doi:10.1038/nature13673

Supplementary Information 8
Phenotypic inference

Karola Kirsanow*

* To whom correspondence should be addressed (kirsanow@uni-mainz.de)

Introduction
We assessed three ancient modern humans (the Loschbour forager, the Motala12 forager, and the Stuttgart farmer) at a panel of SNPs and multi-allelic markers having well-validated phenotypic effects in present-day humans. Many of the markers in our panel have also been affected by natural selection relatively recently in human prehistory.

Walsh et al. [1-4], among others [5-7], demonstrated that it is possible to predict human eye, hair, and skin color phenotypes with accuracy using a small number of DNA variants. Here, we predicted pigmentation phenotypes of the three ancient modern humans using two models [2,3,6-8] that have been validated in present-day populations [1,4,7-9] as well as on skeletal remains [10]. We used these models together with additional SNP and haplotype data to infer the most likely iris, hair and skin pigmentation for the Loschbour, Motala12, and Stuttgart individuals.

We also analyzed Loschbour and Stuttgart at 35 single nucleotide polymorphisms (SNPs) known from genome-wide association studies (GWAS) to be reproducibly associated with susceptibility to the Metabolic Syndrome (MetS) and compared the results of two different diabetes-risk score models incorporating 24 of these SNPs [11,12]. MetS-related SNPs have evidence of being under recent selection[13,14], possibly because of pressures related to changes in diet and climate associated with human migration and the adoption of agriculture.

Finally, we analyzed the Loschbour, Motala12, and Stuttgart samples at a panel of sites having well-described phenotypic effects that have been identified as targeted by selection in recent human prehistory [15-19].

Methods
We analyzed DNA polymorphism data stored in the VCF format [20] using the VCFtools software package (http://vcftoools.sourceforge.net/). For the Loschbour and Stuttgart individuals, we included data from sites not flagged as LowQuality, with genotype quality (GQ) of ≥30, and SNP quality (QUAL) of ≥50. We chose to assess Motala12 because this individual was sequenced at higher average coverage than the other Motala samples. However, the coverage of the Motala12 individual (2.4×) was appreciably lower than that of the Loschbour (22×) and Stuttgart (19×) individuals, and we therefore altered our genotyping methodology to account for the limitations imposed by lower coverage at our sites of interest. Specifically, we included sites having at least 2× coverage which passed visual inspection of the local alignment using samtools tview (http://samtools.sourceforge.net)[21].

We carried out five sets of phenotypic analyses:

(1) We assessed the genotypes of the Loschbour forager and Stuttgart farmer at pigmentation SNPs included in the 8-plex and the Hirisplex pigmentation phenotype prediction models

system (Table S8.1). We assigned probabilities to hair and eye color phenotypes using the enhanced version 1.0 Hirisplex Microsoft Excel macro[8] (Table S8.2) .

(2) We assessed the genotypes of the Loschbour and Motala foragers and the Stuttgart farmer at a panel of SNPs in the *HERC2/OCA2* and *SLC24A5* genes comprising several pigmentation-related haplotypes (Tables S8.3, S8.4, S8.5).

(3) We assessed the genotypes of the Loschbour and Stuttgart individuals at a panel of SNPs associated with risk for Metabolic Syndrome (Table S8.6) and that form the basis for two type 2 diabetes (T2D) risk score models [11,12]. We computed weighted genotype risk scores using the methods described in Meigs (2008) and Cornelis (2009). We additionally genotyped the three ancient modern humans (Table S8.7) at 5 SNPs in the *SLC16A11* gene forming a haplotype associated with T2D risk in which the risk haplotype appears to derive from Neanderthal introgression.

(4) We assessed the genotypes of the Loschbour, Motala12, and Stuttgart individuals at a panel of SNPs with evidence for recent natural selection, including several known to show high allele frequency differentiation between European and East Asian populations (Table S8.8).

(5) We assessed the genotypes of the Loschbour, Motala12, and Stuttgart individuals at 8 SNPS in the *NAT2* gene in order to determine the acetylation phenotype of the three ancient modern humans (Table S8.9).

We caution that the pigmentation phenotype models and metabolic syndrome risk scoring models are not independent. In particular, seven of the eight markers in the 8plex pigmentation model are also included in the Hirisplex model, and four metabolic syndrome-associated SNPs are shared between the two diabetes risk score models.

Results

Pigmentation

For hair color, the integrated results of the genotype-based pigmentation models indicate that there is at least a 98% probability that both the Stuttgart and Loschbour individuals had dark (brown or black) hair. The Hirisplex model assigns the highest probability to black hair color for both individuals (Table S8.2).

