Text S1. Features of eight SNPs analyzed in this study

THREE SNPS ANALYZED USING CAPSEQ AND CG DATA

rs6822844 (autoimmune disease): this SNP resides in a region of strong linkage disequilibrium (LD) at chromosome region 4q27 that contains variation associated with type 1 diabetes (WTCCC 2007; Zhernakova et al. 2007; Maiti et al. 2010), celiac disease (van Heel et al. 2007; Adamovic et al. 2008; Hunt et al. 2008), Graves' disease (Todd et al. 2007), rheumatoid arthritis (Zhernakova et al. 2007; Daha et al. 2009; Teixeira et al. 2009; Hollis-Moffatt et al. 2010; Maiti et al. 2010), Crohn's disease (Zhernakova et al. 2007; Festen et al. 2009; Marquez et al. 2009), ulcerative colitis (Festen et al. 2009; Marquez et al. 2009), juvenile idiopathic arthritis (Albers et al. 2009), psoriasis (Liu et al. 2008), psoriatic arthritis (Liu et al. 2008), early onset psoriasis (Warren et al. 2011), alopecia areata (Petukhova et al. 2010) and systemic lupus erythematosus (Maiti et al. 2010). This LD region contains four genes, of which interleukin 2 *(IL2)* and interleukin 21 (*IL21)* are strong candidates for autoimmune disease risk. In particular, *IL-2* plays an important role in regulatory T cell development and in tolerance (Cheng, Yu, and Malek 2011). Accordingly, *IL-2* deficient mice suffer from autoimmune hemolytic anemia and inflammatory bowel disease (O'Shea, Ma, and Lipsky 2002; Ozaki et al. 2004). Four SNPs in near-complete LD with each other (rs6822844, rs13151961, rs13119723, and rs6840978) have very strong disease association signals. SNP rs6822844 is the most strongly associated variant in studies that tested all four SNPs (van Heel et al. 2007; Hunt et al. 2008) and the risk allele, rs6822844*G, significantly associates with lower IL2 levels (Fichna et al. 2013). For this reason, we focused our study on rs6822844; given the high LD among them, we expect the four SNPs to have similar ages. The protective allele, rs6822844*T, shows evidence of recent positive selection in the HapMap CEU based on patterns of extended haplotype homozygosity (iHS=-1.958). In addition, this SNP shows signals of selection based on the correlation between allele frequency and climate variables (transformed rank =0.034 and 0.022 for winter short wave radiation and maximum summer temperature).

rs3184504 (autoimmune disease): this SNP resides in chromosome region 12q24 that harbors a large number of SNPs significantly associated with several autoimmune diseases, including type 1 diabetes (WTCCC 2007; Barrett et al. 2009), celiac disease (Hunt et al. 2008; Dubois et al. 2010; Zhernakova et al. 2011), rheumatoid arthritis (Stahl et al. 2010; Zhernakova et al. 2011), multiple sclerosis (Alcina et al. 2010), autoimmune hepatitis (de Boer et al. 2014) as

well as immune-related traits, such as eosinophil counts (Gudbjartsson et al. 2009). The non-synonymous SNP rs3184504 (c.784T>C, p. Arg262Trp) in exon 3 of the *SH2B3* gene had the strongest association signals in most studies. Additionally, the derived rs3184504*T allele associated with disease risk was shown to be protective against bacterial infection (Zhernakova et al. 2010). *SH2B3* is an excellent candidate susceptibility gene for diseases of the immune response as it regulates T-cell receptor, growth factor and cytokine receptor–mediated signaling implicated in leukocyte and myeloid cell homeostasis (Li et al. 2000). In addition, *Sh2b3-/-* mice have increased responses to multiple cytokines (Velazquez et al. 2002). More importantly, homozygotes for the risk allele at rs3184504 show an increased production of pro-inflammatory cytokines after stimulation with muramyl dipeptide, a specific ligand of the NOD2 innate immune receptor (Zhernakova et al. 2010). Interestingly, rs3184504 was also found to be associated with variation in blood pressure (Levy et al. 2009; Ehret et al. 2011) and risk to coronary artery disease (Schunkert et al. 2011; Lian et al. 2013; Aghabozorg Afjeh et al. 2014), although the mechanism underlying these associations is not well understood. The risk allele at rs3184504, as well as multiple SNPs strongly correlated with it, have strong selection signals based on extended haplotype homozygosity (iHS= -2.756) and a strong correlation between allele frequency and climate variables (transformed rank = 0.006 and 0.016 for short wave radiation in the winter and absolute latitude), consistent with the action of spatially-varying selective pressures.

