# The Genetic Architecture of Adaptations to High Altitude in Ethiopia

Gorka Alkorta-Aranburu[1], Cynthia M. Beall[2]*, David B. Witonsky[1], Amha Gebremedhin[3], Jonathan K. Pritchard[1,4], Anna Di Rienzo[1]*

1 Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America, 2 Department of Anthropology, Case Western Research University, Cleveland, Ohio, United States of America, 3 Department of Internal Medicine, Faculty of Medicine, Addis Ababa University, Addis Ababa, Ethiopia, 4 Howard Hughes Medical Institute, Chevy Chase, Maryland, United States of America

## Abstract

Although hypoxia is a major stress on physiological processes, several human populations have survived for millennia at high altitudes, suggesting that they have adapted to hypoxic conditions. This hypothesis was recently corroborated by studies of Tibetan highlanders, which showed that polymorphisms in candidate genes show signatures of natural selection as well as well-replicated association signals for variation in hemoglobin levels. We extended genomic analysis to two Ethiopian ethnic groups: Amhara and Oromo. For each ethnic group, we sampled low and high altitude residents, thus allowing genetic and phenotypic comparisons across altitudes and across ethnic groups. Genome-wide SNP genotype data were collected in these samples by using Illumina arrays. We find that variants associated with hemoglobin variation among Tibetans or other variants at the same loci do not influence the trait in Ethiopians. However, in the Amhara, SNP rs10803083 is associated with hemoglobin levels at genome-wide levels of significance. No significant genotype association was observed for oxygen saturation levels in either ethnic group. Approaches based on allele frequency divergence did not detect outliers in candidate hypoxia genes, but the most differentiated variants between high- and lowlanders have a clear role in pathogen defense. Interestingly, a significant excess of allele frequency divergence was consistently detected for genes involved in cell cycle control and DNA damage and repair, thus pointing to new pathways for high altitude adaptations. Finally, a comparison of CpG methylation levels between high- and lowlanders found several significant signals at individual genes in the Oromo.

## Introduction

Hypoxia is a major stress on human physiological processes and a powerful homeostasis system has evolved in animals to cope with fluctuations in oxygen concentration [1]. High altitude (HA) hypoxia, such as that experienced at 2500 m of altitude or greater, engages this system and elicits physiological acclimatization when lowlanders become exposed to hypoxia. In addition to lower oxygen levels, lower biodiversity and extreme day-to-night temperature oscillations challenge HA living. The classic response is an increase in hemoglobin (Hb) concentration that during acclimatization compensates for the unavoidable lowered percent of oxygen saturation ($O_2$ sat) of Hb due to ambient hypoxia. Acclimatization is not completely effective, however. For example, birth weights are lower than at low altitude (LA) as is physical exercise performance [2,3]. In addition, lowlanders residing at altitudes higher than 2500 meters (m) are at risk for chronic health problems arising in part from acclimatization processes. For example, long-term high Hb levels increase blood viscosity as well as the risk of thrombosis and stroke [4,5,6] and poorer pregnancy outcomes [7]. These results taken together suggest that the acclimatization response does not

assure that fitness is unaltered at HA. Several distantly related human populations have survived for 5–50 thousand years (ky) [8,9,10] at altitudes above 2500 m. Indeed, in most cases, sufficient time has elapsed since HA settlement for natural selection to have changed the frequency of adaptive alleles. Interestingly, even though all HA residents are exposed to the same, constant, ambient hypoxia, indigenous highlander populations show distinctive physiological characteristics thought to offset HA stress: Andeans show some reduction in $O_2$ sat, but a marked increase in Hb levels [11], Tibetans present markedly low $O_2$ sat, but relatively little increase in Hb levels [12], and Amhara in Ethiopia present little reduction in $O_2$ sat or increase in Hb levels [13,14]. Whether these phenotypic contrasts reflect different genetic adaptations across populations remains an open question.

Substantial evidence in Tibetan highlanders suggests that variation in Hb levels and $O_2$ sat is adaptive. In the Tibetan population, a major gene effect on $O_2$ sat was detected, with the inferred genotypes associated with higher $O_2$ sat also associated with higher reproductive success [15]; though the locus underlying this effect has not yet been identified, its effect on phenotypes directly related to fitness points to the presence of adaptive

## Author Summary

Although hypoxia is a major stress on physiological processes, several human populations have survived for millennia at high altitudes, suggesting that they have adapted to hypoxic conditions. Consistent with this idea, previous studies have identified genetic variants in Tibetan highlanders associated with reduction in hemoglobin levels, an advantageous phenotype at high altitude. To compare the genetic bases of adaptations to high altitude, we collected genetic and epigenetic data in Ethiopians living at high and low altitude, respectively. We find that variants associated with hemoglobin variation among Tibetans or other variants at the same loci do not influence the trait in Ethiopians. However, we find a different variant that is significantly associated with hemoglobin levels in Ethiopians. Approaches based on the difference in allele frequency between high- and lowlanders detected strong signals in genes with a clear role in defense from pathogens, consistent with known differences in pathogens between altitudes. Finally, we found a few genome-wide significant epigenetic differences between altitudes. These results taken together imply that Ethiopian and Tibetan highlanders adapted to the same environmental stress through different variants and genetic loci.