The results of the 8-plex skin pigmentation model were inconclusive for both the Loschbour and Stuttgart individuals. However, the Loschbour and Stuttgart genotypes at rs1426654 in *SLC24A5* indicate that the Stuttgart individual may have had lighter skin than the Loschbour hunter-gatherer. The Loschbour individual is homozygous for the rs1426654 ancestral allele, while Stuttgart is homozygous for the derived skin-lightening allele [22,23]. This allele has the single greatest effect on skin pigmentation of the SNPs identified to date in present-day populations[22-24].

For eye color, the single most significant determinant is the rs12913832 SNP in the *HERC2* gene. The genotype at this site excludes the possibility that the Stuttgart farmer had blue eyes. Positive iris color determinations are less secure. The Loschbour forager is homozygous for the derived allele at rs12913832, indicating that this individual is likely to have had blue (61% probability) or intermediate

iris color (17% probability). It has been suggested that this mutation arose within the last 6,000 to 10,000 years, and thus the Loschbour individual would have been a relatively early carrier [25].

It should be noted that while these predictive models have been well-validated in present-day European populations, it is possible that heretofore undetected variation in pigmentation genes may have contributed to phenotypic variation in ancient modern humans. Any such variation would not be captured by this analysis.

The number of sites without coverage in the Motala12 sample prevented the inclusion of this individual in the model-based phenotypic inference. However, some inferences can be made from high-impact single pigmentation SNPs and short haplotypes. The Motala12 forager, like the Stuttgart farmer, carries at least one copy of the derived rs1426654 pigmentation-lightening allele, and may thus have had lighter skin pigmentation than the Loschbour forager. We typed the three ancient modern humans at 7 SNPs forming three short haplotypes associated with eye color in present-day worldwide populations (Table S8.3)[26]. The observed reads in the Motala12 forager, like the Loschbour forager match the blue-eye-associated allele at all 7 SNPs. However, this includes two SNPs (rs7495174 and rs1291382) at only 1× coverage. Motala12 carries the blue-eye haplotype at the two BEH3 SNPs, which are in LD with the causal SNP, rs1291382, in present-day Europeans (but not outside of Europe)[26]. However, the Stuttgart farmer is also homozygous for the two blue-associated BEH3 SNPs, despite being homozygous for the ancestral allele at rs1291382. Stronger support for the inference of non-brown eyes for Motala12 is the observation of the derived allele at rs1129038, a site at almost complete LD with rs1291382 in present-day populations[26].

In order to determine the haplotypes carried by the ancient modern humans at major pigmentation loci and compare these ancient 'superalleles' with haplotypes in present-day populations, we genotyped all three ancient modern humans at a number of SNPS in the *OCA2*, *HERC2*, and *SLC24A5* genes:

1. We evaluated the three ancient modern humans at a panel of 13 SNPs in the *OCA2*/*HERC2* region comprising the haplotype shared by 97% of blue-eyed individuals in a present-day study population[25] (Table S8.4).

2. We determined the genotypes of the Loschbour, Stuttgart, and Motala12 individuals at a panel of 16 SNPs in the *SLC24A5* gene comprising the *SLC25A5* haplotype observed in most present-day humans carrying the derived rs1426654 (A111T) allele [27](Table S8.5).

We find that the Loschbour forager is homozygous for the h-1 *HERC2*/*OCA2* haplotype observed in 97% of blue-eyed individuals in a present-day study population from Turkey, Jordan, and Denmark[25]. Due to a combination of missing and heterozygous sites in our unphased genotype data, the haplotypes of the Motala12 forager and Stuttgart farmer could not be conclusively determined. The identification of h-1 in one of the earliest reported carriers of the rs1291382 derived allele supports the inference that h-1 is the founder haplotype for the blue-eye mutation[25].

Examining the *SLC24A5* region, we find that the Stuttgart farmer is homozygous for the C11 haplotype found in 97% of all modern carriers of the derived rs1426654 pigmentation-lightening allele[27]. The A111T mutation is estimated to have arisen at ~22-28 kya[28], with the selective sweep favoring its rise beginning ~19kya (under a dominant model) or ~11kya (under an additive model)[29]. The Loschbour forager does not carry the derived rs1426654 allele. The Motala12 forager, like the Stuttgart farmer, is homozygous for the C11 haplotype. Although three of the SNPs defining C11 were

genotyped at 1× coverage in the Motala12 sample, C11 is the only haplotype matching the possible patterns of variation.