rs12913832 (eye and hair color): this SNP, located at the chromosome region 15q13, is a non-coding variant strongly associated with eye and hair color in GWAS conducted in Europeans (Sulem et al. 2007; Han et al. 2008; Kayser et al. 2008; Eriksson et al. 2010; Liu et al. 2010; Zhang et al. 2013). This variation also affects skin pigmentation (Branicki, Brudnik, and Wojas-Pelc 2009; Cook et al. 2009) and associates with melanoma (Fang et al. 2013), but to a lesser extent compared to eye color. In addition, a non-synonymous polymorphism rs1800414 (c.1844A>G, p.His615Arg) in the oculocutaneous albinism 2 (*OCA2*) gene was found to be significantly associated with skin pigmentation in East Asians (Edwards et al. 2010). We focused on SNP rs12913832, where the derived G allele is most consistently associated with light pigmentation in Europeans (Cook et al. 2009) and occurs at high frequency only in Western Eurasian populations while virtually absent everywhere else (Eiberg et al. 2008) (Figure S1). In contrast, the non-synonymous SNP rs1800414 is found predominantly in East Asians and occurs

at low frequencies (<4%) in some European populations (Donnelly et al. 2011). SNP rs12913832 resides within the *HERC2* gene, but it is a regulatory variant for the nearby *OCA2* gene (Eiberg et al. 2008), which is known to play a role in pigmentation as shown by animal studies (Gardner et al. 1992; Rinchik et al. 1993; Rosemblat et al. 1998) and by analysis of patients with Mendelian forms of oculocutaneous albinism (OMIM#203200). Importantly, molecular studies showed that the region spanning rs12913832 is highly conserved across species (Sturm et al. 2008) and contains an enhancer element that regulates *OCA2* expression in human epidermal melanocytes (Visser, Kayser, and Palstra 2012). In addition, the rs12913832*G allele was shown to alter transcription factor occupancy on the enhancer, chromatin loop formation and *OCA2* expression (Visser, Kayser, and Palstra 2012). The genomic region containing rs12913832 shows strong signatures of natural selection based on allele frequency divergence between European and sub-Saharan African or East Asian populations, on the extended haplotype homozygosity (Voight et al. 2006), on the site frequency spectrum (Williamson et al. 2007) and on patterns of correlation between allele frequency and climate variables including solar radiation (Hancock et al. 2011). In particular, the rs12913832*G allele is associated with extended haplotype homozygosity, consistent with the idea that light eye and hair pigmentation was favored in Europeans. Allele frequencies at this SNP are also strongly correlated with several climate variables, with the strongest signal being with short wave radiation flux in the winter (transformed rank = 0.000042).

FIVE SNPS ANALYZED USING ONLY CG DATA

rs4988235 (lactase persistence): lactase persistence has long been thought to be the result of cultural adaptation (Flatz and Rotthauwe 1973; Itan et al. 2010) to a lifestyle in which dairy products are a major staple of adult diet. The expression of the *LCT* gene in adult life is regulated by an enhancer located 14 kb upstream of the transcriptional start site. In Europeans, the derived allele, rs4988235*T, is strongly associated with lactase persistence. Functional assays have shown that rs4988235*T is a cis-regulatory variant that is located in the binding site for a major transcriptional factor (Oct-1) in intestinal epithelial cells and enhances the expression of *LCT* (Olds and Sibley 2003; Lewinsky et al. 2005). A haplotype carrying rs4988235*T extends for more than 1 Mb in Europeans (Bersaglieri et al. 2004) and shows a strong signature of selection (iHS=-3.870) (Voight et al. 2006).

rs1426654 and **rs16891982** (skin pigmentation): light skin pigmentation has been proposed to be adaptive at high latitudes to maintain appropriate levels of vitamin D synthesis (Jablonski and Chaplin 2000; Jablonski and Chaplin 2010; Yuen and Jablonski 2010). The derived alleles at these missense SNPs, rs1426654*A and rs16891982*G, cause amino acid changes at the *SLC24A5* (p.Ala111Thr) and *SLC45A2* (p.Leu374Phe) genes encoding the NCKX5 and MATP proteins that underlie pigmentation phenotypes in zebrafish (Lamason et al. 2005) and medaka (Fukamachi, Shimada, and Shima 2001), respectively. These derived alleles harbor signatures of recent positive selection based on allele frequency divergence and extended haplotype homozygosity (Soejima et al. 2006; Kimura et al. 2007; Myles et al. 2007; Norton et al. 2007; Sabeti et al. 2007).