variation. With regard to Hb levels, which are surprisingly similar between Tibetan highlanders and lowlanders at sea level [11], population genetics and genotype-phenotype association analyses have identified alleles at two loci (endothelial PAS domain protein 1 (EPAS1) and egl nine homolog 1 (EGLN1)) that are consistently associated with signatures of positive natural selection and with lower Hb levels, suggesting that natural selection in Tibet favored variants that counteract the deleterious effects of long-term acclimatization [16,17,18,19].

An analysis of genome-wide genotype data in Tibetan and Andean highlanders suggested that natural selection acted on largely distinct loci in the two populations [20]. In addition, a recent study comparing Ethiopian Amhara highlanders with other ethnic groups at LA identified yet another set of candidate targets of selection [21]. However, the Tibetans remain unique with regard to the strength of the evidence for natural selection and the marked genetic effects on the Hb level phenotype.

HA populations offer a rare opportunity to investigate the impact of natural selection on the genetic architecture of adaptation because independent realizations of the adaptive process can be examined in different parts of the world. Specifically, the Ethiopian highlands offer a unique opportunity to study HA adaptation because individuals from distinct, but closely related ethnic groups have communities at HA and LA, thus allowing more informative genetic and phenotypic comparisons. In this study, we extended genomic analysis to two Ethiopian ethnic groups, Amhara and Oromo, with the goal of determining whether Tibetans and Ethiopian highlanders share the same adaptations and of elucidating the genetic bases of adaptive HA phenotypes in Ethiopia. We also measured genome-wide methylation levels to explore the contribution of epigenetic modifications to HA adaptations.

## Results

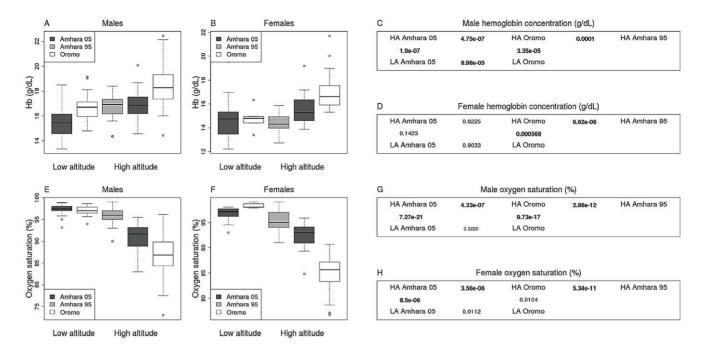### The Ethiopian Amhara and Oromo differ in adaptive phenotypes

We obtained phenotype data in two distinct, but closely related ethnic groups, the Amhara and the Oromo (Texts S1, S2; Figures S1, S2, S3, S4, S5), that include communities of HA and LA residents. All individuals were born and raised at the same altitude where they were sampled. These samples allow comparing phenotypes across altitudes within ethnic groups as well as across ethnic groups. While historical records indicate that the Oromo have moved to HA only in the early 1500 s [22,23], the Amhara have inhabited altitudes above 2500 m for at least 5 ky and altitudes around 2300–2400 m for more than 70 ky [24,25]. Therefore, sufficient time has elapsed for the Amhara to have evolved genetic adaptations to hypoxia.

As shown in Figure 1, the HA samples of both ethnic groups had higher Hb than the LA samples, however the Oromo had twice as much elevation in Hb as the Amhara. The elevation in Hb levels is particularly evident for the measurements in males, raising the possibility that other factors (e.g. menstrual cycle) in females affect the power to detect significant phenotypic differences between groups. With regard to $O_2$ sat, HA Amhara had a 5.6% lower $O_2$ sat compared to LA Amhara while HA Oromo had 10.5% lower $O_2$ sat than their LA counterparts. Therefore, we detected significant phenotypic differences not only between populations from the same ethnic group that live at different altitudes, but also across populations from closely related ethnic groups (Oromo and Amhara) that live at the same altitude. Given the low genetic divergence between these two ethnic groups at the genome-wide level (mean $F_{ST} = 0.0098$), the phenotypic differences between Amhara and Oromo highlanders are unlikely to be due to independent genetic adaptations in these ethnic groups; rather they are likely to reflect genetic adaptations that evolved in the Amhara, due to their longer residence at HA.