Metabolic Syndrome Risk Score

Complex human disease phenotypes are less amenable to genotype-based prediction than externally visible characteristics such as pigmentation. The diabetes risk scoring systems developed to date thus do not have strong predictive power at the population level [11]. Nevertheless, we used these scoring systems to begin to characterize the metabolic genotypes of the Loschbour and Stuttgart ancient modern humans in comparison with the average present-day non-diabetic genotype.

We evaluated the Loschbour and Stuttgart individuals using two different type 2 diabetes (T2D) risk score models (Motala12 could not be included in this analysis because of inadequate coverage at crucial SNPs) (Table S8.6). We find that the two ancient modern humans display metabolic syndrome-associated allele spectra comparable to those observed in present-day Europeans.

The Meigs 2008 model indicates a higher T2D risk for the Loschbour individual relative to Stuttgart. The weighted genotype risk scores for both Loschbour and Stuttgart fall within the overlapping one standard deviation ranges of the present-day diabetic and non-diabetic ranges predicted by this model.

The Cornelis 2009 model predicts a roughly equal risk for both individuals. According to this model, the weighted genotype risk scores of both the Loschbour and Stuttgart individuals (10.7 and 10.6, respectively) are within the 95% CI of that of the median present-day non-diabetic individual (10.4).

Overall, the risk allele is the ancestral allele at 19 out of the 35 MetS-associated SNPs whose genotypes we evaluated. The Loschbour and Stuttgart individuals carried similar numbers of ancestral MetS-associated risk alleles (21 for Loschbour and 19 for Stuttgart), and derived MetS risk alleles (14 for Loschbour and 15 for Stuttgart). Moreover, the MetS risk scores of the ancient forager and farmer do not indicate any significant departures from the MetS risk score averages in present-day Europeans.

We also genotyped the three ancient modern humans at 5 SNPs in the *SLC16A11* gene comprising a haplotype associated with type 2 diabetes risk in a present-day Latin American population and for which the risk haplotype is believed to derive from Neanderthal introgression[30] (Table S8.7). None of the three ancient modern humans carried the risk haplotype.

Other phenotypic characteristics

We also assessed the Loschbour, Motala12, and Stuttgart individuals for their genotype at nine SNPs with well-validated phenotypic associations and evidence for recent positive selection (Table S8.8).

All three ancient modern humans are homozygous for the ancestral alleles at the *LCTa* and *LCTb* polymorphisms and as a result are predicted to have been unable to digest lactose as adults. The *LCTa* mutation has been estimated to have first experienced positive selection between 6,256 and 8,683 years ago in central Europe [31]. Thus, although the allele is associated with the spread of the LBK culture, it is likely to have been uncommon in early LBK populations, consistent with our results.

The heterozygous state of both the Stuttgart and Loschbour individuals at a SNP in the *AGT* gene suggests that they may have had a slightly increased risk of hypertension (Motala12 could not be

genotyped at this locus). The risk allele in the *AGT* gene is an ancestral allele. The derived protective allele is estimated to have arisen 22,500-44,500 years ago [16].

All three ancient modern humans were homozygous for a derived allele at rs2740574 in *CYP3A4*, which is thought to confer protection from certain forms of cancer and is also possibly associated with protection from rickets [32]. Loschbour and Stuttgart are also homozygous for the derived allele at rs776746 in *CYP3A5*, which is estimated to have arisen ~75,000 years ago [33], and which affects drug metabolism (Motala12 could not be genotyped at this locus).

We additionally evaluated the three ancient modern humans for their genotypes at SNPs in *EDAR*, *ADH1B*, *ABCC1*, and *ALDH2* that are known to have high allele frequency differentiation between present-day Europeans and East Asians. All three individuals are homozygous for alleles associated with wet earwax (*ABCC1*) and non-shoveled incisors (*EDAR*), which are phenotypes known to occur at higher frequency in Europeans [34-36]. Both the Loschbour forager and the Stuttgart farmer carried the ancestral alleles at *ALDH2* or *ADH1Ba*, two loci associated with alcohol metabolism which are known to have been under recent positive selection in East Asian populations[18,19,37] . The derived alleles at these SNPs are associated with slower alcohol metabolism and reduced alcohol consumption.The Motala12 forager could not be genotyped at *ALDH2* or *ADH1Ba*, but was homozygous for the ancestral allele at *ADH1Bb*. The Stuttgart individual was also homozygous for the ancestral allele at a third alcohol-metabolism locus under recent positive selection, *ADH1Bb(Arg48His)*[37], (the Loschbour and Motala12 individuals could not be conclusively genotyped at this locus).