rs17810546 and **rs2188962** (autoimmune disease): rs17810546 is upstream of the *IL12A* gene encoding interleukin-12 α chain, or p35, which is a subunit in 2 distinct heterodimeric cytokines: interleukin-12 (IL12) and IL35. rs2188962 is located within the chromosome 5 region associated with inflammatory bowel disease containing the *IRF1* gene as well as the *OCTN1* and *OCTN2* genes (Ma et al. 1999; Rioux et al. 2000; Rioux et al. 2001) and correlates with IRF1 expression in mononuclear cells and $CD14^{+}CD16^{-}$ monocytes (Raj et al. 2013). The functional importance of *IRF1* in Crohn's disease was also corroborated by the significantly increased level of mRNA in patients relative to controls (Huff et al. 2012). Although the derived alleles rs17810546*G and rs2188962*T increase the disease risk, both alleles have strong selection signals based on the extended haplotype homozygosity (iHS = -2.492 at rs17810546 and iHS = -2.600 at rs2188962) (Barreiro and Quintana-Murci 2010; Zhernakova et al. 2010; Huff et al. 2012; Raj et al. 2013).

All iHS values (Voight et al. 2006; Kudaravalli et al. 2009) were obtained from Haplotter (http://haplotter.uchicago.edu) and the climate correlation results (Hancock et al. 2011) from dbCLINE (http://genapps2.uchicago.edu:8081/dbcline/main.jsp).

Text S2. Allele age estimation by an Approximate Bayesian Computation (ABC) approach.

The Approximate Bayesian Computation (ABC) approach implemented in this study for the 8 SNPs with signatures of selection is essentially the same as that in Beaumont *et al* (Beaumont, Zhang, and Balding 2002) including the local linear regression and weighting procedures. The demographic model used was that estimated by Li and Durbin from whole genome sequence data for a CEU individual (Li and Durbin 2011): a 10-fold population reduction from 12,000 to 1,200 individuals (*i.e.,* bottleneck) from 40,000 to 27,500 years ago (*i.e*., from 1,600 to 1,100 generations ago assuming a 25-year generation); and an instantaneous 10-fold population expansion from 12,000 to 120,000 individuals 12,500 years ago (*i.e.,* 500 generations ago). The ABC estimation of the age of the advantageous mutation (t) and the corresponding selection coefficient (s) values of each of the 8 variants chosen in this study (see Text S1) is based on (i) their observed frequency in the HapMap or CG CEU sample (f_{obs}), (ii) their linked neutral genetic variation (G), and (iii) the simulations detailed below.

First, simulated trajectories of selected alleles are generated by Wright-Fisher forward simulation. To start, a value of t is randomly selected from a distribution ranging between 10 and 5,000 generations (*i.e*., 250 to 125,000 years ago), but with a density distribution proportional to population size (see above). Next, a value of s is randomly selected from a uniform distribution ranging between -0.01 and 0.1. Using the selection coefficient s starting t generations ago, a frequency trajectory is generated under a simple additive selective model with genetic drift. The simulated trajectory is accepted based on the final frequency of the simulated allele with probability proportional to the binomial probability of f_{obs}. If the simulated trajectory is rejected, a new s value is sampled from the prior uniform distribution. A new s value is sampled until a simulated trajectory is obtained for each t from the prior distribution ranging between 10 and 5,000 generations, so that the accepted s values were a sample from a posterior distribution given t and f_{obs}.

Next, for each of the accepted trajectories, neutral variation surrounding the selected site (G_i) is generated by coalescent simulation conditional on the trajectory. To model the uncertainty in mutation rate (μ), μ was sampled from a truncated normal distribution with mean equal to the divergence-based μ estimate (from the comparison of human with the other primates) and a variance of the square of the mean; values less than a half of μ and greater than 1.5 times μ were excluded. For each of the SNPs analyzed in this study, the above procedure was repeated until

1.5M t and s pairs with their corresponding frequency trajectories and neutral variation surrounding the selected allele (G_i) were generated. Because of the nature of our CapSeq data, simulation data were modified in the analyses of three SNPs (rs6822844, rs3184504 and rs12913832) to emulate (i) the presence of unsurveyed regions and (ii) the phasing uncertainty for novel and rare variants. Given that the CapSeq data contained unsurveyed regions, simulated variants in these regions were removed. Since repetitive elements were omitted in the array design, genomic regions for which no probes were designed were likely to remain unsurveyed or have lower coverage. We refer to those regions for which oligos were not designed as gaps, which consistently showed lower coverage and are expected to have higher error rates. Looking at our data, we observed that the variants lying in the un-targeted regions, but within 200bp of the targeted regions still have sufficient coverage and could be confidently called. However, variants farther than ~200 bp from the targeted regions were most likely missed. In order to emulate the effect of these gaps, we removed all segregating sites from the simulations that were located farther than 200 bp from the regions for which capture oligos were designed. Simulated chromosomes were paired to reproduce the same number of homozygote or heterozygote individuals sequenced by our study. Correct phasing of the simulated data was known, but to emulate the difficulties of phasing rare variants within our observed sequencing data, singleton and doubleton alleles were randomly assigned to one of two paired chromosomes. We tested how each of the modifications affects the estimation of allele age in Text S3.

