amino acid substitutions (e.g., Zhen *et al.* 2012; van Ditmarsch *et al.* 2013). In the case of
525 convergent pigmentation phenotypes in *Peromyscus*, not only is the same locus targeted, but the

526 same pattern of isoform regulatory change has occurred—highlighting the various ways in which
527 convergent evolution can occur at the molecular level.

528 Our results add an additional layer to the known mechanisms by which *Agouti* can play a
529 role in the evolution of pigmentation phenotypes in *Peromyscus*. *Cis*-regulatory changes in
530 *Agouti* are known to contribute to both the wideband phenotype in *P. maniculatus* (Linnen *et al.*
531 2009; 2013) and the beach mouse phenotype in *P. polionotus* (Steiner *et al.* 2007). In addition, an
532 amino acid change in the *Agouti* coding sequence of *P. maniculatus* wideband mice is strongly
533 associated with light phenotypes and shows strong signatures of selection (Linnen *et al.* 2009;
534 2013). Here, we find that in addition to these changes, differences in *Agouti* isoform regulation
535 have also been involved in the evolution of adaptive pigmentation variation in this genus. It is
536 clear that alternative transcript processing provides a virtually limitless substrate for the
537 generation of functional and structural transcriptomic and proteomic diversity, and genomic
538 studies analyzing rates of alternative splicing and alternative promoter usage have revealed the
539 importance of this mechanism in originating diversity at large taxonomic scales (Barbosa-Morais
540 *et al.* 2012; Merkin *et al.* 2012). Our study, by providing an example that links alternative mRNA
541 processing with adaptation to a known selective pressure in different subspecies within a single
542 genus, highlights the importance of this mechanism as a driver of diversification and adaptation
543 at smaller taxonomic scales as well.

544

545

546 **References**

547 Ahlbrandt TS, Fryberger SG (1980) Eolian deposits in the Nebraska Sand hills. *US Geological*
548 *Survey Professional Paper*, **1120A**, 1–24.

549 Baek D, Davis C, Ewing B, Gordon D, Green P (2007) Characterization and predictive discovery
550 of evolutionarily conserved mammalian alternative promoters. *Genome Research*, **17**, 145–
551 155.
552 Barbosa-Morais NL, Irimia M, Pan Q *et al.* (2012) The evolutionary landscape of alternative
553 splicing in vertebrate species. *Science*, **338**, 1587–1593.
554 Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful
555 approach to multiple testing. *Journal of the Royal Statistical Society Series B*, **57**, 289–300.
556 Bennet D, Lamoreux ML (2003) The color loci of mice – a genetic century. *Pigment Cell*
557 *Research*, **16**, 333–344.
558 Blencowe BJ (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37–47.
559 Campbell KM (1985) *Campbell: Geology of Sarasota County, Florida*. Florida Geological
560 Survey Open File.
561 Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS Biology*, **3**, e245.
562 Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Abser
563 DM, Grimwood, J, Schmutz J, Myers RM, Petrov D, Jonsson B, Schluter D, Bell, MA,
564 Kingsley DM (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent
565 deletion of a *Pitx1* enhancer. *Science*, **15,** 302-305.
566 Dice LR (1940) Ecologic and genetic variability within species of *Peromyscus*. *American*
567 *Naturalist*, **74**, 212–221.
568 Dice LR (1941) Variation of the Deer-mouse *(Peromyscus maniculatus)* on the Sand hills of
569 Nebraska and adjacent areas. Contributions from the Laboratory of Vertebrate Biology, **15,**
570 1-19.
571 Domingues VS, Poh Y-P, Peterson BK *et al.* (2012) Evidence of adaptation from ancestral
572 variation in young populations of beach mice. *Evolution*, **66**, 3209–3223.
573 Golley FB, Morgan EL, Carmon JL (1966) Progression of molt in *Peromyscus polionotus*.
574 *Journal of Mammalogy*, **47**, 145–148.
575 Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association
576 studies and other large-scale problems. *The Annals of Applied Statistics*, **5**, 1780–1815.
577 Haldane J (1948) The theory of a cline. *Journal of Genetics*, **48**, 277–284.
578 Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in