Finally, we assessed the genotypes of the three ancient modern humans at 8 SNPs in the *NAT2* gene in order to determine the acetylation phenotype (rapid>intermediate>slow) of each individual. *NAT2* is involved in the metabolism of a wide variety of xeniobiotics, including a number of carcinogens, and there is evidence from present-day populations for selection favoring the slow-acetylator phenotype, possibly related to dietary changes accompanying the transition to agriculture. [38,39] We inferred acetylation status using three partially independent methods: the 4 SNP panel proposed by Hein *et al.*[40]*;* the NAT2pred online tool (NAT2pred.rit.albany.edu)[41], which consists of the 4 SNP panel plus 3 additional SNPs; and a tag SNP which is in strong LD with a 7 SNP panel in present-day Europeans[42].The three inference methods we employed agreed that the Stuttgart farmer was most likely to have been a slow acetylator, the Loschbour forager was an intermediate acetylator, and the Motala12 forager was a rapid acetylator. The observation of a slow-acetylator phenotype in the sample from an early agriculturalist population supports the inference that selection on the *NAT2* region may be related to the adoption of farming.

Table S8.1. Loshbour, Stuttgart and Motala12 genotypes for SNPs associated with pigmentation

8plex SNPs				
SNP	***Gene***	**Loschbour**	**Stuttgart**	**Motala12**
rs1291382	*HERC2*	G/G	A/A	G/G (1x)
rs1545397	*OCA2*	A/A	A/A	°
rs16891982	*SLC45A2*	C/C	C/C	°
rs885479	*MC1R*	G/G	G/G	G/G (3x)
rs1426654	*SLC24A5*	G/G	A/A	A/A (3x)
rs12896399	*SLC24A4*	G/G*	T/T	°
rs6119471	*ASIP*	C/C	C/C	C/C (3x)
rs12203592	*IRF4*	T/T	C/C	T/T (3x)
Hirisplex SNPs				
SNP	**Gene**	**Loschbour**	**Stuttgart**	**Motala12**
n29insa	*MC1R*	C/C	C/C	C/C (5x)
rs11547464	*MC1R*	G/G	G/G	°
rs885479	*MC1R*	G/G	G/G	G/G (3x)
rs1805008	*MC1R*	C/C	C/C	C/C (2x)
rs1805005	*MC1R*	G/G	G/G	G/G (6x)
rs1805006	*MC1R*	C/C	C/C	C/C (6x)
rs1805007	*MC1R*	C/C	C/C	°
rs1805009	*MCIR*	G/G	G/G	G/G (4x)
y152och	*MC1R*	C/C	C/C	°
rs2228479	*MC1R*	G/G*	G/G	G/G (4x)
rs1110400	*MC1R*	T/T	T/T	°
rs28777	*SLC45A2*	C/A	C/C	A/A (4x)
rs16891982	*SLC45A2*	C/C	C/C	°
rs12821256	*KITLG*	T/T	T/T	T/T (3x)
rs4959270	*EXOC2*	A/A	C/C	A/A (1x)
rs12203592	*IRF4*	T/T	C/C	T/T (3x)
rs1042602	*TYR*	C/C	C/A	C/C (3x)
rs1800407	*OCA2*	C/C	C/C*	C/C (1x)
rs2402130	*SLC24A4*	G/A	A/A	A/A (1x)
rs12913832	*OCA2/HERC2*	G/G	A/A	G/G (1x)
rs2378249	*PIGU/ASIP*	A/A	A/A	A/A (1x)
rs12896399	*SLC24A4*	G/G*	T/T	°
rs1393350	*TYR*	G/G	G/G	G/G (4x)
rs683	*TYRP1*	A/A	A/A	A/A (1x)

*Coverage at each position is given in parentheses for the Motala12 sample. *These SNPs had genotype quality between 20 and 30, but passed other quality filters. °These SNPs could not be genotyped.*

Table S8.2. Hirisplex model probability scores for pigmentation.

	Loschbour	Stuttgart
HAIR	**Probability**	**Probability**
Brown	0.413	0.220
Red	0	0
Black	0.579	0.774
Blond	0.008	0.005
HAIR SHADE	**Probability**	**Probability**
Light	0.022	0.006
Dark	0.978	0.994
EYE	**Probability**	**Probability**
Blue	0.613	0
Intermediate	0.166	0.004
Brown	0.222	0.996

Table S8.3. OCA2/HERC2 haplotypes observed in the Loschbour, Motala, and Stuttgart individuals

