nature genetics

A molecular basis for classic blond hair color in Europeans

Catherine A Guenther[1,2], Bosiljka Tasic[2,3,5], Liqun Luo[2,3], Mary A Bedell[4] & David M Kingsley[1,2]

Hair color differences are among the most obvious examples of phenotypic variation in humans. Although genome-wide association studies (GWAS) have implicated multiple loci in human pigment variation, the causative base-pair changes are still largely unknown[1]. Here we dissect a regulatory region of the *KITLG* gene (encoding KIT ligand) that is significantly associated with common blond hair color in northern Europeans[2]. Functional tests demonstrate that the region contains a regulatory enhancer that drives expression in developing hair follicles. This enhancer contains a common SNP (rs12821256) that alters a binding site for the lymphoid enhancer-binding factor 1 (LEF1) transcription factor, reducing LEF1 responsiveness and enhancer activity in cultured human keratinocytes. Mice carrying ancestral or derived variants of the human *KITLG* enhancer exhibit significant differences in hair pigmentation, confirming that altered regulation of an essential growth factor contributes to the classic blond hair phenotype found in northern Europeans.

Humans show striking differences in hair pigmentation within and among populations. Whereas dark hair color predominates in many African, Asian and southern European populations, lighter hair colors are common in northern Europe. Blond hair, which occurs naturally in a small fraction of humans, has had a notable range of both positive and negative associations in human history. In some cultures, light skin and hair color is stigmatized as a ghost-like abnormality, sign of promiscuity or signature of unusual ancestry[3,4]. In contrast, fair hair was associated with youth and beauty in the earliest written works of ancient Greece[5] and has frequently been imitated (or masked) using bleaches, dyes and wigs in both ancient and modern populations[4]. Despite thousands of years of interest in hair color variation, the molecular basis of common human hair color phenotypes is still incompletely understood. Recent genome-wide surveys have shown that genetic variants linked to eight genes are significantly associated with blond hair color in Europeans[2,6–8]. Some blond-associated variants in humans alter the coding regions of genes known to be involved in pigmentation[1,9]. However, many GWAS signals for pigmentation and other human traits map outside the protein-coding regions of genes[1,10], are enriched in likely regulatory sequences[11] and have been difficult to trace to particular DNA base-pair mutations[12]. Knowing the causative nucleotides underlying human traits may improve genetic predictions compared to common linked markers[13] and facilitate comparison of traits and mutations among both past and present populations[14].

Human *KITLG* (mouse *Kitl*) encodes a secreted ligand for the KIT receptor tyrosine kinase and has an essential role in the development, migration and differentiation of many different cell types in the body, including melanocytes, blood cells and germ cells[15]. Null mutations affecting *Kitl* or *Kit* are lethal in mice, and hypomorphic alleles cause white hair, mast cell defects, anemia and sterility[16–18]. A noncoding SNP (rs12821256) located in a large intergenic region over 350 kb upstream of the *KITLG* transcription start site is significantly associated with blond hair color in Iceland and The Netherlands[2] (**Fig. 1a**). This SNP shows relatively large odds ratios of 1.9–2.4 per allele associated with blond versus brown hair in northern Europeans (multiplicative model[2]). Together with variants in other genes, rs12821256 helps explain 3–6% of the variance in categorical hair color scores[2] and is now one of several markers used for predictive testing of human hair color[19]. The blond-associated A>G substitution at this position is prevalent in northern European populations but virtually absent in African and Asian populations[2,20,21] (**Fig. 1b**), suggesting that regulatory changes associated with an essential signaling gene might contribute to common blond hair color in Europe.

Regulatory mutations in mice confirm the importance of distant upstream sequences in the control of hair color. The *Steel* panda mutation (Sl^{pan}), a viable X-ray–induced *Kitl* allele that reduces pigmentation in both heterozygotes and homozygotes[22], results from a large chromosome inversion upstream of *Kitl*[23]. We sequenced the precise molecular breakpoints of this inversion for the first time to our knowledge (Online Methods), showed that the inversion spans 65.6 Mb and confirmed that the position of the breakpoints would displace mouse sequences orthologous to the blond-associated GWAS peak. Mice homozygous for the mutation are white, and heterozygous mice carrying only a single copy of the Sl^{pan} allele are noticeably lighter than control mice (**Fig. 1c**). Quantitative RT-PCR assays showed that heterozygous mice expressed 61 ± 9.1% of wild-type *Kitl* RNA levels in skin (mean ± s.e.m.; $P = 0.0022$), showing that displacement of a single copy of the distant upstream regulatory sequences for *Kitl* is sufficient to reduce *Kitl* expression and lighten hair color.

To identify the base-pair changes responsible for the association with blond hair in humans, we used a transgenic approach to search for functional enhancers throughout the GWAS-identified candidate

[1]Department of Developmental Biology, Stanford University School of Medicine, Stanford, California, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, California, USA. [3]Department of Biology, Stanford University, Stanford, California, USA. [4]Department of Genetics, University of Georgia, Athens, Georgia, USA. [5]Present address: Allen Institute for Brain Science, Seattle, Washington, USA. Correspondence should be addressed to D.M.K. (kingsley@stanford.edu).

Received 30 November 2013; accepted 1 May 2014; published online 1 June 2014; doi:10.1038/ng.2991