For each of the 8 SNPs, the 1500 simulations (out of the above mentioned 1.5 M) that better fitted our observed data were chosen to obtain the posterior distributions for t and s. First, a set of summary statistics (*SSs*) (*i.e.,* $1/L_H$, M_H, and R_H; see main text and Text S4 for more details) standardized by their mean and standard deviation were calculated for the observed data (G) as well as for the above mentioned 1.5 M simulation data (G_i) sets. The *SSs* are based only on the genetic variation in the chromosomes bearing the selected allele. A simulation is accepted if $d(S(G), S(G_i))$ based on the 3 *SS*s between the observed and the simulated data is less than δ, where $d()$ is the Euclidean distance and δ is the one-sided tolerance limit. The value of δ was chosen so that 0.1% of the 1.5M t values are accepted. Thus, the t and s values of those final 1,500 simulations form the posterior distributions of t and s, and we characterized the posteriors using the mode-based point estimates and (ii) 95% credible intervals (CI).

Finally, we performed posterior predictive checks of the model for each SNP. To this end, we generated trajectories by drawing t from the marginal posterior distribution and, for each t value, by sampling s from the same uniform distribution used in the ABC and accepted those where the frequency at present is compatible with the observed frequency. The distribution of *SSs* was calculated from data simulated using the accepted trajectories (Figures S9 and S10). Most of the observed *SS*s from the eight SNPs are within the 95% confidence interval of the simulated distributions (Figure S9 and S10). We found an exception for one of the three *SS*s in the lactase allele.

For three of the SNPs (rs4988235, rs17810546 and rs3184504), the modal allele ages in the prior and posterior distributions were similar. To evaluate whether this result depends on the choice of prior for t, we repeated the estimation using a uniform prior on t varying between 40 and 5000 generations ago. As shown in Figure S11, the posterior distributions for rs4988235 and rs17810546 are similar with those obtained using a prior with density proportional to N. For rs3184504, the posterior distribution is slightly shifted towards older time, but the mode remains consistent with an origin during the spread of agriculture. Therefore, our posterior estimation depends on the prior distributions only weakly and it does not affect our conclusions.

Text S3. Choice of summary statistics.

We conducted an initial investigation of the correlation between age (t) and a broad range of summary statistics (*SS*s) under a simple demographic model with a constant population size (N = 10,000). We sampled t from a distribution that is proportional to the population size (i.e., uniform distribution between 20 and 5,000). Ten thousand simulations were produced with number of chromosomes n = 28, uniform recombination rate (r = 1 cM/Mbp), additive dominance (h = 0.5) and with the selected allele at a current frequency of 50%. For each simulated data set, we calculated *SS*s that have been widely used to detect signatures of natural selection, *i.e*. nucleotide diversity, number of segregating sites, Tajima's D, Fay and Wu's theta, Fay and Wu's H, number of singletons and the iHS score of the selected variant (Nei and Li 1979; Tajima 1989; Fay and Wu 2000; Voight et al. 2006), using: i) all chromosomes in the sample (*i.e*., chromosomes carrying both the ancestral and the derived allele at the selected site), ii) only the chromosomes carrying the derived allele at the focal variant, or iii) only the chromosomes carrying the derived allele at the focal variant but considering only the selected region (*i.e*., the region defined by extended haplotype homozygosity (EHH) > 0.05). The length of the selected region and the number of mutations accumulated were also calculated in the second and third conditions and used as *SSs*. Then, we calculated Spearman's correlation coefficients between allele age (t) and each one of the *SS*s (see table below).

Spearman rank correlation coefficients between different *SSs* and t			
*SS*s	i) All chromosomes	Chromosomes carrying the derived allele at the focal variant	
		ii) Entire region	iii) Selected region
Nucleotide diversity	0.4725	0.7551	0.7907
Number of segregating sites	0.4131	0.6556	0.8398
Tajima's D	0.3251	0.7070	0.0924
Fay and Wu's H	0.5801	0.6685	0.8056
Number of singletons	0.3980	0.1021	0.8455
iHS	0.4424	*NA*	*NA*
Number of mutations	*NA*	0.5103	0.8399
Length of selected region	*NA*	*NA*	0.8995

The highest Spearman's correlation values were observed between *t* and *SS*s calculated for the selected region (i.e. condition *iii*). Among the most strongly correlated *SSs*, we chose the length, number of singletons and number of mutations of the selected region. To reduce the high correlation (>0.9) among them, we scaled the number of mutations by the physical length of the selected region (M_H) and the number of singletons by the total number of segregating sites in the selected region (R_H). We also took the inverse of the length of the selected region ($1/L_H$), which is expected to be linearly related to *t*.