We also measured pulse and calculated arterial oxygen content, but these phenotypes did not show significant differences across ethnic groups or altitudes and were omitted from further analyses. For details on the phenotypic variation in Amhara and Oromo, see Text S3, Tables S1 and S2 and Figure S6.
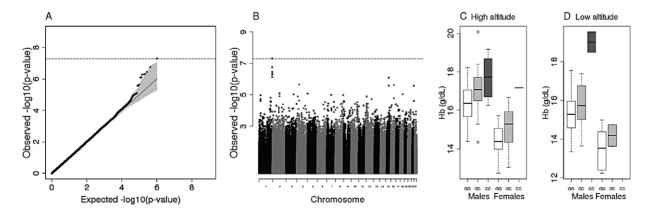
### Genome-wide association signals in Ethiopia

To learn about the genetic bases of variation in Hb level and $O_2$ sat in Ethiopia, we tested for an association between SNP genotype in the 260 unrelated Ethiopian samples and Hb levels or $O_2$ sat. We considered the total Ethiopian sample (i.e. HA and LA Amhara and Oromo) as well as each ethnic group and each altitude separately (Figures S7, S8, S9, S10, S11, S12, S13, S14, S15). No genome-wide significant signal was observed for either Hb levels or $O_2$ sat in the Oromo and in the total Ethiopian sample and for $O_2$ sat in the Amhara (Figures S7, S8, S9, S10, S11, S12, S13, S14, S15 and Tables S3, S4, S5, S6, S7, S8, S9, S10, S11, S12, S13, S14, S15, S16, S17, S18, S19, S20). Likewise, no excess of low p-values was observed in these association analyses relative to null expectations obtained by permutations (Figures S7, S8, S9, S10, S11, S12, S13, S14, S15). In contrast, the Amhara showed an excess of low association p-values in the analysis of Hb levels compared to expectations obtained by permutations (Figure 2A), indicating a genetic contribution to variation in Hb levels. In addition, one SNP (rs10803083) on chromosome 1 was associated with variation in Hb levels ($p = 4.96 \times 10^{-8}$) in Amhara; this association is genome-wide significant after correction for the 985,385 tests performed (Figure 2B and Table S3). In addition, the next six most strongly associated SNPs are in strong LD with rs10803083 ($r^2 \geq 0.69$). There are no known genes within 600 kb of these SNPs, and the closest genes, i.e. phospholipase D family member 5 (PLD5) and centrosomal protein 170 kDa (CEP170), are not obvious candidate genes for variation in Hb levels. These SNPs do not reside within an ultra conserved sequence element. Notably, the effect size of

**Figure 1. Hemoglobin and oxygen saturation measurements.** Box plots describe variation in the Amhara 05 (dark grey boxes), Amhara 95 (grey boxes) and Oromo (white boxes) for Hb concentration (g/dL) among males (A) and females (B) and for O$_2$ sat also among males (E) and females (F). Box plots show the median (horizontal line), interquartile range (box), and range (whiskers), except the extreme values represented by circles. Statistically significant differences after multiple test correction between groups (unpaired two-sided two-sample t-test) are bolded in C, D, G, and H.
doi:10.1371/journal.pgen.1003110.g001

SNP rs10803083 (~0.83 g/dL, Figure 2C) is half that of the *EGLN1* SNPs [16] but comparable to that of Hb associated *EPAS1* SNPs in Tibetans [17]. Although SNP rs10803083 reaches genome-wide significance levels, replication studies will be needed to further assess the evidence for an association with Hb levels.

Our sample size is small relative to traditional GWAS, thus liming the power of our analysis. In addition, due to the correlation between linked SNPs, the Bonferroni correction we applied is overly conservative. Therefore, many true Hb concentration variants may not reach genome-wide levels of significance. In this regard, it is interesting to note that the second strongest association signal (rs2899662) is located within the established

hypoxia-candidate gene retinoid-related orphan receptor alpha (*RORA*). *RORA* encodes a protein that induces the transcriptional activation of hypoxia-inducible-factor-1alpha (*HIF-1α*) [26], thus it plays a significant role in the same pathway where adaptations were detected in Tibetans. This SNP is, therefore, an excellent candidate for variation in Hb levels. Additional SNPs of interest for follow up analyses (Table S3) include: the solute carrier family 30 member 9 (*SLC30A9*), which is regulated by hypoxia [27], the collagen type VI alpha 1 (*COL6A1*), which is associated with performance during endurance cycling [28] and is a HIF response gene [29], and the hepatocyte growth factor (*HGF*), which is induced by hypoxia [30], activates *HIF1* DNA binding [31], plays



**Figure 2. Hemoglobin association test within Amhara.** The QQplot represents the excess of strong association with Hb among Amhara individuals (A). The observed −log10 p-value distribution is ranked from smallest to largest and plotted (y-axis) against the expected −log10 p-value (y-axis) in black. The grey area indicates the 95% confidence interval (see methods). Genome-wide (GW) significance level (after multiple test correction) is indicated by the dashed line. The Manhattan plot (B) shows the GW significance achieved by a set of high-LD SNPs in chromosome 1. The box plots describe the correlation between hemoglobin levels and the 3 genotypes of the top and GW significant SNP (rs10803083) among high (C) and low altitude (D) Amhara.
doi:10.1371/journal.pgen.1003110.g002

a role in angiogenesis and protects against hypoxia induced cell injury [32,33].