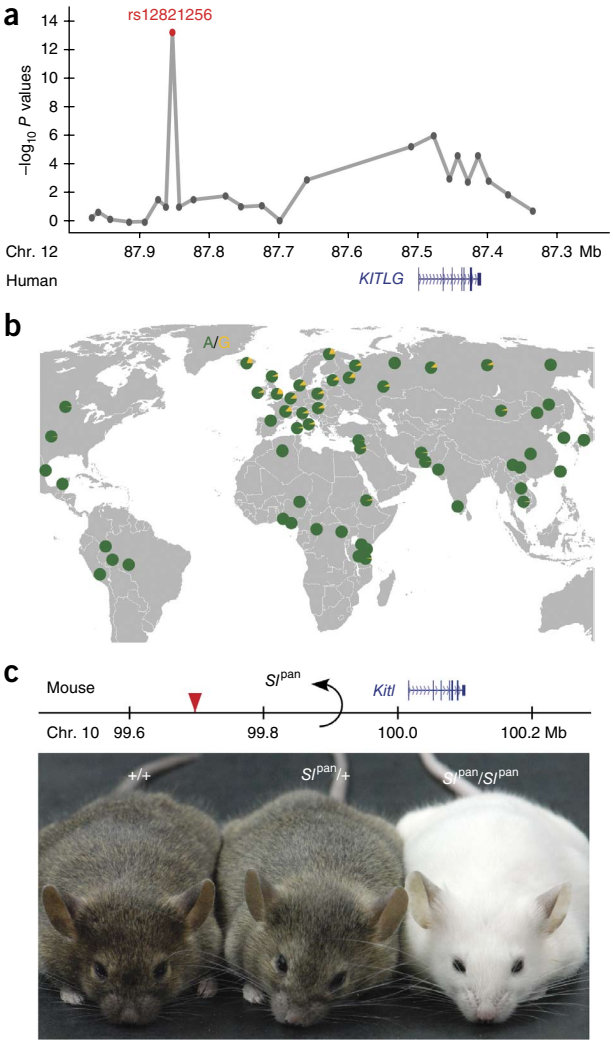

Figure 1 A distant regulatory region upstream of the *KITLG* gene controls hair pigmentation in humans and mice. (**a**) SNPs on human chromosome 12 are associated with blond hair in Europeans (modified from ref. 2). The peak association is found at rs12821256 (red), which is 355 kb upstream of the *KITLG* transcription start site. (**b**) Frequency distribution of rs12821256 in different populations. The G allele associated with blond hair (yellow) is most prevalent in northern Europe. Green color represents the frequency of the ancestral A allele. (**c**) The Sl^{pan} allele at the mouse *Kitl* locus consists of a 65.6-Mb chromosome inversion (GRCm38/mm10, chr. 10 breakpoints: 34.301–99.916 Mb) that displaces upstream sequences orthologous to rs12821256 (red triangle). Mice heterozygous ($Sl^{pan}/+$) and homozygous (Sl^{pan}/Sl^{pan}) for this allele have lighter-colored coats than control (+/+) mice, demonstrating that alteration of even a single copy of the region upstream of *Kitl* can reduce hair pigmentation.

interval surrounding rs12821256 (ref. 2). Three segments of human DNA completely spanning the 17.1-kb blond-associated region (as defined by the nearest flanking non-significant markers in the GWAS) were separately cloned upstream of a minimal promoter and *lacZ* reporter gene (**Fig. 2a**). Only the 6.7-kb region, H2, drove consistent reporter expression in transgenic mouse embryos (**Fig. 2b–d** and **Supplementary Fig. 1**). *lacZ* expression was visible in the kidney (n = 14/15 embryos) and in developing hair follicles (n = 13/15 embryos). We subsequently tested two smaller clones from the H2 region, each of which overlapped major peaks of mammalian sequence conservation. One of these fragments, H2b, drove consistent expression in kidney (n = 12/13 embryos; **Fig. 2f** and **Supplementary Fig. 2**), and the other fragment, HFE (for hair follicle enhancer), drove consistent expression in developing hair follicles (n = 8/11 embryos; **Fig. 2e** and **Supplementary Fig. 3**). Histological analysis demonstrated that expression with the HFE fragment occurred in the epithelial cells of developing hair and skin (**Fig. 2g**,**h**), corresponding to a known site of endogenous *Kitl* expression[24] that has an important role in attracting melanocytes to the developing epidermis and hair follicles[25].

The location of the HFE overlaps SNP rs12821256 (**Fig. 2a**), which showed the strongest association with blond hair color[2]. Complete genome sequences for 270 individuals showed no additional linked sequence changes in the HFE that are common in northern Europeans and absent in Africans (Online Methods). To test the effect of the rs12821256 polymorphism, we constructed derivatives of the H2 clone containing the ancestral allele (A; H2-ANC), the blond-associated allele (G; H2-BLD) or an 11-bp deletion overlapping the rs12821256 SNP position (H2-DEL) (**Fig. 3**). No obvious qualitative differences in *lacZ* expression patterns were noted between H2-ANC and H2-BLD transgenic embryos (**Fig. 3a**,**b** and **Supplementary Figs. 1** and **4**). However, for the H2-DEL construct, relatively weak *lacZ* expression was observed in hair, even in transgenic embryos with strong expression in kidney, as expected if the sequences surrounding rs12821256 modulate expression in hair but not kidney (**Fig. 3c** and **Supplementary Fig. 5**).

Although gross differences in *lacZ* staining are readily observed in transgenic mice, quantitative changes can be difficult to score owing to random variation in transgene copy number and integration site from embryo to embryo. We therefore cloned the ancestral (HFE-ANC), blond-associated (HFE-BLD) or 11-bp deleted (HFE-DEL) version of the hair follicle enhancer upstream of a luciferase reporter gene and examined quantitative levels of expression in the HaCaT human keratinocyte cell line[26] (**Fig. 3d**). The HFE-ANC construct resulted in a 17-fold increase over background activity, confirming that the HFE is active in human keratinocytes (Stouffer's Z method, $P = 1 \times 10^{-24}$; $n = 8$ biological replicates). Cells with the HFE-DEL construct showed a marked 73% reduction in luciferase activity compared to cells with HFE-ANC (Stouffer's Z method, $P = 8 \times 10^{-16}$; $n = 5$). The HFE-BLD construct with the rs12821256 base-pair change also showed a small but consistent reduction in activity, which was 22% less than with the HFE-ANC construct (Stouffer's Z method, $P = 1 \times 10^{-8}$; $n = 8$).

Examination of human Encyclopedia of DNA Elements (ENCODE) regulatory tracks[11] identified a robust TCF/LEF signal that overlapped the position of rs12821256 in chromatin immunoprecipitation and sequencing (ChIP-seq) experiments with TCF7L2 in HCT116 colorectal carcinoma epithelial cells (**Fig. 4a**). LEF1 is a major WNT-activated transcription factor that binds DNA through a high-mobility-group (HMG) domain[27–29]. *LEF1* is known to be expressed during hair follicle development and regeneration[30,31], and *Lef1* knockout mice have defects in hair and whisker formation[32]. Interestingly, *Lef1* knockout mice appear light-colored owing to local failure of melanogenesis in mutant hair follicles[32]. Conversely, hyperactivation of WNT signal transduction in epithelial cells leads to increased *Kitl* expression and hyperpigmentation in mice[33]. Finally, LEF1 is known to induce sharp DNA bends[27,34] and stimulate DNA looping interactions[35,36]. Thus, LEF1 is an excellent candidate for an upstream factor that could act on distant regulatory enhancers of *KITLG* and other melanogenic genes[37]. The rs12821256 SNP alters a well-conserved base within a sequence that resembles a consensus LEF binding motif (**Fig. 4b**). Examination of the UniProbe collection of protein-binding