Lastly, the correlation between *t* and those three *SSs* ($1/L_H$, M_H and R_H) was assessed under a more realistic demographic model (Text S2) while also considering the region specific characteristics summarized in Table S6. Table S7 reports the correlation values between *t* and *SS*s and among *SS*s.

Text S4. Sensitivity of the age estimation to demographic history, sample size, gaps and phasing.

Another set of simulations were performed to understand the effect of the complex European population demographic history (Li and Durbin 2011) on the age estimate. We tested the performance of the ABC method for 3 variants assessed in the CapSeq and CG data (see different simulation parameters in Table S6) by simulating selection events at four specific time points (*t*) 2000, 1200, 800 and 400 generations ago (*i.e.,* 50000, 30000, 20000 and 10000 years ago assuming a 25-year generation time) representing pre-bottleneck, bottleneck, recovery and expansion periods of the European population (see Figure S7A-D, S7E-H and S7I-L for rs6822844, rs3184504 and rs12913831, respectively).

First, for each of the four *t* values, 100 data sets were simulated (i) by sampling 28 chromosomes with the same ratio of the number of selected sequences to that of non-selected sequences in our data (to the 14 CEU individuals studied); (ii) by removing genetic variation within un-surveyed gaps (to emulate lack of complete sequence data); and (iii) by keeping phased variants with a count greater than two, but randomly placing singleton and doubleton alleles on the paired chromosomes (to emulate phasing error). These 100 data sets for each *t* value were treated as "observed data". Next, as described in Text S2, 900K simulated trajectories were created by randomly sampling a value of *t* ranging between 10 and 5,000 generations and a value of *s* ranging between -0.01 and 0.1. These 900K simulations were used to estimate $\hat{t}$ given the observed data generated for each of the 100 simulations mentioned above, for which the *t* used to simulate data was known. Once $\hat{t}$ was estimated for each of the 100 starting simulations, the relative error of the ABC estimate is represented by $\log_2\left(\frac{\hat{t}}{t}\right)$ in Figure S7, so that a value of 1.0 or -1.0 corresponds to a two-fold over- or under-estimation, respectively. As shown in Figure S7, the ABC method is more accurate for younger selection events, with a significant increase of the variance for older selection events (rs6822844 and rs3184504). No statistically significant difference was observed between the relative error distributions of the simulated selection events before and during the bottleneck for rs6822844 and rs3184504, except for rs12913831 (p = 0.024).

Second, because the sample size of the CapSeq data was small we assessed if an increase of sample size would have reduced bias or variance in allele age estimation: n = 28 *versus* n = 128. This number of sampled chromosomes was chosen to match the sample size of the CG data.

A total of 900K simulated data sets were produced per subregion; and a random set of 128 chromosomes were chosen after which gap areas were masked and phasing error was emulated (see Figure S7-orange boxplots for results). A non-significant reduction of bias and variance is observed, but overall, no significant difference is observed between n = 28 (Figure S7-blue boxplots) and n = 128 (Figure S7-orange boxplots). When focused on n = 128, no statistically significant difference was observed between the relative error distributions of the two oldest simulated selection events for rs6822844 and rs3184504, except for rs12913831 (p = 0.0003).

Third, simulations were also used to assess the effect of (i) the presence of gaps observed in the CapSeq data and (ii) the phasing uncertainty for novel and rare variants. One set of 900K simulations were run, which was then modified to create two final simulation sets. In the first simulation set, singleton and doubleton alleles were randomly assigned to one of the two chromosomes, but none of the variants within our sequencing gaps were excluded. The presence of gaps did not affect the performance of the age estimation (Figure S7, blue *vs.* green boxplots). In the second set, variants within the gap were excluded, but all retained simulated variants were phased. A weak but not significant increase in the variance of the error was observed when phasing error was not emulated (Figure S7, blue *vs*. red boxplots).

REFERENCES

Adamovic, S., S. S. Amundsen, B. A. Lie, A. H. Gudjonsdottir, H. Ascher, J. Ek, D. A. van Heel, S. Nilsson, L. M. Sollid, and A. Torinsson Naluai. 2008. Association study of IL2/IL21 and FcgRIIa: significant association with the IL2/IL21 region in Scandinavian coeliac disease families. Genes Immun **9**:364-367.

Aghabozorg Afjeh, S. S., S. M. Ghaderian, R. Mirfakhraie, M. Piryaei, and H. Zaim Kohan. 2014. Association Study of rs3184504 C>T Polymorphism in Patients With Coronary Artery Disease. Int J Mol Cell Med **3**:157-165.

Albers, H. M., F. A. Kurreeman, G. Stoeken-Rijsbergen et al. 2009. Association of the autoimmunity locus 4q27 with juvenile idiopathic arthritis. Arthritis Rheum **60**:901-904.