## Comparing the genetic architecture of Hb levels between Ethiopians and Tibetans

The above association analyses allow comparing the genetic architecture of Hb levels between Ethiopians and Tibetans [16,17,18].

First, we focused on the *EPAS1* and *EGLN1* SNPs that were previously associated with variation in Hb levels in Tibetans, with effect sizes of 0.8 g/dL and 1.7 g/dL, respectively [16,17]. None of these SNPs were significantly associated with Hb levels in Ethiopians (Table 1). Because we have complete or nearly complete power to detect a genotype-phenotype association in our Ethiopian samples (see Text S4 and Table S21), we infer that the SNPs associated with variation in Hb levels in the Tibetans do not make a contribution in Ethiopians.

Second, because the association signal in the Tibetans may be due to an untyped variant that is tagged by different SNPs in Tibetans and Ethiopians, we also considered all SNPs within 10 kb of the *EPAS1* and the *EGLN1* genes and repeated this analysis applying a Bonferroni correction for the number of tests performed. None of the *EPAS1* or *EGLN1* SNPs was significantly associated with Hb levels in Ethiopians. Based on power analyses (Figure 3), we can exclude associated variants with the same effect size as in Tibetans if the MAF in Ethiopia is greater than 10% and 5%, respectively, for the *EPAS1* and *EGLN1* genes (see Figure S16 for Amhara and Oromo). Therefore, this more comprehensive analysis suggests that genes shown to contribute to variation in Hb levels in Tibetans either do not influence variation in the

Ethiopian populations or if they do, their effect sizes are lower than those reported for the Tibetans.

Third, we considered all SNPs within 10 kb of the candidate genes in the "Response to Hypoxia" Gene Ontology (GO) category (26 genes). None of these SNPs is significantly associated with Hb levels after multiple test correction (p<0.05/ 1309 = $3.81 \times 10^{-5}$). Because of the larger number of SNPs tested, this analysis has a relatively high multiple testing burden. Nonetheless, we find that we have greater than 80% power to detect a SNP significantly associated with Hb levels and effect size 0.8 g/dL if its MAF is at least 20% and 100% power if its effect size is 1.7 g/dL Hb (Figure 3C and Figure S16 for Oromo and Amhara). Therefore, we conclude that variation within the "Response to Hypoxia" GO category genes is unlikely to have the same effect on Hb levels in Ethiopians as that observed in Tibetans.

## Assessing genetic differences between high and low altitude populations

A widely used family of approaches for the detection of beneficial alleles uses information about the haplotype structure around the selected site [34,35]. However, these approaches have adequate power only in the case of new advantageous alleles that were driven to high frequency by natural selection, i.e. ≥70% [34,36]. Because the largest allele frequency differences observed between HA and LA among Amhara or Oromo is less than 40%, these approaches are unlikely to be powerful in this setting. Therefore, to identify alleles that contribute to genetic adaptations to HA in Ethiopians, we used two complementary approaches that focus on the divergence of allele frequency between HA and LA