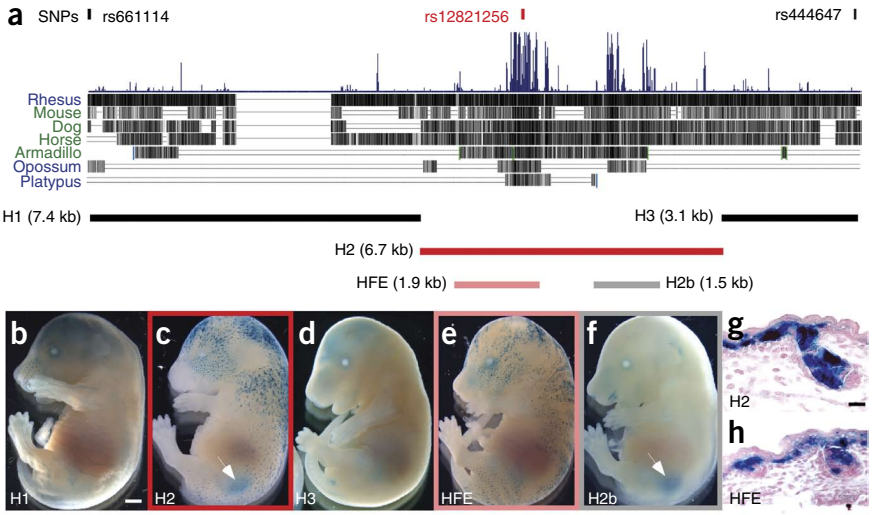

Figure 2 The human blond-associated region contains a functional hair follicle enhancer. (**a**) A 17.1-kb region bounded by SNPs rs444647 and rs661114 defines the candidate interval for blondness[2]. Within this region, a large block of mammalian sequence conservation (blue peaks) overlaps peak marker rs12821256. Five human fragments were cloned upstream of a *lacZ* reporter gene and tested for *in vivo* enhancer activity in transgenic mice. (**b**–**f**) Representative transgenic embryos generated by pronuclear microinjection with the different *lacZ* constructs, processed at embryonic day (E) 16.5 to show *lacZ* gene activity (blue staining). Scale bar, 1 mm. (**b**) H1. (**c**) H2. (**d**) H3. (**e**) HFE (for hair follicle enhancer). (**f**) H2b. Pictures in **b**–**f** are representative of 17, 15, 11, 11 and 13 independent transgenic embryos, respectively. Of the three clones spanning the entire interval, only the 6.7-kb clone, H2, produced consistent *lacZ* expression in skin and kidney (arrow). Analysis of two subclones of H2 separated HFE skin (**e**) and H2b kidney (**f**, arrow) enhancers. (**g**,**h**) Cross-sections (6 µm) through E16.5 dorsal skin from H2 (**g**) and HFE (**h**) transgenic embryos counterstained with nuclear fast red. Strong *lacZ* expression is visible in the basal epithelium and developing hair follicles. Scale bar, 30 µm.

microarray data confirmed that oligonucleotides matching the HFE ChIP-seq motif bind directly to TCF/LEF *in vitro* and oligonucleotides containing the rs12821256 A>G change show less experimental binding[38,39] (**Supplementary Table 1**), consistent with the observed differences in activity of the two HFEs in our assays. To further quantify the LEF1 responsiveness of the *KITLG* HFE sequences, we cotransfected HaCaT keratinocytes with different mini-promoter constructs containing canonical, human ancestral or human blond-associated binding motifs together with a *LEF1* expression plasmid (**Fig. 4c**). As expected, the canonical LEF1 response sequence showed the highest level of activation. Similarly, the ancestral binding motif found in the human *KITLG* enhancer also mediated strong LEF1 responsiveness. In contrast, the blond-associated rs12821256 base-pair change in the motif reduced but did not eliminate LEF1 responsiveness *in vitro* (**Fig. 4c**).

To determine whether small quantitative changes in enhancer activity were sufficient to alter hair color *in vivo*, we generated matched lines of transgenic mice that expressed *Kitl* cDNA under the control of an 864-bp version (HE) of either the ANC or BLD hair enhancer (**Fig. 5a**). To eliminate possible differences due to transgene copy number, orientation or integration site, we used the ϕC31 integrase system to generate single-copy integrants at the *H11P3* locus on mouse chromosome 11 (ref. 40). We verified that integration occurred at the same position in both transgenic lines. Therefore, any phenotypic differences can be attributed to the status of a single base pair present in the HE—the presence of an A allele (ANC-*Kitl*) or a G allele (BLD-*Kitl*) corresponding to rs12821256.

Quantitative RT-PCR analysis of *Kitl* mRNA expression from postnatal day (P) 8 dorsal skin samples confirmed higher *Kitl* mRNA expression from both site-specific integrants compared to controls (**Fig. 5b**). The BLD-*Kitl* insertion increased *Kitl* expression by only 79% of the increase mediated by the ANC-*Kitl* insertion (**Fig. 5b**), consistent with the quantitative differences also observed between the two enhancers in cell culture. Notably, this quantitative difference was also sufficient to produce obvious visual differences in hair color. Both the ANC-*Kitl* and BLD-*Kitl* transgenic mice had altered hair pigmentation compared to control mice (**Fig. 5c** and **Supplementary Fig. 6**). However, the coats of BLD-*Kitl*/+ mice appeared to be of significantly lighter color than the coats of ANC-*Kitl*/+ mice and showed lower