579 vertebrates. *Heredity*, **97**, 222–234.
580 Hoekstra HE, Coyne JA (2007) The locus of evolution: Evo devo and the genetics of adaptation.
581 *Evolution*, **61**, 995–1016.
582 Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP (2006) A single amino acid
583 mutation contributes to adaptive beach mouse color pattern. *Science*, **313**, 101–104.
584 Jackson I (1994) Molecular and developmental genetics of mouse coat color. *Annual Review of*
585 *Genetics*, **28**, 189–217.
586 Kim E, Goren A, Ast G (2008) Alternative splicing: current perspectives. *BioEssays*, **30**, 38–47.
587 Kingsley EP, Manceau M, Wiley CD, Hoekstra HE (2009) Melanism in *Peromyscus* is caused by
588 independent mutations in *Agouti*. *PLoS One*, **4**, e6435.
589 Kowalko JE, Rohner N, Linden TA *et al.* (2013) Convergence in feeding posture occurs through
590 different genetic loci in independently evolved cave populations of *Astyanax mexicanus*.
591 *PNAS*, **110**, 16933–16938.
592 Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187–208.
593 Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene*,
594 **299**, 1–34.
595 Landry J-R, Mager DL, Wilhelm BT (2003) Complex controls: the role of alternative promoters
596 in mammalian genomes. *Trends in Genetics*, **19**, 640–648.
597 Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE (2009) On the origin and spread of an adaptive
598 allele in deer mice. *Science*, **325**, 1095–1098.
599 Linnen CR, Poh Y-P, Peterson BK *et al.* (2013) Adaptive evolution of multiple traits through
600 multiple mutations at a single gene. *Science*, **339**, 1312–1316.
601 Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time
602 quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, **25**, 402–408.
603 Loope DB, Swinehart JB (2000) Thinking like a dune field: Geologic history in the Nebraska
604 Sand Hills. *Great Plains Research*, **10**, 5–35.
605 Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE (2010) Convergence in
606 pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of*
607 *the Royal Society*, **365**, 2439-2450.
608 Manceau M, Domingues VS, Mallarino R, Hoekstra HE (2011) The developmental role of

609 Agouti in color pattern evolution. *Science*, **331**, 1062–1065.
610 Medenbach J, Seiler M, Hentze MW (2011) Translational control via protein-regulated upstream
611 open reading frames. *Cell*, **145**, 902–913.
612 Merkin J, Russell C, Chen P, Burge CB (2012) Evolutionary dynamics of gene and isoform
613 regulation in mammalian tissues. *Science*, **338**, 1593–1599.
614 Mullen LM, Hoekstra HE (2008) Natural selection along an environmental gradient: a classic
615 cline in mouse pigmentation. *Evolution*, **62**, 1555–1570.
616 Muller-Rover S, Paus R (2001) A comprehensive guide for the accurate classification of murine
617 hair follicles in distinct hair cycle stages. *Journal of Investigative Dermatology* **117**, 3–15.
618 Nadeau NJ, Minvielle F, Ito S, Inoue-Murayama M, Gourichon D, Follett SA, Burke T, Mundy
619 NI (2008) Characterization of Japanese quail yellow as a genomic deletion upstream of
620 the avian homologue of the mammalian ASIP (agouti) gene. *Genetics* **178**, 777–786.
621 Nielsen R, Williamson S, Kim Y *et al.* (2005) Genomic scans for selective sweeps using SNP
622 data. *Genome Research*, **15**, 1566–1575.
623 Paus R, Cotsarelis G (1999) The biology of hair follicles. *New England Journal of Medicine*,
624 **341**, 491–497.
625 Poh YP, Domingues VS, Hoekstra HE, Jensen JD (2014) On the prospect of identifying adaptive
626 loci in recently bottlenecked populations. *PLoS One*, **9**, e110579.
627 Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R,
628 Tabin CJ (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution
629 of albinism. *Nature Genetics*, **38**, 107-111.
630 Purcell S, Neale B, Todd-Brown K, Thomas L (2007) PLINK: a tool set for whole-genome
631 association and population-based linkage analyses. *The American Journal of Human*
632 *Genetics*, **81**, 559–575.
633 Reed RW, Papa R, Martin A, Hines HM, Counterman BA, Pardo-Diaz C, Jiggins CD,
634 Chamberlain NL, Kronforst MR, Chen R, Halder G, Nijhout HF, McMillan WO (2011)
635 *Optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*,
636 **333**, 1137-1141.
637 Rieder S, Taourit S, Mariat D, Langlois B, Guérin G (2001) Mutations in the agouti (ASIP), the
638 extension (MC1R), and the brown (TYRP1) loci and their association to coat color

phenotypes in horses (*Equus caballus*). *Mammalian Genome*, **12**, 450–455.

Rosenblum EB, Parent CE, Brandt EE (2014) The molecular basis of phenotypic convergence. *Annual Reviews of Systematics and Ecology*, **45**, 203-226.

Rosenstiel P, Huse K, Franke A *et al.* (2007) Functional characterization of two novel 5' untranslated exons reveals a complex regulation of NOD2 protein expression. *BMC Genomics*, **8**, 472–480.

Schmutz SM, Berryere TG (2007) Genes affecting coat colour and pattern in domestic dogs: a review. *Animal Genetics*, **38**, 539–549.

Seo K, Mohanty TR, Choi T, Hwang I (2007) Biology of epidermal and hair pigmentation in cattle: a mini-review. *Veterinary Dermatology*, **18**, 392–400.