**Table 1.** Association test of Tibetan *EGLN1* and *EPAS1* SNPs within Amhara, Oromo, and combined Ethiopians.
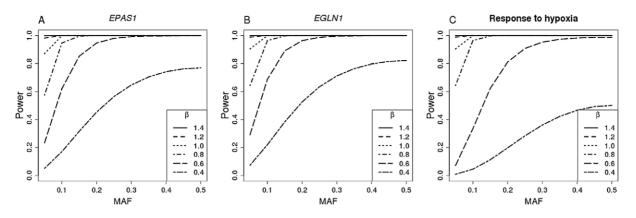
| SNP | Gene | Tibetans[1] β[2] | Amhara MAF | β[2] | p | Power[3] | Oromo MAF | β[2] | p | Power[3] | Ethiopians MAF | β[2] | p | Power[3] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rs961154 | EGLN1 | 1.70 | 0.39 | 0.08 | 0.58 | 100 | 0.41 | 0.13 | 0.57 | 100 | 0.40 | 0.12 | 0.34 | 100 |
| rs2790859 | EGLN1 | 1.70 | 0.39 | 0.08 | 0.58 | 100 | 0.41 | 0.13 | 0.57 | 100 | 0.40 | 0.12 | 0.34 | 100 |
| rs1992846 | EPAS1 | 0.84 | 0.36 | 0.00 | 0.99 | 100 | 0.42 | −0.39 | 0.13 | 100 | 0.38 | −0.14 | 0.29 | 100 |
| rs7594278 | EPAS1 | 0.52 | 0.31 | 0.06 | 0.71 | 98 | 0.24 | −0.08 | 0.77 | 80 | 0.28 | −0.01 | 0.93 | 100 |
| rs6544887 | EPAS1 | 0.79 | 0.34 | −0.05 | 0.74 | 100 | 0.34 | −0.11 | 0.63 | 100 | 0.34 | −0.07 | 0.58 | 100 |
| rs17035010 | EPAS1 | 0.84 | 0.39 | −0.07 | 0.66 | 100 | 0.30 | −0.02 | 0.93 | 100 | 0.36 | −0.04 | 0.80 | 100 |
| rs3768729 | EPAS1 | 0.80 | 0.44 | 0.10 | 0.52 | 100 | 0.44 | −0.08 | 0.74 | 100 | 0.44 | 0.04 | 0.77 | 100 |
| rs7583554 | EPAS1 | 0.94 | 0.46 | 0.10 | 0.52 | 100 | 0.41 | 0.16 | 0.53 | 100 | 0.44 | 0.13 | 0.34 | 100 |
| rs7583088 | EPAS1 | 0.92 | 0.30 | 0.10 | 0.51 | 100 | 0.22 | −0.07 | 0.79 | 100 | 0.27 | 0.02 | 0.87 | 100 |
| rs11678465 | EPAS1 | 0.85 | 0.30 | 0.10 | 0.51 | 100 | 0.22 | −0.06 | 0.83 | 100 | 0.27 | 0.03 | 0.84 | 100 |
| rs6712143 | EPAS1 | 0.94 | 0.40 | 0.16 | 0.28 | 100 | 0.32 | −0.09 | 0.74 | 100 | 0.37 | 0.09 | 0.52 | 100 |
| rs4953342 | EPAS1 | 0.90 | 0.21 | −0.12 | 0.50 | 100 | 0.12 | −0.19 | 0.57 | 99 | 0.18 | −0.16 | 0.33 | 100 |
| rs2121266 | EPAS1 | 1.02 | 0.38 | 0.14 | 0.34 | 100 | 0.30 | −0.20 | 0.43 | 100 | 0.35 | 0.01 | 0.95 | 100 |
| rs9973653 | EPAS1 | 0.52 | 0.35 | 0.01 | 0.95 | 100 | 0.30 | −0.18 | 0.47 | 100 | 0.33 | −0.05 | 0.70 | 100 |
| rs1374749 | EPAS1 | 0.88 | 0.47 | −0.18 | 0.18 | 100 | 0.49 | −0.29 | 0.20 | 100 | 0.48 | −0.22 | 0.07 | 100 |
| rs4953353 | EPAS1 | 0.97 | 0.31 | −0.01 | 0.93 | 100 | 0.31 | −0.19 | 0.46 | 100 | 0.31 | −0.08 | 0.54 | 100 |
| rs6756667 | EPAS1 | 0.93 | 0.37 | 0.02 | 0.89 | 100 | 0.32 | 0.04 | 0.86 | 100 | 0.35 | 0.03 | 0.82 | 100 |
| rs7571218 | EPAS1 | 0.71 | 0.50 | 0.17 | 0.22 | 100 | 0.49 | −0.19 | 0.46 | 100 | 0.50 | 0.06 | 0.64 | 100 |

[1]Genotype-phenotype association β coefficients for *EGLN1* were obtained from Simonson *et al* [16] while those for *EPAS1* were obtained from Beall *et al* [17].
[2]β indicates the observed linear coefficient for the relationship between SNP genotype and Hb levels.
[3]Power refers to the probability of detecting a significant association (p<0.05) between SNP genotype and Hb level given the MAF and the sample size in the Ethiopian populations assuming that the β coefficient is as high or higher as that observed in Tibetans.
doi:10.1371/journal.pgen.1003110.t001

**Figure 3. Power plots.** The effect of $\beta$ and MAF on the power of association tests based on the Ethiopian sample size (corrected for the number of SNPs tested within 10 kb from gene) is illustrated for *EPAS1* (A, 72 SNPs)), *EGLN1* (B, 38 SNPs) and any gene within the Response to Hypoxia gene ontology category (C, 1309 SNPs).
doi:10.1371/journal.pgen.1003110.g003

populations. One of these approaches was previously used to successfully identify adaptive alleles in Tibetan highlanders [18].

The first approach is based on the population branch statistic (PBS) [18], which summarizes information about the allele frequency change ($PBS_{A\_BC}$) at a given locus in the history of a population (population A) since its divergence from two populations (population B and C) so that a high $PBS_{A\_BC}$ value represents a marked change in allele frequency on the branch leading to population A. This approach was previously used to detect advantageous alleles in Tibetans relative to Han Chinese and Europeans [18]. We tested for an excess of high allele frequency differentiation (i.e. large PBS values) on the branch leading to the Ethiopian populations in SNPs within candidate genes for response to hypoxia (i.e., genes within "Response to Hypoxia" GO category) relative to SNPs in all other genes (Table S22 lists all the population trios tested). Specifically, we calculated the ratio of the proportion of SNPs in hypoxia genes *versus* the proportion of SNPs in all other genes in the top 0.5%, 1% and 5% of the distribution of PBS values and used bootstrap resampling to assess the significance of the excess of large PBS values. Although an excess was observed in most population trios (Table S22), this excess was rarely statistically significant; this finding suggests that levels of linkage disequilibrium in hypoxia genes tend to be higher than in other genes and that this feature may be a confounder in tests for selection [18]. A significant excess of large PBS values in hypoxia genes was observed only in the HA Amhara and the entire Amhara sample (Table S22 and Figure S17A and S17B), thus suggesting that HA Amhara indeed evolved genetic adaptations to hypoxic environments. When we extended this analysis of these same population trios to additional gene classifications (i.e. BioCarta, KEGG, Gene Ontology), we found significant enrichments for SNPs in gene sets related to cell cycle control, response to DNA damage and DNA repair (Table 2).