Haplotype	Blue-eye allele	SNP	Loschbour	Stuttgart	Motala12
BEH1	A	rs4778138	A/A	A/G	A/A (3x)
BEH1	C	rs4778241	C/C	A/C	C/C (5x)
BEH1	A	rs7495174	A/A*	A/A	A/A (1x)
BEH2	T	rs1129038	T/T*	C/C	T/T (3x)
BEH2	G	rs1291382	G/G	A/A	G/G (1x)
BEH3	C	rs916977	C/C	C/C	C/C (2x)
BEH3	T	rs1667394	T/T	T/T	T/T (4x)

*Genotypes of the Loschbour, Stuttgart, and Motala12 individuals at the sites comprising three haplotypes associated with blue eyes in modern populations[26]. Coverage at each position is given in parentheses for the Motala12 sample.*These sites had genotype quality between 20 and 30 but passed other quality filters.*

Table S8.4. 13-SNP OCA2/HERC2 genotypes of the Loschbour, Motala, and Stuttgart individuals

SNP	Loschbour	Stuttgart	Motala12
rs4778241	C/C	A/C	C/C (5x)
rs1129038	T/T*	C/C	T/T (3x)
rs12593929	A/A	A/A	A/A (2x)
rs12913832	G/G	A/A	G/G (1x)
rs7183877	C/C	C/C	°
rs3935591	C/C	C/C	T/C (7x)
rs7170852	A/A	A/A	A/A (7x)
rs2238289	A/A	A/A	A/A (3x)
rs3940272	G/G*	°	°
rs8028689	T/T	T/T	T/T (2x)
rs2240203	T/T	T/T	T/T (5x)
rs11631797	G/G	G/G*	°
rs916977	CC	CC	C/C (2x)
Haplotype	**h-1**	*****	*****

Genotypes of the Loschbour, Stuttgart, and Motala12 individuals at the 13 sites comprising the haplotype found at high frequency in present blue-eyed individuals, along with haplotype asignment[25]. Coverage at each position is given in parentheses for the Motala12 sample. These SNPs had genotype quality <30 but passed other quality filters; °the individual could not be genotyped at this locus.*

Table S8.5. 16-SNP SLC25A5 genotypes of the Loschbour, Motala12, and Stuttgart individuals

SNP	Loschbour	Stuttgart	Motala12
rs1834640	A/G	A/A	A/A(2x)
rs2675345	A/G*	A/A	A/A(3x)
rs2469592	A/G	A/A	A/A(4x)
rs2470101	T/C	T/T	T/T(5x)
rs938505	C/T	C/C	C/C(2x)
rs2433354	C/T	C/C	C/C(1x)
rs2459391	A/G	A/A	A/A(5x)
rs2433356	A/G	G/G	G/G(4x)
rs2675347	A/G	A/A	A/A(1x)
rs2675348	A/G	A/A	A/A(2x)
rs1426654	G/G	A/A	A/A(3x)
rs2470102	A/G*	A/A	A/A(5x)
rs16960631	A/A	A/A	A/A(3x)
rs2675349	A/G	A/A	A/A(1x)
rs3817315	C/T	C/C	C/C(2x)
rs7163587	T/C	C/C	C/C(4x)
Haplotype	*	**C11**	**C11**

*Genotypes of the Loschbour, Stuttgart, and Motala12 individuals at the 16 sites comprising the SLC25A5 haplotype observed in most modern humans carrying the derived rs1426654 (A111T) allele, along with haplotype assignment (SLC25A5 haplotype assignment was not possible for the Loschbour forager) [27]. Coverage at each position is given in parentheses for the Motala12 sample.*These SNPs had genotype quality <30 .*

Table S8.6. Metabolic syndrome SNPs assessed in Loschbour and Stuttgart, by risk score model.