Shabalina SA, Ogurtsov AY, Spiridonov NA, Koonin EV (2014) Evolution at protein ends: major contribution of alternative transcription initiation and termination to the transcriptome and proteome diversity in mammals. *Nucleic Acids Research*, **42**, 7132–7144.

Song KY, Hwang CK, Kim CS *et al.* (2007) Translational repression of mouse mu opioid receptor expression via leaky scanning. *Nucleic Acids Research*, **35**, 1501–1513.

Stapor FW Jr, Mathews TD (1991) Barrier-island progradation and Holocene sea-level history in southwest Florida. *Journal of Coastal Research*, **7**, 815–838.

Steiner CC, Weber JN, Hoekstra HE (2007) Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biology*, **5**, e219.

Steiner CC, Rompler H, Boettger LM, Schoneberg T, Hoekstra HE (2009) The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution*, **26**, 35-45.

Stern DL, Orgogozo V (2008) The loci of evolution: how predictable is genetic evolution? *Evolution*, **62**, 2155–2177.

Thornton KR, Jensen JD, Becquet C, Andolfatto P (2007) Progress and prospects in mapping recent selection in the genome. *Heredity*, **98**, 340–348.

van Ditmarsch D, Boyle KE, Sakhtah H *et al.* (2013) Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Reports*, **4**, 697–708.

Vignieri SN, Larson JG, Hoekstra HE (2010) The selective advantage of crypsis in mice. *Evolution*, **64**, 2153–2158.

669 Voisey J, Box NF, van Daal A (2001) A polymorphism study of the human Agouti gene and its
670 association with MC1R. *Pigment Cell Research*, **14**, 264–267.
671 Vrieling H, Duhl DM, Millar SE, Miller KA, Barsh GS (1994) Differences in dorsal and ventral
672 pigmentation result from regional expression of the mouse agouti gene. *PNAS*, **91**, 5667–
673 5671.
674 Weng J-K, Akiyama T, Bonawitz ND *et al.* (2010) Convergent evolution of syringyl lignin
675 biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. *The*
676 *Plant Cell*, **22**,1033–1045.
677 Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular
678 evolution in a long-term experiment with *Escherichia coli*. *PNAS*, **103**, 9107–9112.
679 Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P (2012) Parallel molecular
680 evolution in an herbivore community. *Science*, **337**, 1634–1637.
681

682 **Data accessibility**

683 Museum voucher numbers from specimens used for phenotypic measurements will be provided
684 prior to publication. A file with gene expression values (CT values) for all individuals and
685 genes/isoforms will be provided.

686

687 **Acknowledgements**

688 We thank Judy Chupasko and Mark Omura for help with specimen preparation, and Nikki
689 Hughes for providing logistical support. RM and the molecular work was supported by a Swiss
690 National Science Foundation grant to HEH. TAL was supported by a Herchel Smith-Harvard
691 Undergraduate Research Fellowship. CRL was supported by a Ruth Kirschstein National
692 Research Service Award from NIH. HEH is an Investigator of the Howard Hughes Medical
693 Institute.

694

Author contributions

RM, CRL, and HEH conceived the project and designed the experiments. RM, TAL, and CRL performed experiments and analyzed data. All authors wrote the paper.

Figure legends

Figure 1. Environmental and phenotypic differences between ancestral and derived habitats and mice in Nebraska and Florida. (**A**) Photographs of wild-type *Peromyscus maniculatus bairdii* (WT), wideband *P. m. nebrascensis* (WB), *P. polionotus subgriseus* mainland, and *P. p. leucocephalus* beach mice (top), typical habitats (bottom), and soil (bottom inset). (**B**) Dorsal brightness (total dorsal reflectance) measured in mice from ancestral (black circles) and derived (white circles) habitats. (**C**) Reflectance of soil samples from ancestral (black circles) and derived (white circles) habitats. Soil samples are from Cherry County, Sparks, Nebraska vs. Cherry County, Schlagel, Nebraska; Jackson County, Graceville, Florida vs. Okaloosa County, Santa Rosa Island (SRI) Florida. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-tailed *t*-tests; $n = 4$ (mice) and $n = 5$ (soil).

Figure 2. Characterization of *Agouti* isoform expression in *P. maniculatus*. (**A**) *Agouti* isoforms expressed in the ventrum and dorsum of *M. musculus* pups (strain A^W) and *P. maniculatus* (pups and adults). (**B**) Map of the *Agouti* locus in *M. musculus* (above) and *P. maniculatus* (below). Colors in (A) and (B) represent ventral specific non-coding exons (blue), hair cycle-specific non-coding exons (red), and the novel non-coding exons (green) reported here. Coding exons (black) are common to all isoforms.