Interestingly, however, the SNPs with the highest PBS values are found in genes with a well-established role in pathogen response (Tables S23 and S24). More specifically, the SNPs with the highest PBS values are located within the major histocompatibility complex class II DR alpha (*HLA-DRA*). Moreover, the null allele (*FY\*0*) at the Duffy blood group locus, which protects against *Plasmodium vivax* malaria [37] and predicts white blood cell and neutrophil counts [38], has the second highest value. Consistent with expectations based on the protective effects of the *FY\*0* allele against malaria, its frequency is lower at HA compared to LA, where malaria is endemic (51.5% *versus* 74.1%). Therefore, these results suggest that,

in Ethiopian populations, differences in pathogen loads between LA and HA environments result in stronger selective pressures compared to differences in oxygen levels.

SNPs with large, even though not extreme PBS scores and lying within genes known to play an important role in hypoxia are of potential interest for follow up studies. These genes include: Cullin3 (*CUL3*), which potentiates *HIF-1* signaling [39], as well as adrenergic beta receptor kinase 1 (*ADRBK1*) [40], coronin actin binding protein 1B (*CORO1B*) [41], anti-silencing function 1 homolog B (*ASF1B*) [42] and MAPK-activated protein kinase MK2 (*MAPKAPK2*) [43], which are all down-regulated under hypoxia (Tables S23 and S24). None of those large PBS SNPs were significantly associated with Hb or $O_2$ sat (Tables S25 and S26), but a SNP within utrophin A (*UTRN*) - rs7753021 - reached nominal levels of significance with $O_2$ sat ($p = 0.005$; Table S26). *UTRN* expression correlates with oxidative capacity [44] and increases with chronic physical training [45]. Slow-twitch muscles, which are associated with endurance performance, have high levels of *UTRN* [46].

In a complementary analysis, we developed a multiple regression (MR) approach to identify SNPs that show high allele frequency differentiation in HA populations relative to predictions based on a large set of worldwide population samples. This method should also be able to predict allele frequencies appropriately in a situation where the target population is admixed. In this approach, we used allele frequency data from 61 LA populations (including the HGDP and several other populations) to predict the expected allele frequencies in the HA Amhara. We focused on the HA Amhara because they have lived at HA for a longer period of time and exhibit distinct patterns of Hb and $O_2$ sat levels compared to the Oromo (Figure 1). In addition, we omitted the LA Amhara in an attempt to reduce the effect of gene flow between altitudes, which could potentially reduce our power to detect adaptive divergence. We used all SNPs to estimate the best-fitting regression coefficients for each population: that is, these are the coefficients that generate the lowest mean square error in predicting the HA Amhara allele frequencies. The populations with the largest regression coefficients in the Amhara regression model are from geographically proximate populations in East Africa (Maasai, Luhya and LA Oromo) and from the Middle East and Southern Europe (see Figure S18). We reasoned that changes in allele frequencies due to high altitude adaptation would be detectable as departures (i.e. large residuals) from the predicted allele frequencies based on all other populations

As for the PBS analysis, we tested for an excess of SNPs with high allele frequency differentiation using the MR statistic for

**Table 2.** Biological pathways for which a significant excess of genic relative to all other genic SNPs are observed for all three tail cut-offs of the Amhara PBS, Amhara High PBS, or Amhara High MR distributions, respectively.