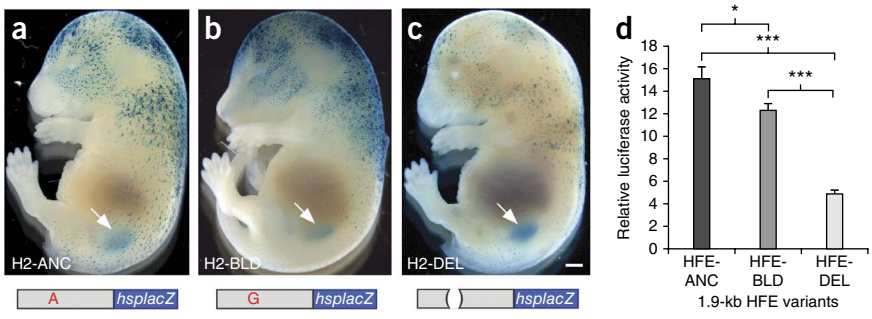

Figure 3 Variant hair follicle enhancers produce altered levels of gene expression. (**a**–**c**) Representative E16.5 transgenic embryos, generated by pronuclear injection with different 6.7-kb H2-*lacZ* constructs (shown below), processed for *lacZ* gene activity (blue). The full H2 region was used for these experiments, as expression in kidney provided a control for successful integration and expression of constructs, even if expression in hair was disrupted. The clones tested were H2-ANC with the A allele at rs12821256 (**a**), H2-BLD with the G allele at rs12821256 (**b**) and H2-DEL with an 11-bp deletion that removes the rs12821256 position (**c**). *lacZ* gene activity was observed in developing hair follicles and kidney (arrows) in all transgenic embryos. Although no consistent difference was noted between H2-ANC ($n = 15$) and H2-BLD ($n = 9$) embryos, H2-DEL embryos ($n = 8$) showed reduced *lacZ* activity in skin but normal kidney expression. Scale bar, 1 mm. (**d**) Expression analysis of different 1.9-kb HFE-luciferase reporters in the human HaCaT keratinocyte cell line. Bars represent the mean increase in luciferase gene activity relative to an empty vector control measured 48 h after transfection for a typical experiment. The enhancers tested differed only by the following: HFE-ANC, A at rs12821256; HFE-BLD, G at rs12821256; HFE-DEL, 11-bp deletion removing rs12821256. Both the HFE-BLD and HFE-DEL constructs exhibited significantly reduced activity in HaCaT keratinocytes compared to the HFE-ANC plasmid. Error bars represent s.e.m. $^{*}P < 0.05$, $^{***}P < 5 \times 10^{-4}$, unpaired *t* test.

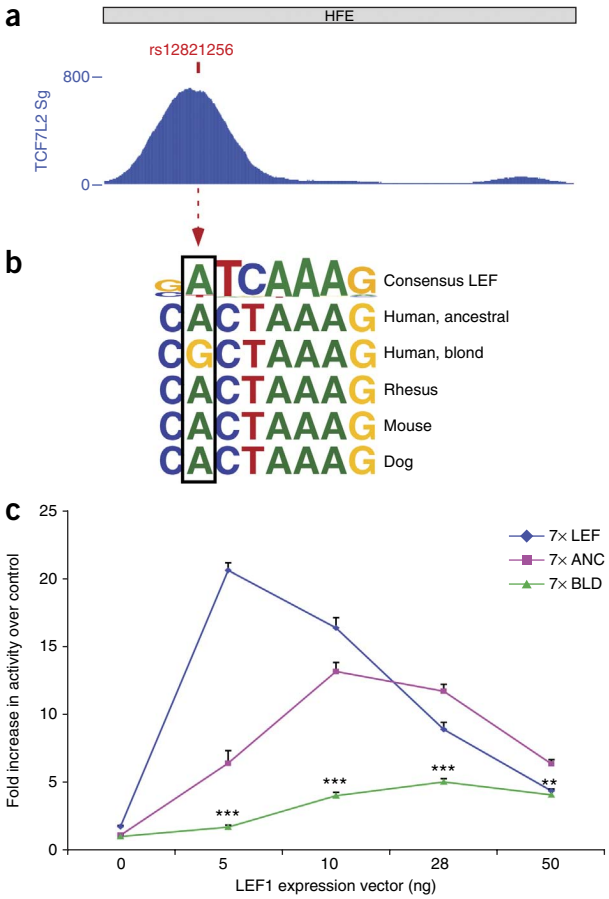

Figure 4 The blond-associated allele at rs12821256 alters a TCF/LEF binding site and reduces LEF responsiveness in keratinocytes. (**a**) A 4-kb window centered on the blond-associated SNP shows that TCF7L2 ChIP-seq reads from the ENCODE Project[11] accumulate over rs12821256 in the 1.9-kb HFE (NCBI36/hg18, chr. 12: 87,852,100–87,853,992; HCT116 cells, TCF7L2 Sg data set). (**b**) The sequence surrounding rs12821256 resembles a consensus TCF/LEF binding motif[39]. The blond-associated allele in humans changes a highly conserved, consensus-matching A nucleotide to a non-consensus G nucleotide within the predicted TCF/LEF binding motif. (**c**) Response of mini-promoters to increasing levels of LEF1 protein 48 h after cotransfection into HaCaT keratinocytes. The three luciferase reporter constructs tested contained seven tandem copies of an artificial consensus LEF binding site (7× LEF) (SuperTOPFlash[43]), seven copies of the human ancestral binding site (7× ANC) or seven copies of the blond-associated sequence variant (7× BLD). All three mini-promoters demonstrated elevated activity in response to increased amounts of LEF1 protein. The magnitude of the response to moderate LEF1 levels corresponds with the predicted binding capabilities of the variant LEF sites, with 7× LEF >> 7× ANC >> 7× BLD. Note that the 7× BLD human variant shows significantly lower activation than the 7× ANC human sequence at every level of LEF1 tested (5 ng, $P < 0.0001$; 10 ng, $P < 0.0001$; 28 ng, $P < 0.0001$; 50 ng, $P < 0.001$, Mann-Whitney test). A representative experiment is shown of $n = 2$ biological replicates. Error bars, s.e.m. $^{**}P < 0.005$, $^{***}P < 5 \times 10^{-4}$.

pigmentation density in hair shafts ($P < 0.028$; **Supplementary Fig. 7**). Taken together, these data demonstrate that a single-base change in a *KITLG* regulatory sequence is sufficient to significantly alter the activity of a functional hair follicle enhancer. Modulated *KITLG* expression produces different hair pigmentation *in vivo* and provides a molecular link between a regulatory change in an essential developmental signaling gene and a classic morphological phenotype in humans. Additional studies will be required to determine whether the rs12821256 polymorphism alters the number, size, dendricity, or differentiation and pigment synthesis of melanocytes positive for expression of the KIT receptor, all features that are known to respond to different levels of KITLG signaling in human skin grafts[41]. However, our results with the rs12821256 polymorphism already illustrate how changes in tissue-specific enhancers can localize phenotypes to particular body regions. Some human pigmentation variants alter general aspects of pigment biosynthesis, producing changes in all melanocytes, and therefore have pleiotropic effects on hair, skin and eye color. However, it is well known that hair and eye color can also vary independently, producing common human phenotypes such as light-haired individuals with brown eyes or brown-haired individuals with blue eyes. The rs12821256 variant alters an enhancer that is active specifically in the hair follicle environment, providing a simple genetic explanation for previous observations that this SNP is associated with changes in hair pigmentation but not eye pigmentation in northern Europeans[2].