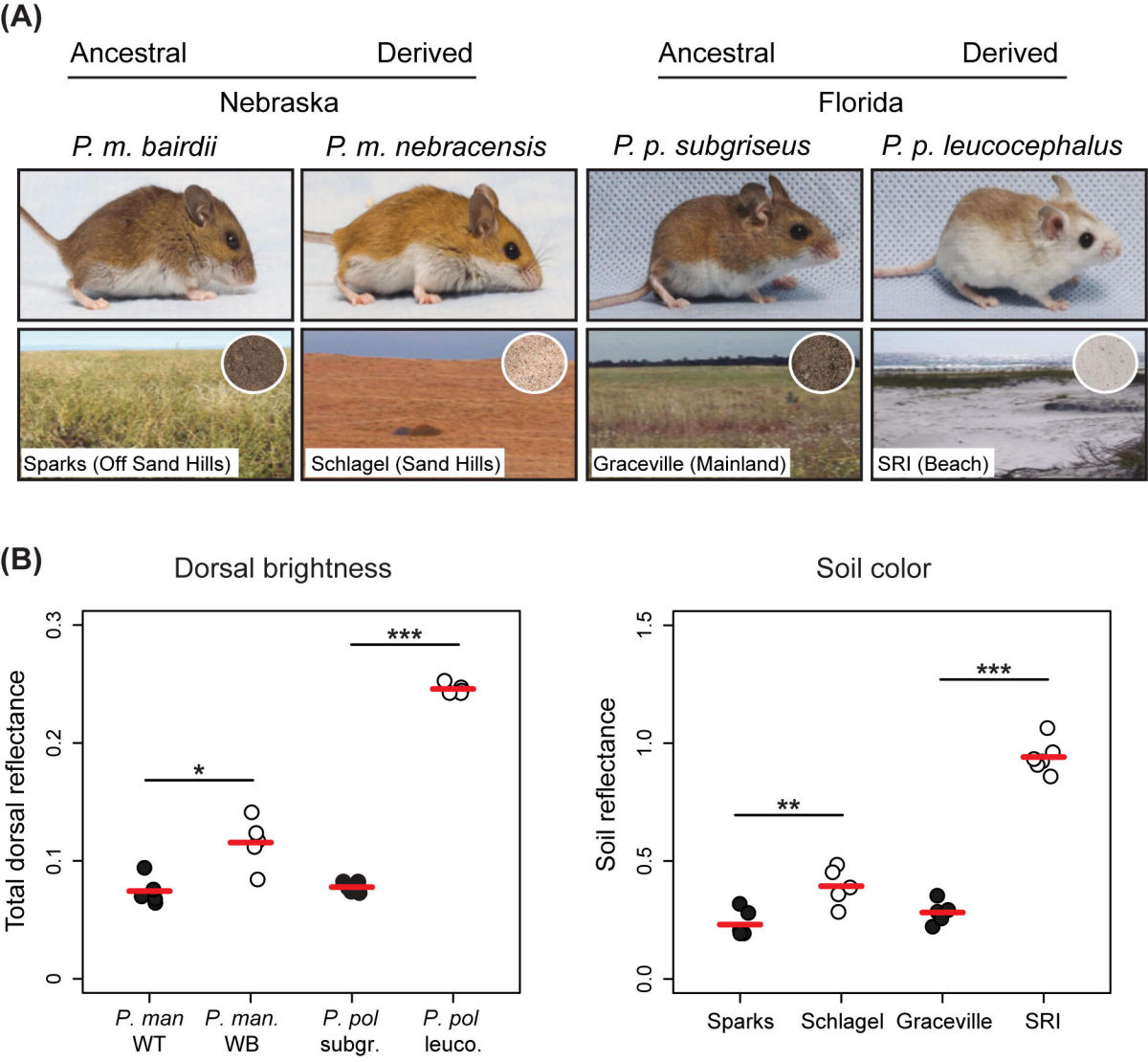
Figure 3. Differential expression of *Agouti* isoforms between wideband and wild-type mice. (**A**) qPCR in adults (> 60 days) shows that the overall expression of *Agouti* mRNA in wideband mice (light circles) is 18-fold higher than wild-type mice (dark circles). (**B**) Isoform-specific mRNA levels of isoform 1C are 17-fold higher in wideband than wild-type mice, but no significant differences in 1D or 1E expression. (**C**) qPCR measurements of overall *Agouti* mRNA levels at different time points after hair removal show that wideband and wild-type mice differ in the expression of total *Agouti* at days 3, 7, and 9 after depilation. (**D**) Isoform specific measurements show that isoform 1C differs between the two strains at days 3, 7, and 9, whereas isoforms 1D and 1E do not differ between the two strains, with the exception of 1E, which differs at day 5. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two-tailed *t*-tests; $n = 4$ (for each strain in (A) and (B)) and $n = 3$ (for each strain and time point in (C) and (D)); red bars indicate the mean.