Alcina, A., K. Vandenbroeck, D. Otaegui et al. 2010. The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes Immun **11**:439-445.

Barreiro, L. B., and L. Quintana-Murci. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet **11**:17-30.

Barrett, J. C., D. G. Clayton, P. Concannon et al. 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet **41**:703-707.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics **162**:2025-2035.

Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, and J. N. Hirschhorn. 2004. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet **74**:1111-1120.

Branicki, W., U. Brudnik, and A. Wojas-Pelc. 2009. Interactions between HERC2, OCA2 and MC1R may influence human pigmentation phenotype. Ann Hum Genet **73**:160-170.

Cheng, G., A. Yu, and T. R. Malek. 2011. T-cell tolerance and the multi-functional role of IL-2R signaling in T-regulatory cells. Immunol Rev **241**:63-76.

Cook, A. L., W. Chen, A. E. Thurber et al. 2009. Analysis of cultured human melanocytes based on polymorphisms within the SLC45A2/MATP, SLC24A5/NCKX5, and OCA2/P loci. J Invest Dermatol **129**:392-405.

Daha, N. A., F. A. Kurreeman, R. B. Marques, G. Stoeken-Rijsbergen, W. Verduijn, T. W. Huizinga, and R. E. Toes. 2009. Confirmation of STAT4, IL2/IL21, and CTLA4 polymorphisms in rheumatoid arthritis. Arthritis Rheum **60**:1255-1260.

de Boer, Y. S., N. M. van Gerven, A. Zwiers et al. 2014. Genome-wide association study identifies variants associated with autoimmune hepatitis type 1. Gastroenterology **147**:443-452.e445.

Donnelly, M. P., P. Paschou, E. Grigorenko et al. 2011. A global view of the OCA2-HERC2 region and pigmentation. Hum Genet.

Dubois, P. C., G. Trynka, L. Franke et al. 2010. Multiple common variants for celiac disease influencing immune gene expression. Nat Genet **42**:295-302.

Edwards, M., A. Bigham, J. Tan, S. Li, A. Gozdzik, K. Ross, L. Jin, and E. J. Parra. 2010. Association of the OCA2 polymorphism His615Arg with melanin content in east Asian populations: further evidence of convergent evolution of skin pigmentation. PLoS Genet **6**:e1000867.

Ehret, G. B.P. B. MunroeK. M. Rice et al. 2011. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature **478**:103-109.

Eiberg, H., J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K. W. Kjaer, and L. Hansen. 2008. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet **123**:177-187.

Eriksson, N., J. M. Macpherson, J. Y. Tung, L. S. Hon, B. Naughton, S. Saxonov, L. Avey, A. Wojcicki, I. Pe'er, and J. Mountain. 2010. Web-based, participant-driven studies yield novel genetic associations for common traits. PLoS Genet **6**:e1000993.

Fang, S., J. Han, M. Zhang, L. E. Wang, Q. Wei, C. I. Amos, and J. E. Lee. 2013. Joint effect of multiple common SNPs predicts melanoma susceptibility. PLoS One **8**:e85642.

Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics **155**:1405-1413.

Festen, E. A., P. Goyette, R. Scott et al. 2009. Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis. Gut **58**:799-804.

Fichna, M., M. Zurawek, P. Fichna, I. Ziółkowska-Suchanek, D. Januszkiewicz, and J. Nowak. 2013. Polymorphic variant at the IL2 region is associated with type 1 diabetes and may affect serum levels of interleukin-2. Mol Biol Rep **40**:6957-6963.

Flatz, G., and H. W. Rotthauwe. 1973. Lactose nutrition and natural selection. Lancet **2**:76-77.

Fukamachi, S., A. Shimada, and A. Shima. 2001. Mutations in the gene encoding B, a novel transporter protein, reduce melanin content in medaka. Nat Genet **28**:381-385.

Gardner, J. M., Y. Nakatsu, Y. Gondo, S. Lee, M. F. Lyon, R. A. King, and M. H. Brilliant. 1992. The mouse pink-eyed dilution gene: association with human Prader-Willi and Angelman syndromes. Science **257**:1121-1124.

Gudbjartsson, D. F., U. S. Bjornsdottir, E. Halapi et al. 2009. Sequence variants affecting eosinophil numbers associate with asthma and myocardial infarction. Nat Genet **41**:342-347.

Han, J., P. Kraft, H. Nan et al. 2008. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. PLoS Genet **4**:e1000074.

Hancock, A. M., D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin, R. Sukernik, G. Utermann, J. K. Pritchard, G. Coop, and A. Di Rienzo. 2011. Adaptations to climate-mediated selective pressures in humans. PLoS Genet **7**:e1001375.