Metabolic syndrome associated SNPs			
SNP	**Gene**	**Loschbour**	**Stuttgart**
rs7923837	*HHEX*	G/G	A/A
rs5015480	*HHEX/IDE*	C/C	T/T
rs3802678	*GBF1*	A/A	A/T
rs6235	*PCSK1*	C/C	G/G
rs7756992	*CDKAL1*	A/G	A/G
rs6446482	*WFS1*	C/G	C/G
rs11037909	*EXT2*	T/C	T/C
rs6698181	*PKN2*	T/T	C/T
rs17044137	*FLJ39370*	T/A	T/A
rs12255372	*TCF7L2*	G/G	G/G
rs7480010	*LOC387761*	A/A	A/A
rs11634397	*ZFAND6*	A/G	G/G
rs10946398	*CDKAL1*	A/C	C/C
rs8050136	*FTO*	A/A	C/A
Meigs 2008			
SNP	**Gene**	**Loschbour**	**Stuttgart**
rs7901695°	*TCF7L2*	T/T	°
rs7903146°	*TCF7L2*	C/C	C/C
rs1470579	*IGF2BP2*	A/C	A/A
rs10811661	*CDKN2A/B*	T/C	T/T
rs864745	*JAZF1*	T/C	C/C
rs5219	*KCNJ11*	*	T/C
rs5215*	*KCNJ11*	C/T	C/T
rs12779790	*CDC123/CAMK1D*	A/G	A/A
rs7578597	*THADA*	T/T	T/T
rs7754840	*CDKAL1*	G/C	C/C
rs7961581	*TSPAN8/LGR5*	T/T	C/T
rs4607103	*ADAMTS9*	C/C	C/C
rs1111875	*HHEX*	C/C	T/T
rs10923931	*NOTCH2*	G/T	G/T
rs13266634	*SLC30A8*	C/C	C/C
rs1153188	*DCD*	T/T	T/A
rs1801282	*PPARG*	C/C	C/C
rs9472138	*VEGFA*	C/C	C/C
rs10490072	*BCL11A*	T/C	T/T
rs689	*INS*	A/T	A/T
Weighted genotype risk score		**118.0**	**101.6**
Cornelis 2009			
SNP	**Gene**	**Loschbour**	**Stuttgart**
rs564398	*CDKN2A/B*	C/T	T/T
rs10010131	*WFS1*	A/G	A/G
rs7754840	*CDKAL1*	G/C	C/C
rs4402960	*IGF2BP2*	G/T	G/G
rs1801282	*PPARG*	C/C	C/C
rs5219	*KCNJ11*	*	T/C
rs5215*	*KCNJ11*	C/T	C/T
rs1111875	*HHEX*	C/C	T/T
rs13266634	*SLC30A8*	C/C	C/C
rs10811661	*CDKN2A/B*	T/C	T/T
rs7901695	*TCF7L2*	T/T	T/T°
Rs7903146°	*TCF7L2*	C/C	C/C
Weighted genotype risk score		**10.6**	**10.7**

**For the purpose of computing the Weighted Genotype Risk Score, we use rs5215 as a proxy for rs5219, which failed to pass the quality filter for the Loschbour sample. These two SNPs are in strong LD (r^2=0.90)[43]*

in present-day populations. °rs7903146 was used as a proxy for rs7901695, which for the Stuttgart individual failed to pass the quality filter. The two SNPs are in strong LD (r^2=0.98) [44] in present-day populations.

Table S8.7. SLC16A11 genotypes of the Loschbour, Stuttgart, and Motala12 individuals

SNP	Loschbour	Stuttgart	Motala12
rs75493593 (P443T)	G/G*	G/G	G/G (2x)
rs75418188 (G340S)	C/C*	C/C*	C/C (2x)
rs13342232 (L187L)	A/A	A/A	A/A (3x)
rs13342692 (D127G)	T/T	T/T	T/T (7x)
rs117767867 (V113I)	C/C	C/C	C/C (3x)

*Genotypes of the Loschbour, Stuttgart, and Motala12 individuals at 5 SNPs in the SLC16A11 gene comprising a haplotype associated with type 2 diabetes risk in a modern Latin American population[30]. Coverage at each position is given in parentheses for the Motala12 sample.*These sites had genotype quality between 20 and 30, but passed other quality filters.*

Table S8.8. Loschbour, Stuttgart, and Motala12 genotypes for SNPs known to be under selection in modern humans

SNP	Gene	Loschbour	Stuttgart	Motala12
rs182549	*LCTb*	C/C	C/C	C/C(3x)
rs4988235	*MCM6/LCTa*	G/G	G/G	G/G(2x)
rs699	*AGT*	A/G	A/G	°
rs4590952	*KITLG*	A/G	G/G	G/G(2x)
rs2740574	*CYP3A4*	T/T	T/T	T/T(3x)
rs776746	*CYP3A5*	C/C	C/C	C/C(1x)
rs3827760	*EDAR*	A/A	A/A	A/A(3x)
rs17822931	*ABCC1*	C/C	C/C	C/C (3x)
rs671	*ALDH2*	G/G	G/G	GG(1x)
rs3811801	*ADH1Ba*	G/G	G/G	G/G(2x)
rs1229984	*ADH1Bb*	°	C/C	C/C(1x)

Coverage at each position is given in parentheses for the Motala12 sample. °The individual could not be genotyped at this locus.