Figure 4. *Agouti* isoforms differ in luciferase production. (**A**) Relative luciferase levels produced by transcripts carrying the three different 5' UTRs expressed in *P. maniculatus*. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, two tailed *t*-tests; $n = 6$ per construct. (**B**) mRNA levels of luciferase did not differ between the constructs demonstrating that differences seen in (A) occur at the posttranscriptional level; ANOVA, $n = 3$. Only results from wild-type sequences are shown. (**C**) (Above) The sequence of exon 1D (black font) in wild-type *P. maniculatus*. Upstream start codons are boxed in red, the beginning of exon 2 is in green font, and the functional ATG is boxed in blue. (Below) Quantification of relative luciferase levels from transcripts carrying the *P. maniculatus* wild-type exon 1D and those carrying a version of exon 1D where each upstream ATG site has been mutated to ACG, relative to a control lacking a 5'

UTR. * P < 0.05, **P < 0.01, ***P < 0.001, two tailed *t*-tests; *n* = 5 per construct; red bars indicate the mean. In all cases, luciferase levels are normalized relative to background levels.

Figure 5. Genetic variation near exon 1C is associated with dorsal color, *Agouti* expression, and signatures of selection in a natural population of *P. maniculatus*. (**A, B**) Strength of statistical association (-log p-value) between dorsal brightness (A) and *Agouti* expression (B) for 466 SNPs (circles) tested in *n* = 91 (A) or 88 (B) mice. SNPs significant after a 10% FDR correction are indicated in red. The gray bar highlights the location of exon 1C, and the dashed lines indicate a 20-kb region centered on this exon. Color (A) and expression (B) both have multi-SNP peaks of association centered on exon 1C. (**C**) Strength of evidence favoring a selection model over a neutral model (likelihood ratios) as a function of location in *Agouti*. Likelihood surfaces are shown for light (red) and dark (black) haplotypes, as determined by the strongest associated SNP in (A). Dashed lines indicate significance thresholds, determined via neutral simulations, for light (red) and dark (black) haplotypes. Data for (A) and (C) are from Linnen *et al.* 2013.

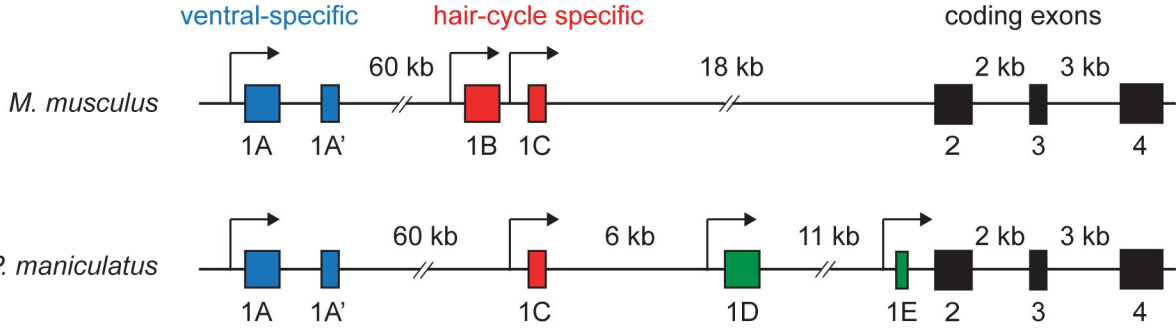
Figure 6. Differential *Agouti* and *Agouti* isoform expression in dark and light *P. polionotus*. (**A**) Mice with the light beach allele (*Agouti LL*; light circles) have two-fold higher expression of *Agouti* than mainland mice (*Agouti DD* mice; dark circles), as determined by qPCR. (**B**) Expression of isoform 1C is four-fold higher in *Agouti LL* mice (light circles) compared to mainland mice (dark circles), whereas there were no significant differences in 1D or 1E expression. * P < 0.05, **P < 0.01, ***P < 0.001, two-tailed *t*-tests; *n* = 4 mice for each strain; red bars indicate the mean.