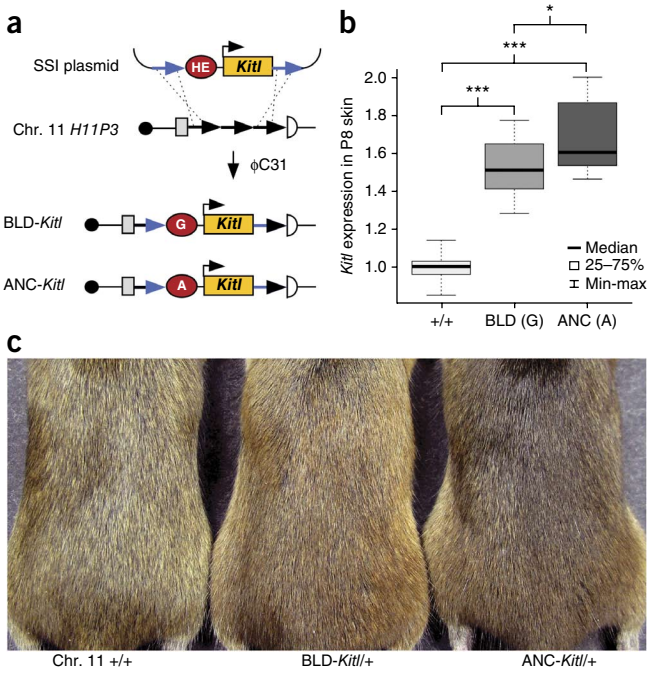

Figure 5 Mouse lines differing at a single base-pair position in the *KITLG* hair enhancer (HE) show obvious differences in hair color. (**a**) Schematic of the site-specific integration (SSI) strategy used to create matched BLD-*Kitl* and ANC-*Kitl* insertions in mice. The blond-associated or ancestral HE was cloned upstream of an *Hspa1b* minimal promoter–*Kitl* transgene (Online Methods). Blue arrows denote flanking *attB* sites that recombine with tandem *attP* sites (black arrows) in the mouse chromosome 11 *H11P3* locus upon pronuclear injection of a mix containing each SSI plasmid with ϕC31 mRNA. (**b**) Box plots representing quantitative RT-PCR analysis of *Kitl* RNA expression in P8 dorsal skin. Both BLD-*Kitl*/+ and ANC-*Kitl*/+ heterozygotes exhibit significantly higher levels of epidermal *Kitl* than control mice. However, mice carrying the blond-associated variant transgene produce 21% less *Kitl* than mice with the matched ancestral transgene. Mann-Whitney *P* values: BLD versus +/+ = 5×10^{-9}; ANC versus +/+ = 7×10^{-10}; BLD versus ANC = 0.03146. $^{*}P < 0.05$, $^{***}P < 5 \times 10^{-4}$. (**c**) Representative 2-month-old mice exhibiting the hair color phenotypes associated with a single copy of each SSI transgene. The mice pictured from left to right are wild type (FVB/C57BL/6J F_1 hybrid) and BLD-*Kitl*/+ and ANC-*Kitl*/+ heterozygotes. Mice carrying the blond-associated allele at rs12821256 are notably lighter than mice carrying the ancestral allele at the *KITLG* HE. Representative picture of four matched replicates, with additional quantification in **Supplementary Figure 7**.

The current data highlight why it is still so difficult to identify the causal basis of human trait associations. The rs12821256 SNP maps more than 350 kb from *KITLG*, acts at a specific anatomical site whose active enhancers have not yet been characterized in large-scale studies of human chromatin marks[11], alters a sequence that does not perfectly match a LEF1 consensus binding site and only causes an approximately 20% reduction in the activity of a previously unrecognized hair follicle enhancer. However, our results also illustrate how these difficulties can now be overcome using information from human population surveys, large-scale genome annotation projects and transcription factor interaction databases in combination with detailed functional tests of enhancer activity in cell lines and in mice. Well-matched animal models may ultimately be necessary to confirm the specific nucleotides that underlie many other human trait associations[42], making it possible to link phenotypic variation to particular base-pair changes in either protein-coding sequences or the complex regulatory sequences that surround many mammalian genes.

URLs. Allele Frequency Database (ALFRED), http://alfred.med.yale.edu/; MetaP program to combine *P* values, http://compute1.lsrc.duke.edu/softwares/MetaP/metap.php; UniProbe, http://the_brain.bwh.harvard.edu/uniprobe.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank R. Moon (University of Washington) for the XE237 *LEF1* expression plasmid, R. Nusse (Stanford University) for the SuperTOPFlash plasmid, C. Lowe (Stanford University) for help with statistical and 1000 Genomes Project analysis and members of the Kingsley laboratory for useful comments on the manuscript. This work was supported in part by the University of Georgia Research Foundation (M.A.B.) and by US National Institutes of Health grants GM65393 (M.A.B.), R01-NS050835 (L.L.) and a Center of Excellence in Genomic Science award 5P50HG2568 (D.M.K.). L.L. and D.M.K. are investigators of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