| Classification | Gene Set | Amhara PBS tail cut-off | | | Amhara High PBS tail cut-off | | | Amhara High MR tail cut-off | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 0.01 | 0.01 | 0.05 | 0.01 | 0.01 | 0.05 | 0.01 | 0.01 | 0.05 |
| GO BP | RESPONSE TO HYPOXIA | 3.40* | 3.41** | 1.84* | 3.31 | 2.9 | 1.92* | 2.58 | 2.2 | 1.75** |
| GO BP | CELL CYCLE CHECKPOINT | 5.72** | 3.40** | 1.83** | 0.48 | 0.96 | 0.96 | 0.68 | 0.36 | 0.9 |
| GO BP | DNA INTEGRITY CHECKPOINT | 6.30** | 3.86** | 2.00* | 0 | 0.82 | 0.94 | 0 | 0.07 | 0.85 |
| GO BP | DNA DAMAGE RESPONSE AND SIGNAL TRANSDUCTION | 4.83** | 2.91** | 1.59* | 2.02 | 1.79 | 1.07 | 1.12 | 1.26 | 1.18 |
| GO BP | DNA DAMAGE CHECKPOINT | 6.47** | 4.00** | 1.97** | 0 | 0.9 | 0.94 | 0 | 0.09 | 0.81 |
| GO BP | SENSORY ORGAN DEVELOPMENT | 10.29* | 6.08* | 2.31* | 2.72 | 1.36 | 0.87 | 1.9 | 0.95 | 1.04 |
| BioCarta | SALMONELLA PATHWAY | 15.61* | 9.63** | 2.89* | 0 | 0 | 0.91 | 0 | 0 | 0.31 |
| KEGG | CELL CYCLE | 1.1 | 1.1 | 1.39* | 2.93** | 2.70** | 1.60** | 4.39* | 3.35* | 1.34 |
| KEGG | CHRONIC MYELOID LEUKEMIA | 1.23 | 1.24 | 1.34 | 2.72* | 2.02** | 1.44* | 2.67* | 2.09* | 1.1 |
| GO BP | CELLULAR RESPIRATION | 0 | 1.01 | 0.69 | 4.55* | 3.07* | 2.20** | 1.11 | 2.25 | 2.42* |
| GO BP | AXON GUIDANCE | 0.35 | 0.62 | 0.8 | 1.97* | 1.72** | 1.28* | 0.47 | 0.95 | 0.93 |
| GO BP | CHROMOSOME ORGANIZATION AND BIOGENESIS | 0.69 | 0.77 | 0.87 | 1.46 | 2.22* | 1.46* | 3.93** | 3.17** | 1.99** |
| GO BP | DNA REPAIR | 1.76 | 1.2 | 1.22 | 1.83 | 1.6 | 1.36* | 2.23* | 2.41** | 1.64** |
| GO BP | ESTABLISHMENT AND OR MAINTENANCE OF CHROMATIN ARCHITECTURE | 0.73 | 0.91 | 0.83 | 1.78 | 2.09 | 1.39* | 3.31* | 2.57* | 1.74** |
| GO BP | HISTONE MODIFICATION | 1.76 | 1.76 | 0.95 | 1.58 | 2.28 | 1.26 | 5.58* | 4.57* | 2.34** |
| GO BP | CHROMATIN MODIFICATION | 0.96 | 0.97 | 0.84 | 2.4 | 2.4 | 1.37 | 4.29* | 3.34* | 1.75* |
| GO BP | ALCOHOL METABOLIC PROCESS | 1.44 | 1.18 | 0.93 | 2.1 | 1.84 | 1.15 | 3.48** | 2.51* | 1.53* |

\* and \*\*denote support from ≥95% and 99% of bootstrap replicates, respectively.

doi:10.1371/journal.pgen.1003110.t002

genes within "Response to Hypoxia" GO category relative to SNPs in all other genes and we used a bootstrap procedure to assess the significance of the observed excess. An excess was observed for all tail cut-offs, but only one reached statistical significance (Table 2). Other gene sets that showed a significant enrichment of SNPs with strong MR signals include chromosome organization and biogenesis, DNA repair, histone modification. These findings are consistent with the pattern observed in the PBS analysis, indicating that they are robust to the choice of populations used in the test (Table 2).

Among the SNPs with the largest MR scores, there are several SNPs in hypoxia genes, which may be of potential interest for follow up studies (Table S27). The SNP (rs12510722), which shows the 4th highest MR scores, lies within the alcohol dehydrogenase 6 (*ADH6*) gene, whose expression is affected by pseudohypoxia [47]. The SNPs with the 8th and the 15th highest MR score (rs2660342 and rs2660343, respectively) are within 100 kb from solute carrier family 30 member 9 (*SLC30A9*) and transmembrane protein 33 (*TMEM33*). *SLC30A9* is up-regulated by hypoxia [27] while *TMEM33* is down-regulated under hypoxia and up-regulated after knockdown of HIF1A [41].

## Assessing epigenetic differences between high and low altitude populations

Methylation is an epigenetic modification that is known to play a crucial role in the cellular response to hypoxia [48]. Since HA adaptation could be in part maintained by methylation, we measured methylation levels at 27,578 CpG sites in 17 HA and 17 LA Amhara and 17 HA and 17 LA Oromo. CpG methylation levels were tested in DNA extracted from blood in the Amhara and from saliva in the Oromo. To avoid confounding due to differences in methylation across tissues, we performed the comparison across altitudes within each ethnic group.