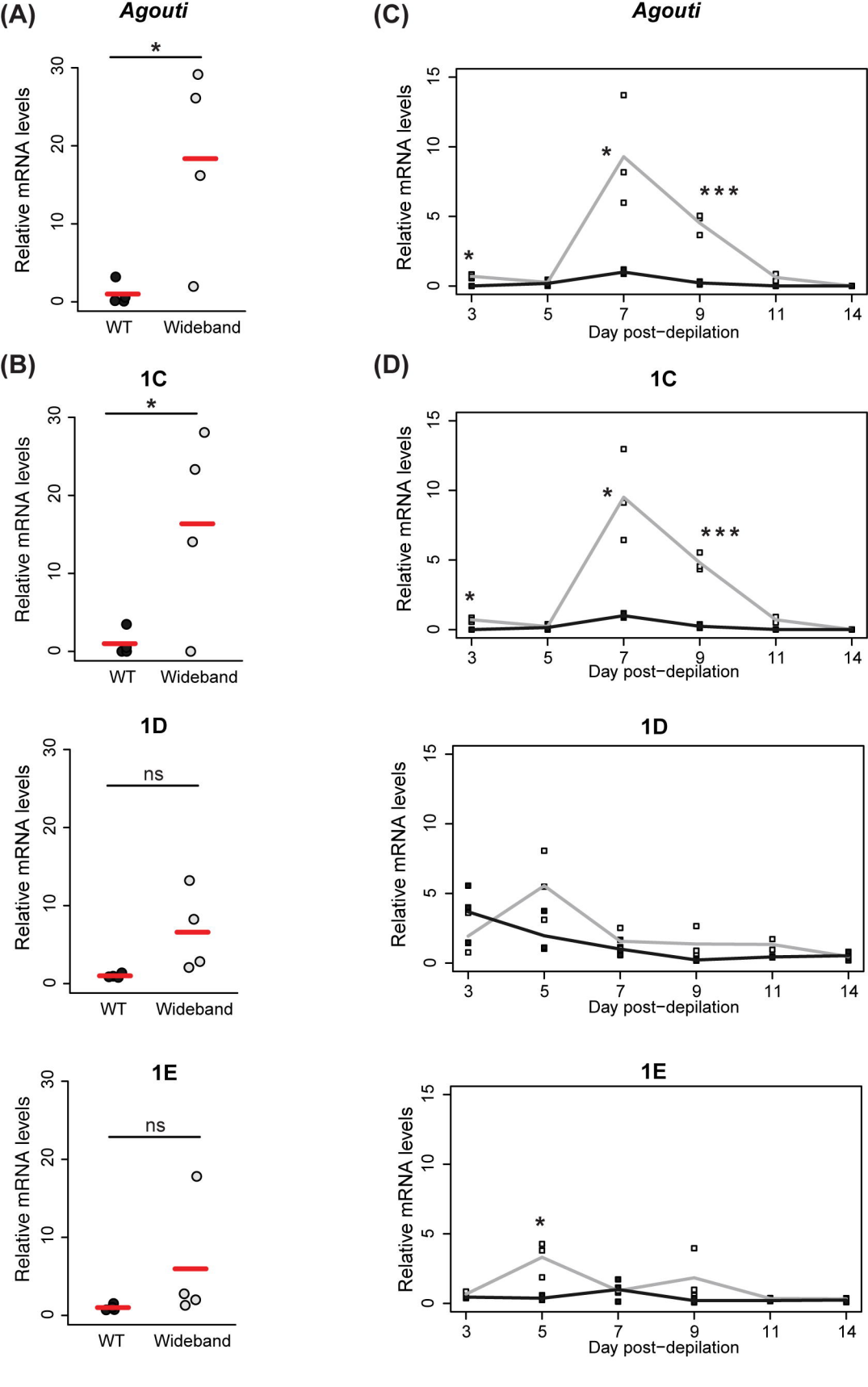

(A)
Ancestral
Derived
Ancestral
Derived
Nebraska
Florida
P. m. bairdii
P. m. nebracensis
P. p. subgriseus
P. p. leucocephalus
Sparks (Off Sand Hills)
Schlagel (Sand Hills)
Graceville (Mainland)
SRI (Beach)
(B)
Dorsal brightness
Total dorsal reflectance
0
0.1
0.2
0.3
*
***
P. man WT
P. man. WB
P. pol subgr.
P. pol leuco.
Soil color
Soil reflectance
0.0
0.5
1.0
1.5
**
***
Sparks
Schlagel
Graceville
SRI

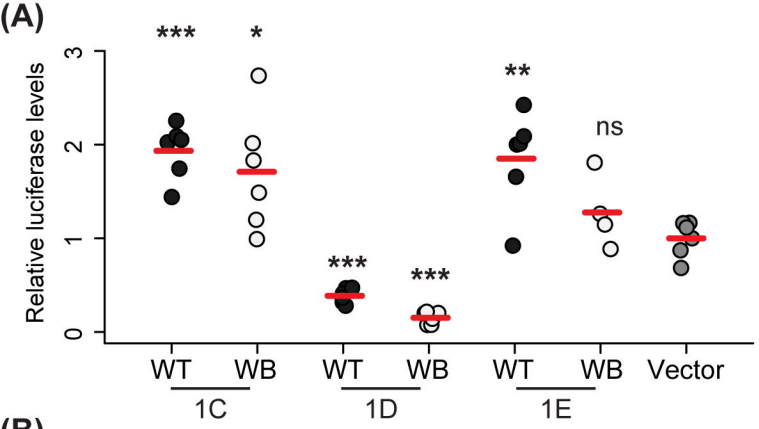
(A)