C.A.G. and D.M.K. conceived and oversaw the project. M.A.B. isolated and sequenced the Sl^{Pan} breakpoint. B.T. and L.L. provided advice, reagents and mice for generating site-specific integrants. C.A.G. performed the gene expression analysis in Sl^{Pan} mutants, carried out the transgenic analysis of the blond-associated GWAS interval, identified the hair follicle enhancer and performed *in vitro* and *in vivo* tests of the effects of the rs12821256 polymorphism. C.A.G. and D.M.K. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Sturm, R.A. Molecular genetics of human pigmentation diversity. *Hum. Mol. Genet.* **18**, R9–R17 (2009).
2. Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nat. Genet.* **39**, 1443–1452 (2007).
3. Cruz-Inigo, A.E., Ladizinski, B. & Sethi, A. Albinism in Africa: stigma, slaughter and awareness campaigns. *Dermatol. Clin.* **29**, 79–87 (2011).
4. Pitman, J. *On Blondes* (Bloomsbury Publishing, New York, 2003).
5. Homer. *The Iliad of Homer* (University of Chicago Press, Chicago, 2011).
6. Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
7. Sulem, P. *et al.* Two newly identified genetic determinants of pigmentation in Europeans. *Nat. Genet.* **40**, 835–837 (2008).
8. Zhang, M. *et al.* Genome-wide association studies identify several new loci associated with pigmentation traits and skin cancer risk in European Americans. *Hum. Mol. Genet.* **22**, 2948–2959 (2013).
9. Kenny, E.E. *et al.* Melanesian blond hair is caused by an amino acid change in TYRP1. *Science* **336**, 554 (2012).
10. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
11. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
12. Praetorius, C. *et al.* A polymorphism in *IRF4* affects human pigmentation through a tyrosinase-dependent MITF/TFAP2A pathway. *Cell* **155**, 1022–1033 (2013).
13. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).
14. Olalde, I. *et al.* Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature* **507**, 225–228 (2014).
15. Morrison-Graham, K. & Takahashi, Y. Steel factor and c-kit receptor: from mutants to a growth factor system. *Bioessays* **15**, 77–83 (1993).
16. Russell, E.S. Hereditary anemias of the mouse: a review for geneticists. *Adv. Genet.* **20**, 357–459 (1979).
17. Nocka, K. *et al.* Molecular bases of dominant negative and loss of function mutations at the murine c-*kit*/white spotting locus: W^{37}, W^{v}, W^{41} and *W*. *EMBO J.* **9**, 1805–1813 (1990).
18. Bedell, M.A., Copeland, N.G. & Jenkins, N.A. Multiple pathways for *Steel* regulation suggested by genomic and sequence analysis of the murine *Steel* gene. *Genetics* **142**, 927–934 (1996).
19. Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7**, 98–115 (2013).
20. Rajeevan, H. *et al.* ALFRED: the ALelle FREquency Database. Update. *Nucleic Acids Res.* **31**, 270–271 (2003).
21. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
22. Beechey, C.V., Loutit, J.F. & Searle, A.G. Panda, a new *Steel* allele. *Mouse News Lett.* **74**, 92 (1986).
23. Bedell, M.A. *et al.* DNA rearrangements located over 100 kb 5′ of the *Steel* (Sl)-coding region in Steel-panda and Steel-contrasted mice deregulate Sl expression and cause female sterility by disrupting ovarian follicle development. *Genes Dev.* **9**, 455–470 (1995).
24. Peters, E.M.J., Tobin, D.J., Botchkareva, N., Maurer, M. & Paus, R. Migration of melanoblasts into the developing murine hair follicle is accompanied by transient c-Kit expression. *J. Histochem. Cytochem.* **50**, 751–766 (2002).
25. Jordan, S.A. & Jackson, I.J. MGF (KIT ligand) is a chemokinetic factor for melanoblast migration into hair follicles. *Dev. Biol.* **225**, 424–436 (2000).
26. Boukamp, P. *et al.* Normal keratinization in a spontaneously immortalized aneuploid human keratinocyte cell line. *J. Cell Biol.* **106**, 761–771 (1988).
27. Giese, K., Amsterdam, A. & Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* **5**, 2567–2578 (1991).
28. Travis, A., Amsterdam, A., Belanger, C. & Grosschedl, R. LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor α enhancer function. *Genes Dev.* **5**, 880–894 (1991).
29. Waterman, M.L., Fischer, W.H. & Jones, K.A. A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer. *Genes Dev.* **5**, 656–669 (1991).
30. Zhou, P., Byrne, C., Jacobs, J. & Fuchs, E. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev.* **9**, 700–713 (1995).
31. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
32. van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1–deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
33. Zhang, Y. *et al.* Activation of β-catenin signaling programs embryonic epidermis to hair follicle fate. *Development* **135**, 2161–2172 (2008).
34. Love, J.J. *et al.* Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
35. Yun, K., So, J.S., Jash, A. & Im, S.H. Lymphoid enhancer binding factor 1 regulates transcription through gene looping. *J. Immunol.* **183**, 5129–5137 (2009).
36. Jash, A., Yun, K., Sahoo, A., So, J.S. & Im, S.H. Looping mediated interaction between the promoter and 3′ UTR regulates type II collagen expression in chondrocytes. *PLoS ONE* **7**, e40828 (2012).
37. Visser, M., Kayser, M. & Palstra, R.J. *HERC2* rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the *OCA2* promoter. *Genome Res.* **22**, 446–455 (2012).
38. Berger, M.F. *et al.* Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.* **24**, 1429–1435 (2006).
39. Newburger, D.E. & Bulyk, M.L. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* **37**, D77–D82 (2009).
40. Tasic, B. *et al.* Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proc. Natl. Acad. Sci. USA* **108**, 7902–7907 (2011).
41. Grichnik, J.M., Burch, J.A., Burchette, J. & Shea, C.R. The SCF/KIT pathway plays a critical role in the control of normal human melanocyte homeostasis. *J. Invest. Dermatol.* **111**, 233–238 (1998).
42. Kamberov, Y.G. *et al.* Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* **152**, 691–702 (2013).
43. Veeman, M.T., Slusarski, D.C., Kaykas, A., Louie, S.H. & Moon, R.T. Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680–685 (2003).

ONLINE METHODS

Primers. All primers used in construct design, genotyping and quantitative RT-PCR are listed in **Supplementary Table 2**.

Mouse strains and transgenics. The *Sl*^{pan} allele was maintained on the C3H background[23]. Additional C3H pups at P7 were obtained from Jackson Laboratories. *lacZ* reporter DNA constructs were prepared for microinjection as described[44]. Pronuclear microinjection, collection of E16.5 embryos and fixation in cold 4% paraformaldehyde in 1x PBS for 1 h were carried out at Taconic or Cyagen Biosciences. Fixed embryos were shipped to Stanford University, where they were hemisected, fixed for an additional 30 min and processed for *lacZ* activity as described[45]. Skin samples were dehydrated, embedded in paraffin, sectioned, dewaxed and stained with nuclear fast red (Vector Labs).