Table S8.9. NAT2 genotypes and inferred acetylation status of the Loschbour, Stuttgart, and Motala12 individuals

SNP	Loschbour	Stuttgart	Motala12
rs1801279 (191G>A)	G/G	G/G	GG (5x)
rs1801280 (341T>C)	T/C	C/C	TT (3x)
rs1799930 (590G>A)	G/G	G/A°	GG (4x)
rs1799931 (857G>A)	G/G	G/G*	GG (1x)
rs1495741(tag)	G/A	A/A	GG (3x)
rs1041983(282C>T)	C/C	C/C	CC (4x)
rs1799929(481C>T)	C/T°	T/T	CT (2x)
rs1208(803A<G)	G/A	G/G	AA (2x)
Acetylation status	**Intermediate**	**Slow**	**Rapid**

*Genotypes of the Loschbour, Stuttgart, and Motala12 individuals at 8 SNPs in the NAT2 gene associated with acetylation status (rapid>intermediate>slow)[40-42,45] Coverage at each position is given in parentheses for the Motala12 sample.. °These sites had Qual <30 *This site had genotype quality between 20 and 30 but passed other quality filters.*

References

1 Walsh, S. *et al.* DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Science International: Genetics* **6**, 330-340, doi:http://dx.doi.org/10.1016/j.fsigen.2011.07.009 (2012).

2 Walsh, S. *et al.* The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic science international. Genetics* **7**, 98-115, doi:10.1016/j.fsigen.2012.07.005 (2013).

3 Walsh, S. *et al.* IrisPlex: A sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics* **5**, 170-180, doi:http://dx.doi.org/10.1016/j.fsigen.2010.02.004 (2011).

4 Walsh, S. *et al.* Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Science International: Genetics* **5**, 464-471, doi:http://dx.doi.org/10.1016/j.fsigen.2010.09.008 (2011).

5 Branicki, W. *et al.* Model-based prediction of human hair color using DNA variants. *Human Genetics* **129**, 443-454, doi:10.1007/s00439-010-0939-8 (2011).

6 Hart, K. L. *et al.* Improved eye- and skin-color prediction based on 8 SNPs. *Croatian Medical Journal* **54**, 248-256, doi:10.3325/cmj.2013.54.248 (2013).

7 Spichenok, O. *et al.* Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Science International: Genetics* **5**, 472-478, doi:10.1016/j.fsigen.2010.10.005 (2011).

8 Walsh, S. *et al.* Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. *Forensic Science International: Genetics* **9**, 150-161, doi:10.1016/j.fsigen.2013.12.006 (2014).

9 Pneuman, A., Budimlija, Z. M., Caragine, T., Prinz, M. & Wurmbach, E. Verification of eye and skin color predictors in various populations. *Legal Medicine* **14**, 78-83, doi:10.1016/j.legalmed.2011.12.005 (2012).

10 Draus-Barini, J. *et al.* Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. *Investigative Genetics* **4**, 3 (2013).

11 Cornelis, M. *et al.* Joint effects of common genetic variants on the risk for type 2 diabetes in U.S. men and women of European ancestry. *Annals of Internal Medicine* **150**, 541-550 (2009).

12 Meigs, J. B. *et al.* Genotype score in addition to common risk factors for prediction of type 2 diabetes. *New England Journal of Medicine* **359**, 2208-2219, doi:10.1056/NEJMoa0804742 (2008).

13 Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nature Genetics* **40**, 340-345, doi:Doi 10.1038/Ng.78 (2008).

14 Corona, E., Dudley, J. T. & Butte, A. J. Extreme evolutionary disparities seen in positive selection across seven complex diseases. *PLoS ONE* **5**, e12236 (2010).

15 Zeron-Medina, J. *et al.* A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell* **155**, 410-422, doi:10.1016/j.cell.2013.09.017 (2013).

16 Nakajima, T. *et al.* Natural selection and population history in the human angiotensinogen gene (AGT): 736 complete AGT sequences in chromosomes from around the world. *The American Journal of Human Genetics* **74**, 898-916, doi:10.1086/420793 (2004).

17 Thompson, E. E. *et al.* CYP3A variation and the evolution of salt-sensitivity variants. *American Journal of Human Genetics* **75**, 1059-1069, doi:10.1086/426406 (2004).

18 Li, H. *et al.* Diversification of the ADH1B gene during expansion of modern humans. *Annals of Human Genetics* **75**, 497-507, doi:10.1111/j.1469-1809.2011.00651.x (2011).

19 Oota, H. *et al.* The evolution and population genetics of the ALDH2 locus: random genetic drift, selection, and low levels of recombination. *Annals of Human Genetics* **68**, 93-109, doi:10.1046/j.1529-8817.2003.00060.x (2004).

20 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158, doi:10.1093/bioinformatics/btr330 (2011).