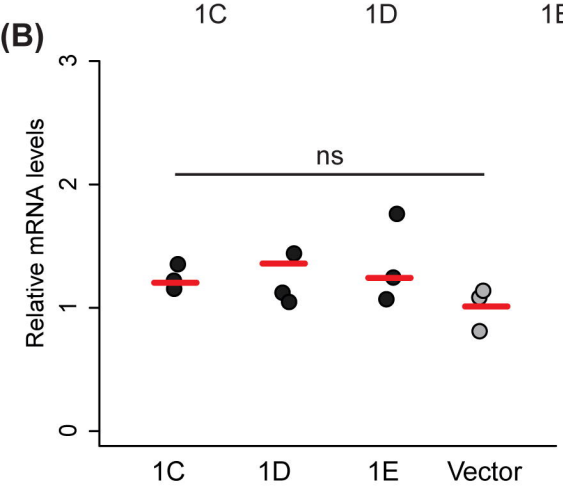
	M. musculus (A^w)	*P. maniculatus*
Dorsum	1B, 1C, 1B1C	1C, 1D, 1E
Ventrum	1A, 1A', 1A1A' 1B, 1C, 1B1C	1A, 1A', 1A1A' 1C

(B)




(A)
Agouti
Relative mRNA levels
WT
Wideband
(B)
1C
(C)
Agouti
Day post−depilation
(D)
1C
1D
ns
1E