The HE-*Kitl* SSI plasmids were individually mixed with ϕC31 mRNA and injected into the pronuclei of *H11P3* FVB embryos by the Stanford transgenics facility as described[40]. Genomic DNA from putative founder and offspring mice were analyzed with primer pair PR387 and PR425 (ref. 40), as well as primer pair PR522 (ref. 40) and Kg1576, and with *Kitl* primers Kg1580 and Kg1581 to screen for site-specific and random integrations. HE-*Kitl*/+ mice, with identical insertion sites, on an FVB background were crossed to C57BL/6J mice to examine gene expression and pigmentation phenotypes. Skin samples from 2-month-old FVB/C57BL/6J F_1 hybrids were fixed with 4% paraformaldehyde in 1× PBS at 4 °C for 48 h, washed, dehydrated and embedded in paraffin. Sections (6 µm) were collected and processed as above. An independent pair of HE-*Kitl*/+ mouse lines, which also had insertion of single copies of the ANC-*Kitl* or BLD-*Kitl* transgene using different combinations of *attP* sites at the *H11P3* locus, were generated to confirm hair phenotypes and *Kitl* mRNA expression and to quantify pigmentation density in hair shafts. All procedures were carried out in accordance with protocols approved by the Stanford University Institutional Animal Care and Use Committee. Both male and female embryos and mice were used, and sample size was not predetermined.

Characterization of the *Sl*^{pan} breakpoint. Cloning of a genomic fragment containing the *Sl*^{pan} breakpoint was described previously[23]. The portion of this fragment that contains the distal inversion breakpoint was sequenced using conventional methods and aligned to the mouse genome (December 2011 assembly, GRCm38/mm10). Inversion breakpoints occurred between nucleotides 99,915,895 and 99,915,897 (approximately 100 kb upstream of the *Kitl* transcriptional start site at 100,015,824) and between nucleotides 34,300,646 and 34,300,649 (consistent with previous genetic and cytological data mapping the inversion to a distant location on proximal chromosome 10; ref. 23).

Plasmids. The 7× LEF plasmid (SuperTOPFlash, originally published as M50 Super 8× TOPFlash[43]) was a gift from the Nusse laboratory (Stanford University). The *LEF1* expression plasmid XE237 Human LEF-1 pCS2+ was purchased from Addgene (plasmid 16709). We digested XE237 with EcoRI and XbaI, blunted the ends with Klenow and religated it to generate the control expression vector pCS2+.

The genomic fragments in H2, H2b, H3, HFE and HFE-ANC were amplified from human genomic DNA (Clontech, 636401) and cloned into the NotI site of the Not5′hsp*lacZ* vector[44] or the XhoI site of the firefly luciferase reporter plasmid pTA-*Luc* (Clontech).

To generate clone H1, 4.7 kb of the 7.4-kb genomic fragment was amplified from human genomic DNA (Clontech) and cloned into the PCR BLUNT II TOPO vector (Invitrogen). An additional 2.7 kb corresponding to NCBI36/hg18 chr. 12: 87,854,729–87,857,439 was synthesized by GenScript and cloned into the PUC19 vector. The GenScript clone was digested with EcoRI, and the 4.7-kb EcoRI fragment from the TOPO clone was inserted. The resulting 7.4-kb H1 insert was liberated by NotI digestion and cloned into the Not5′hsp*lacZ* vector.

The genomic fragments in H2-BLD and HFE-BLD were amplified from Icelandic human DNA (Coriell, GM15755) and cloned into the Not5′hsp*lacZ* or pTA-*Luc* vectors, respectively.

The H2-DEL insert was generated by amplification of two PCR products, H2del1 and H2del2, from human genomic DNA (Clontech). These PCR intermediates were joined together in a third round of PCR to generate the 11-bp deletion within the larger 6.7-kb clone. Similarly, the 11-bp deletion within the 1.9-kb HFE-DEL insert was formed by PCR joining of two PCR intermediates, HFEdel1 and HFEdel2.

The 7× ANC and 7× BLD inserts were generated by annealing overlapping oligonucleotides, filling in the ends with Klenow and TOPO cloning the resulting products. Clones were sequenced to identify 7× inserts, which were liberated by XhoI digestion and were cloned into pTA-*Luc*.

To build the HE-*Kitl* plasmids, an *attB* site was amplified from pBT316 (ref. 40) and cloned into the SacI site of pBSKS+ to make 5′attBpBS. The *Hspa1b* promoter and SV40 polyadenylation fragments with ends overlapping *Kitl* cDNA were amplified from Not5′hsp*lacZ*. *Kitl* cDNA with ends overlapping the *Hspa1b* promoter and SV40 polyadenylation fragments was amplified from Thermo Scientific clone MMM1013_65096. Next, the *Kitl* and SV40 polyadenylation fragment PCR intermediates were joined in a second round of PCR. Finally, the *Hspa1b* promoter and *Kitl*–SV40 polyadenylation fragment intermediates were combined in a third round of PCR. A 2.2-kb fragment was gel purified and TOPO cloned. The *Hspa1b* promoter–*Kitl*–SV40 polyadenylation fragment TOPO plasmid was digested with NotI and SalI, and the insert was cloned into 5′attBpBS to generate *attB-Kitl*. A second *attB* site was added by digesting *attB-Kitl* with SalI and ligating an *attB* site liberated from the minicircle vector pBT346 (ref. 40) by SalI digestion to make *attB-Kitl-attB*. Minimal 864-bp ANC and BLD hair enhancers were amplified from Clontech or Icelandic human DNA. PCR products were digested with NotI and cloned into the 5′ NotI site in *attB-Kitl-attB* to make the final ANC-*Kitl* and BLD-*Kitl* constructs.

Sanger sequencing confirmed the sequence validity of all DNA constructs.

Human population genotypes and sequences. Allele frequencies for rs12821256 were compiled from the 1000 Genomes Project data set[21] and ALFRED[20]. Phased sequences from 93 Finnish, 89 UK and 88 African (Nigerian Yoruba) individuals showed no additional common base-pair changes within the HFE region that were linked with rs12821256.