Hollis-Moffatt, J. E., M. Chen-Xu, R. Topless et al. 2010. Only one independent genetic association with rheumatoid arthritis within the KIAA1109-TENR-IL2-IL21 locus in Caucasian sample sets: confirmation of association of rs6822844 with rheumatoid arthritis at a genome-wide level of significance. Arthritis Res Ther **12**:R116.

Huff, C. D., D. J. Witherspoon, Y. Zhang et al. 2012. Crohn's disease and genetic hitchhiking at IBD5. Mol Biol Evol **29**:101-111.

Hunt, K. A., A. Zhernakova, G. Turner et al. 2008. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet **40**:395-402.

Itan, Y., B. L. Jones, C. J. Ingram, D. M. Swallow, and M. G. Thomas. 2010. A worldwide correlation of lactase persistence phenotype and genotypes. BMC Evol Biol **10**:36.

Jablonski, N. G., and G. Chaplin. 2010. Colloquium paper: human skin pigmentation as an adaptation to UV radiation. Proc Natl Acad Sci U S A **107 Suppl 2**:8962-8968.

Jablonski, N. G., and G. Chaplin. 2000. The evolution of human skin coloration. J Hum Evol **39**:57-106.

Kayser, M., F. Liu, A. C. Janssens et al. 2008. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. Am J Hum Genet **82**:411-423.

Kimura, R., A. Fujimoto, K. Tokunaga, and J. Ohashi. 2007. A practical genome scan for population-specific strong selective sweeps that have reached fixation. PLoS One **2**:e286.

Kudaravalli, S., J. B. Veyrieras, B. E. Stranger, E. T. Dermitzakis, and J. K. Pritchard. 2009. Gene expression levels are a target of recent natural selection in the human genome. Mol Biol Evol **26**:649-658.

Lamason, R. L., M. A. Mohideen, J. R. Mest et al. 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science **310**:1782-1786.

Levy, D., G. B. Ehret, K. Rice et al. 2009. Genome-wide association study of blood pressure and hypertension. Nat Genet **41**:677-687.

Lewinsky, R. H., T. G. Jensen, J. Moller, A. Stensballe, J. Olsen, and J. T. Troelsen. 2005. T-13910 DNA variant associated with lactase persistence interacts with Oct-1 and stimulates lactase promoter activity in vitro. Hum Mol Genet **14**:3945-3953.

Li, H., and R. Durbin. 2011. Inference of human population history from individual whole-genome sequences. Nature **475**:493-496.

Li, Y., X. He, J. Schembri-King, S. Jakes, and J. Hayashi. 2000. Cloning and characterization of human Lnk, an adaptor protein with pleckstrin homology and Src homology 2 domains that can inhibit T cell activation. J Immunol **164**:5199-5206.

Lian, J., Y. Huang, R. S. Huang et al. 2013. Meta-analyses of four eosinophil related gene variants in coronary heart disease. J Thromb Thrombolysis **36**:394-401.

Liu, F., A. Wollstein, P. G. Hysi et al. 2010. Digital quantification of human eye color highlights genetic association of three new loci. PLoS Genet **6**:e1000934.

Liu, Y., C. Helms, W. Liao et al. 2008. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet **4**:e1000041.

Ma, Y., J. D. Ohmen, Z. Li, L. G. Bentley, C. McElree, S. Pressman, S. R. Targan, N. Fischel-Ghodsian, J. I. Rotter, and H. Yang. 1999. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. Inflamm Bowel Dis **5**:271-278.

Maiti, A. K., X. Kim-Howard, P. Viswanathan et al. 2010. Confirmation of an association between rs6822844 at the Il2-Il21 region and multiple autoimmune diseases: evidence of a general susceptibility locus. Arthritis Rheum **62**:323-329.