21 Li, H. The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics.* **25**, 2078-2079 (2009).

22 Lamason, R. L. *et al.* SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. *Science* **310**, 1782-1786, doi:10.1126/science.1116238 (2005).

23 Beleza, S. *et al.* Genetic architecture of skin and eye color in an African-European admixed population. *PLoS Genetics* **9**, e1003372, doi:10.1371/journal.pgen.1003372 (2013).

24 Stokowski, R. P. *et al.* A genomewide association study of skin pigmentation in a South Asian population. *American Journal of Human Genetics* **81**, 1119-1132 (2007).

25 Eiberg, H. *et al.* 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. *Human Genetics* **123**, 177-187 (2008).

26 Donnelly, M. P. *et al.* A global view of the OCA2-HERC2 region and pigmentation. *Human genetics* **131**, 683-696 (2012).

27 Canfield, V. A. *et al.* Molecular phylogeography of a human autosomal skin color locus under natural selection. *G3: Genes|Genomes|Genetics* **3**, 2059-2067, doi:10.1534/g3.113.007484 (2013).

28 Mallick, C. B. *et al.* The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. *PLoS Genetics* **9**, e1003912 (2013).

29 Beleza, S. *et al.* The timing of pigmentation lightening in Europeans. *Molecular Biology and Evolution* (2012).

30 The Sigma Type 2 Diabetes Consortium *et al.* Sequence variants in SLC16A11 are a common risk factor for type 2 diabetes in Mexico. *Nature*, doi:10.1038/nature12828 (2013).

31 Itan, Y., Powell, A., Beaumont, M. A., Burger, J. & Thomas, M. G. The origins of lactase persistence in Europe. *PLoS Computational Biology* **5**, e1000491, doi:10.1371/journal.pcbi.1000491 (2009).

32 Schirmer, M. *et al.* Genetic signature consistent with selection against the CYP3A4*1B allele in non-African populations. *Pharmacogenetics and Genomics* **16**, 59-71 (2006).

33 Bains, R. K. *et al.* Molecular diversity and population structure at the Cytochrome P450 3A5 gene in Africa. *BMC Genetics* (2013).

34 Kimura, R. *et al.* A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *The American Journal of Human Genetics* **85**, 528-535 (2009).

35 Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918, doi:10.1038/nature06250 (2007).

36 Ohashi, J., Naka, I. & Tsuchiya, N. The impact of natural selection on an ABCC11 SNP determining earwax type. *Molecular Biology and Evolution* **28**, 849-857, doi:10.1093/molbev/msq264 (2011).

37 Peng, Y. *et al.* The ADH1B Arg47His polymorphism in East Asian populations and expansion of rice domestication in history. *BMC Evolutionary Biology* **10**, 15 (2010).

38 Patin, E. Deciphering the ancient and complex evolutionary history of human arylamine N-acetyltransferase genes. *American Journal of Human Genetics* **78**, 423-436 (2006).

39 Magalon, H. *et al.* Population genetic diversity of the NAT2 gene supports a role of acetylation in human adaptation to farming in Central Asia. *European Journal of Human Genetics* **16**, 243-251, doi:10.1038/sj.ejhg.5201963 (2008).

40 Hein, D. W. & Doll, M. A. Accuracy of various human NAT2 SNP genotyping panels to infer rapid, intermediate and slow acetylator phenotypes. *Pharmacogenomics* **13**, 31-41, doi:10.2217/pgs.11.122 (2012).

41 Kuznetsov, I. B., McDuffie, M. & Moslehi, R. A web server for inferring the human N-acetyltransferase-2 (NAT2) enzymatic phenotype from NAT2 genotype. *Bioinformatics* **25**, 1185-1186, doi:10.1093/bioinformatics/btp121 (2009).

42 Garcia-Closas, M. *et al.* A single nucleotide polymorphism tags variation in the arylamine N-acetyltransferase 2 phenotype in populations of European background. *Pharmacogenetics and Genomics* **21**, 231-236, doi:10.1097/FPC.0b013e32833e1b54 (2011).

43 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

44 Vcelak, J. *et al.* T2D risk haplotypes of the TCF7L2 gene in the Czech population sample: the association with free fatty acids composition. *Physiological Research* **61**, 229-240 (2012).

45 Selinski, S. *et al.* Genotyping NAT2 with only two SNPs (rs1041983 and rs1801280) outperforms the tagging SNP rs1495741 and is equivalent to the conventional 7-SNP NAT2 genotype. *Pharmacogenetics and Genomics* **21**, 673-678, doi:10.1097/FPC.0b013e3283493a23 (2011).