Luciferase reporter assays. The human keratinocyte cell line HaCaT (CLS) was cultured in DMEM supplemented with 10% FBS, 2 mM L-glutamine and 1% penicillin-streptomycin and were not tested for mycoplasma. Cells were seeded into 24-well plates at a density of 1 x10^5 cells/well. We transfected 370 ng of pTA-*Luc* or enhancer clone HFE-ANC, HFE-BLD or HFE-DEL into cells using Lipofectamine 2000 (Invitrogen) with 30 ng of pRL-*tk* (Promega). To generate LEF1 response curves, 0 (220) ng, 5 (215) ng, 10 (210) ng, 28 (192) ng or 30 (190) ng of XE237 (pCS2+) plasmid was cotransfected into cells using Lipofectamine 2000 with 150 ng of pTA-*Luc*, 7× LEF, 7× ANC or 7× BLD constructs and 30 ng of pRL-*tk*. After 5 h of transfection, cell culture medium was replaced with standard medium supplemented with 2.8 mM calcium chloride (Sigma). Cell lysates were collected 48 h later and assayed on a Glo/Max Multi+ Detection system (Promega) using the Dual-Luciferase Reporter Assay System (Promega) according to the manufacturer's instructions.

Quantitative RT-PCR analysis. Dorsal skin samples from P8 or P21 FVB/C57BL/6J F_1 hybrids with +/+, BLD-*Kitl*/+ or ANC-*Kitl*/+ genotypes and P8 wild-type (C3H), *Sl*^{pan}/+ or *Sl*^{pan}/*Sl*^{pan} genotypes were placed in FastPrep Lysing Matrix A vials (MP Biomedicals) with 1 ml of TRI Reagent (Ambion). Tissue was homogenized using an MP FastPrep-24 with 4 × 30-s pulses at 6.0 m/s, with a 2-min incubation on ice following each 30-s pulse. After tissue disruption, total RNA was prepared according to the manufacturer's instructions. Contaminating genomic DNA was removed from RNA samples before cDNA synthesis using DNase I (Invitrogen). cDNA was transcribed from 500 ng of total RNA using SuperScript III First-Strand Synthesis SuperMix (Invitrogen). Quantitative RT-PCR was performed with *Kitl*-specific primers (Kg1580 and Kg1581) and *Actb*-specific primers (Kg1588 and Kg1589). All reactions were carried out in triplicate using Brilliant Green Low ROX Master Mix (Agilent Technologies) with a 2-step protocol (40 cycles of 95 °C for 15 s and 60 °C for 1 min) on an Mx3005P PCR system (Stratagene).

Quantification of pigmentation levels in lines with site-specific transgene insertion. Dorsal hair samples were collected from FVB/C57BL/6J F_1 hybrids, BLD-*Kitl*/+ (line 2) and ANC-*Kitl*/+ (line 2) heterozygotes at P21. Multiple

zigzag hairs were separated, attached to a piece of tape and affixed to a slide. A coverslip was applied using AquaMount, and photographs were taken of the middle portion of the hair between the distal-most bends of 15 hairs per mouse using a Zeiss axiophot (32× objective, 3,200 K lamp) with an exposure time of 16 ms. Images were imported into Adobe Photoshop and were measured to determine average pigment density per hair (defined as the total number of pixels corresponding to pigmented areas divided by total pixels comprising the hair shaft).

Statistical analyses. For cell culture experiments, the average activity of a construct in a given experiment was calculated using 6–8 individually transfected wells. Raw firefly luciferase values were normalized to *Renilla* luciferase values to control for variation in transfection efficiency. Elevated HFE-driven or mini-promoter expression represents activity over baseline pTA-*Luc* expression. All experiments were replicated (HFE enhancers, $n = 5–8$; 7× mini-promoters, $n = 2$). InStat software was used to determine the statistical significance of HFE-driven expression in unpaired, one-tailed *t* tests following verification of normality. MetaP[46] was used to combine *P* values from replicate HFE experiments. The Stouffer's Z combined *P* value, which corrects for sample size, was reported. The Mann-Whitney test, which makes no assumptions regarding data distribution, was used to determine significance in mini-promoter experiments. All graphs and plots show mean ± s.e.m.

For quantitative RT-PCR experiments, *Kitl* expression in mice from 3 litters for each transgene was analyzed to give a total of 22 wild-type, 11 BLD-*Kitl*/+ and 13 ANC-*Kitl*/+ samples. To control for variability in experiments run on different days, gene expression in all three genotypic classes was normalized to expression in a single dorsal skin sample from a 2-month-old wild-type mouse. Relative *Kitl* expression in each skin sample was then normalized to *Actb* levels, using the PFAFFL method[47]. *Kitl* expression levels in the two sets of wild-type littermates were not significantly different from each other (InStat two-tailed unpaired *t* test, $P = 0.1666$). Increases in *Kitl* expression in BLD-*Kitl*/+ and ANC-*Kitl*/+ mice were calculated relative to expression levels in their wild-type littermates. Plotting and statistical analysis were performed using R. Mann-Whitney rank-sum tests were used to determine significance.

Quantitative RT-PCR experiments on dorsal skin samples taken at P21 from the second pair of SSI lines was performed on mice from two litters for a total of four wild-type, three BLD-*Kitl*/+ and three ANC-*Kitl*/+ samples. Samples were treated as above, and statistical significance was determined using InStat software to perform unpaired, one-tailed *t* tests.

For quantitative RT-PCR experiments on Sl^{pan} mice, *Kitl* expression in three litters (two Sl^{pan} litters and one C3H litter) was analyzed to give a total of three Sl^{pan}/Sl^{pan} homozygotes, five $Sl^{pan}/+$ heterozygotes and six wild-type C3H mice. Gene expression of all samples was normalized to expression in a single dorsal skin sample from a 2-month-old wild-type mouse. Relative *Kitl* expression in each skin sample was then normalized to *Actb* levels, using the PFAFFL method[47]. Mann-Whitney tests were performed using InStat software to determine statistical significance.

For pigment density analysis, 15 individual zigzag hairs were measured from each of 4 wild-type, 3 BLD-*Kitl*/+ and 3 ANC-*Kitl*/+ mice. The reported pigmentation levels for the wild-type, BLD and ANC genotypic classes represent the averages of the mean pigmentation values for each mouse of the corresponding genotype. InStat software was used to determine statistical significance using unpaired, one-tailed *t* tests.

44. DiLeone, R.J., Russell, L.B. & Kingsley, D.M. An extensive 3′ regulatory region controls expression of *Bmp5* in specific anatomical structures of the mouse embryo. *Genetics* **148**, 401–408 (1998).
45. Mortlock, D.P., Guenther, C. & Kingsley, D.M. A general approach for identifying distant regulatory elements applied to the *Gdf6* gene. *Genome Res.* **13**, 2069–2081 (2003).
46. Whitlock, M.C. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **18**, 1368–1373 (2005).
47. Pfaffl, M.W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).

doi:10.1038/ng.2991