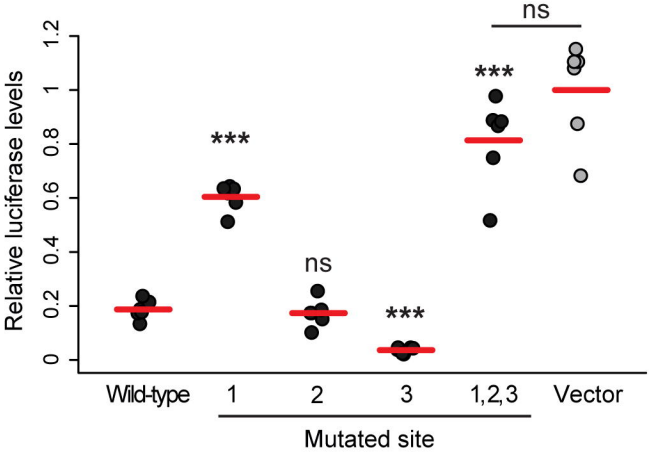


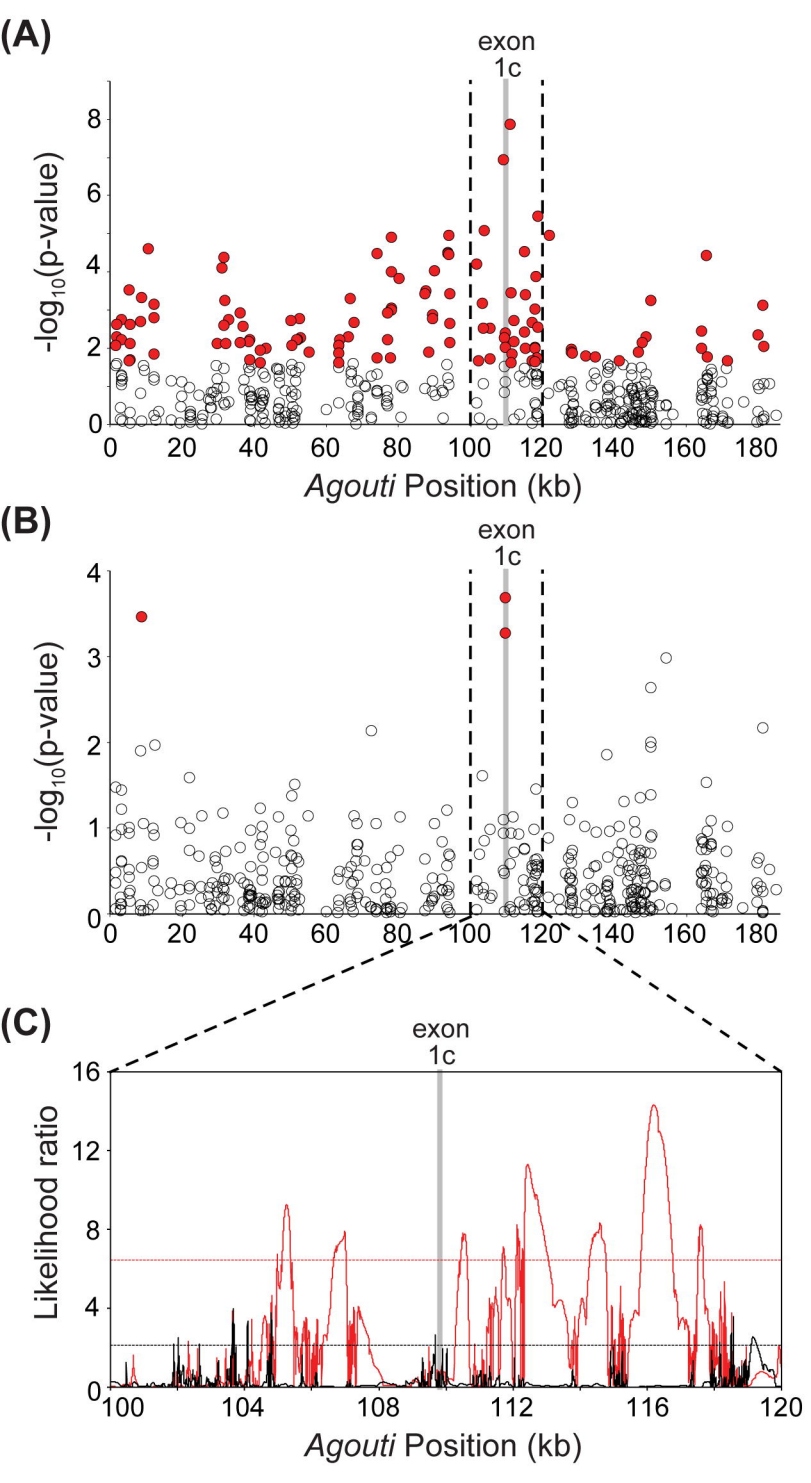


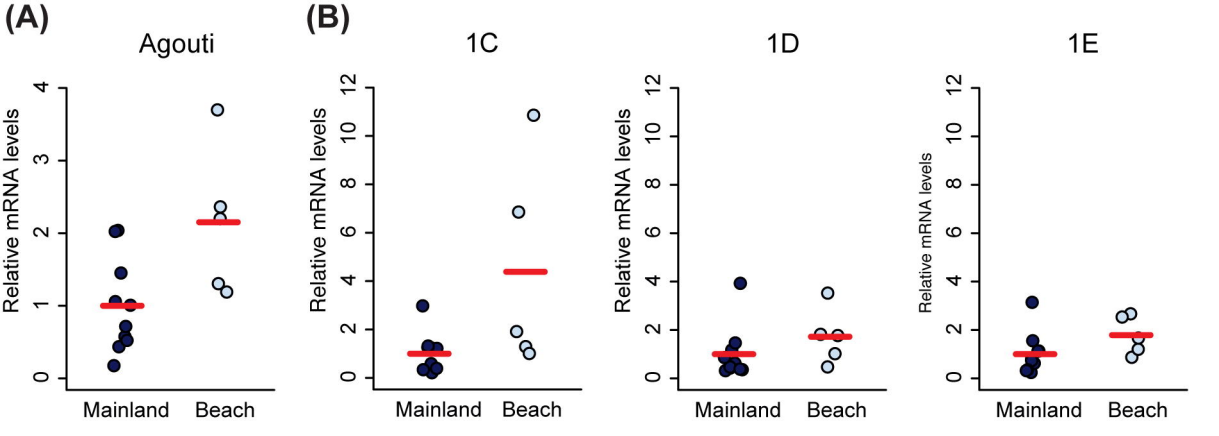
(C)

1D WT

ATA**ATG**GCCTCGCGCCTCGCCTCCCAGGCGCTAGAACTGC**ATG**CA
GGTGCGCAGCATCACTCCTCAGGCCGGTCTTCCTGCTTCCCGTCC
ACTCCTGGGAAAGTGCAGAGCTAGGTCTGGG**ATG**ACTTCTTCCAC
GGCCGTACTTCTGCAGCTTTTCAGGATG




(A)
exon
1c
-log10(p-value)
0
2
4
6
8
0
20
40
60
80
100
120
140
160
180
Agouti Position (kb)
(B)
exon
1c
-log10(p-value)
0
1
2
3
4
0
20
40
60
80
100
120
140
160
180
(C)
exon
1c
Likelihood ratio
0
4
8
12
16
100
104
108
112
116
120
Agouti Position (kb)


(A)
Agouti
Relative mRNA levels
0
1
2
3
4
Mainland
Beach
(B)
1C
Relative mRNA levels
0
2
4
6
8
10
12
Mainland
Beach
1D
Relative mRNA levels
0
2
4
6
8
10
12
Mainland
Beach
1E
Relative mRNA levels
0
2
4
6
8
10
12
Mainland
Beach