Marquez, A., G. Orozco, A. Martinez et al. 2009. Novel association of the interleukin 2-interleukin 21 region with inflammatory bowel disease. Am J Gastroenterol **104**:1968-1975.
Myles, S., M. Somel, K. Tang, J. Kelso, and M. Stoneking. 2007. Identifying genes underlying skin pigmentation differences among human populations. Hum Genet **120**:613-621.
Nei, M., and W. H. Li. 1979. Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. Proceedings of the National Academy of Sciences of the United States of America **76**:5269-5273.
Norton, H. L., R. A. Kittles, E. Parra, P. McKeigue, X. Mao, K. Cheng, V. A. Canfield, D. G. Bradley, B. McEvoy, and M. D. Shriver. 2007. Genetic evidence for the convergent evolution of light skin in Europeans and East Asians. Mol Biol Evol **24**:710-722.
O'Shea, J. J., A. Ma, and P. Lipsky. 2002. Cytokines and autoimmunity. Nat Rev Immunol **2**:37-45.
Olds, L. C., and E. Sibley. 2003. Lactase persistence DNA variant enhances lactase promoter activity in vitro: functional role as a cis regulatory element. Hum Mol Genet **12**:2333-2340.
Ozaki, K., R. Spolski, R. Ettinger et al. 2004. Regulation of B cell differentiation and plasma cell generation by IL-21, a novel inducer of Blimp-1 and Bcl-6. J Immunol **173**:5361-5371.
Petukhova, L., M. Duvic, M. Hordinsky et al. 2010. Genome-wide association study in alopecia areata implicates both innate and adaptive immunity. Nature **466**:113-117.
Raj, T., M. Kuchroo, J. M. Replogle, S. Raychaudhuri, B. E. Stranger, and P. L. De Jager. 2013. Common risk alleles for inflammatory diseases are targets of recent positive selection. Am J Hum Genet **92**:517-529.
Rinchik, E. M., S. J. Bultman, B. Horsthemke, S. T. Lee, K. M. Strunk, R. A. Spritz, K. M. Avidano, M. T. Jong, and R. D. Nicholls. 1993. A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. Nature **361**:72-76.
Rioux, J. D., M. J. Daly, M. S. Silverberg et al. 2001. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet **29**:223-228.
Rioux, J. D., M. S. Silverberg, M. J. Daly et al. 2000. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet **66**:1863-1870.
Rosemblat, S., E. V. Sviderskaya, D. J. Easty, A. Wilson, B. S. Kwon, D. C. Bennett, and S. J. Orlow. 1998. Melanosomal defects in melanocytes from mice lacking expression of the pink-eyed dilution gene: correction by culture in the presence of excess tyrosine. Exp Cell Res **239**:344-352.
Sabeti, P. C.P. VarillyB. Fry et al. 2007. Genome-wide detection and characterization of positive selection in human populations. Nature **449**:913-918.
Schunkert, H.I. R. KonigS. Kathiresan et al. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat Genet **43**:333-338.
Soejima, M., H. Tachida, T. Ishida, A. Sano, and Y. Koda. 2006. Evidence for recent positive selection at the human AIM1 locus in a European population. Mol Biol Evol **23**:179-188.
Stahl, E. A., S. Raychaudhuri, E. F. Remmers et al. 2010. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet **42**:508-514.
Sturm, R. A., D. L. Duffy, Z. Z. Zhao, F. P. Leite, M. S. Stark, N. K. Hayward, N. G. Martin, and G. W. Montgomery. 2008. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. Am J Hum Genet **82**:424-431.
Sulem, P., D. F. Gudbjartsson, S. N. Stacey et al. 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat Genet **39**:1443-1452.
Tajima, F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics **123**:585-595.
Teixeira, V. H., C. Pierlot, P. Migliorini et al. 2009. Testing for the association of the KIAA1109/Tenr/IL2/IL21 gene region with rheumatoid arthritis in a European family-based study. Arthritis Res Ther **11**:R45.
Todd, J. A., N. M. Walker, J. D. Cooper et al. 2007. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet **39**:857-864.
van Heel, D. A., L. Franke, K. A. Hunt et al. 2007. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet **39**:827-829.
Velazquez, L., A. M. Cheng, H. E. Fleming, C. Furlonger, S. Vesely, A. Bernstein, C. J. Paige, and T. Pawson. 2002. Cytokine signaling and hematopoietic homeostasis are disrupted in Lnk-deficient mice. J Exp Med **195**:1599-1611.
Visser, M., M. Kayser, and R. J. Palstra. 2012. HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter. Genome Res **22**:446-455.

Voight, B. F., S. Kudaravalli, X. Wen, and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol **4**:e72.

Warren, R. B., R. L. Smith, E. Flynn, J. Bowes, S. Eyre, J. Worthington, A. Barton, and C. E. Griffiths. 2011. A systematic investigation of confirmed autoimmune loci in early-onset psoriasis reveals an association with IL2/IL21. Br J Dermatol **164**:660-664.

Williamson, S. H., M. J. Hubisz, A. G. Clark, B. A. Payseur, C. D. Bustamante, and R. Nielsen. 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet **3**:e90.

WTCCC. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature **447**:661-678.

Yuen, A. W., and N. G. Jablonski. 2010. Vitamin D: in the evolution of human skin colour. Med Hypotheses **74**:39-44.

Zhang, M., F. Song, L. Liang et al. 2013. 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.

Zhernakova, A., B. Z. Alizadeh, M. Bevova et al. 2007. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. Am J Hum Genet **81**:1284-1288.

Zhernakova, A., C. C. Elbers, B. Ferwerda et al. 2010. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. Am J Hum Genet **86**:970-977.

Zhernakova, A., E. A. Stahl, G. Trynka et al. 2011. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet **7**:e1002004.