In Oromo, four CpG sites reached significance after multiple test correction ($p<1.85\times10^{-6}$), but the closest genes are not known hypoxia candidate genes: apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G (*APOBEC3G*), metallothionein 1G (*MT1G*), paired-like homeodomain 2 (*PITX2*) and olfactory receptor family 2 subfamily K member 2 (*OR2K2*) (Table S28). Interestingly, *APOBEC3G* codes for a well-established cellular antiviral protein and a specific inhibitor of human immunodeficiency virus-1 (HIV-1) infectivity [49,50]. The *MT1G* gene also has a role in HIV-1 infection because it upregulates MT1G expression in immature dendritic cells, which in turn facilitates the expansion of HIV-1 infection [51]. Although the prevalence of HIV was not surveyed in our fieldwork, HIV/ AIDS is known to be a major health problem in Ethiopia [52,53]. While these functions for *APOBEC3G* and *MT1G* point to a role for methylation in defense against pathogens, *MT1G* also plays a role in the response to hypoxia as its promoter is induced by vascular endothelial growth factor (VEGF), which in turn contributes to the prosurvival and angiogenic functions of VEGF [54]. Likewise, expression of *PITX2* is required for normal hematopoiesis [55,56], raising the interesting scenario that methylation of this gene may influence beneficial phenotypes in the response to hypoxia. Some of the CpG sites with nominally significant ($p<6.7\times10^{-5}$) differences in methylation between HA

and LA are close to genes that are differentially expressed in response to hypoxia; these genes include: toll-like receptor 6 (*TLR6*) [57], mif two 3 homolog 1 (*SUMO1*) [27]; phosphodiesterase 4A (*PDE4A*) [58] and human immunodeficiency virus type I enhancer binding protein 2 (*HIVEP2*) [42].

In Amhara, no CpG site showed a methylation difference that reached significance after multiple test correction (Table S29). However, we note that the $3^{rd}$ most significant differentially methylated CpG site ($p = 8.03 \times 10^{-05}$) was closest to Glutathione-S-Transferase (*GSTP1*) whose expression is increased by prolonged hypoxia [59] and whose loss of expression correlates with methylation in prostate cancer [60]. Hypoxia also regulates the expression of genes close to other differentially methylated CpG sites in Amhara ($p < 1.5 \times 10^{-3}$): protein regulator of cytokinesis 1 (*PRC1*) [42], protein tyrosine phosphatase receptor type O (*PTPRO*) [41], ring finger protein 146 (*RNF146*) [27] and Ras-related GTP binding D (*RRAGD*) [41].

Finally, no significant excess of methylation differences between LA and HA populations was observed at the genome-wide level in Oromo or Amhara (Figure S19) nor did we find a significant enrichment of methylation differences between LA and HA populations for gene sets defined by BioCarta or KEGG pathways and by Gene Ontology categories (data not shown).

## Discussion

HA human populations across the world allow studying independent realizations of the adaptive process in response to the same selective pressure, i.e. hypoxia, thus providing an excellent opportunity to investigate how natural selection shapes the genetic architecture of adaptive traits. To make progress on these enduring questions, we have sampled two closely related ethnic groups in the Ethiopian highlands that include both HA and LA residents, thus allowing comparisons across altitudes within and between ethnic groups. Of these two groups, the Oromo have moved to HA only 500 years ago [61,62], thus making it unlikely that genetic adaptations evolved in this group. In contrast, the Amhara have a history of HA residence of at least 5 ky and possibly as far as 70 ky [24,25]. Because previously identified selection signals [63,64,65,66] occurred within a similar period of time, including HA adaptations [18], we conclude that enough time has elapsed since the Amhara moved to HA for genetic adaptations to have taken place. Consistent with this idea, we observe significant phenotypic differences between Amhara highlanders and the more recent HA residents, i.e. the Oromo. While HA Amhara are characterized by mildly elevated Hb levels (similar to Tibetans) and no or mildly reduced $O_2$ sat [13], the HA Oromo sample resembles acclimatized lowlanders with a response characterized by elevated Hb concentration and marked reduction in $O_2$ sat. Our data indicates that Amhara and Oromo are very similar at the genome-wide level, therefore, the observed phenotypic differences are likely due to the different histories of HA occupation. In addition to these phenotypic comparisons, our genomic analyses of these two ethnic groups resulted in several important observations that shed new light on the biology of HA adaptations and that are discussed in detail below.

First, in a GWAS of Hb levels in Amhara, we find a genome-wide significant signal of association as well as an excess of low p-values. In addition, the second most strongly associated SNP is found within the *RORA* gene, which belongs to HIF1 pathway and is, therefore, an excellent candidate gene for hypoxia response phenotypes. Additional strong associations were observed in other candidate hypoxia genes, such as *COL6A1*, *SLC30A9*, and *HGF*. We looked at the Amhara data of Scheinfeldt *et al* [21] to test for

replication of the association signal at these genes. Two of them showed suggestive associations with Hb levels ($p = 0.06$ and $p = 0.15$ for *SLC30A9* and *RORA*, respectively), with β values ($-0.60$ and $1.31$, respectively) consistent with ours ($-0.67$ and $0.92$, respectively). It should be noted that the replication test was performed in only 21 Amhara samples in Scheinfeldt *et al* [21] for which age and BMI data were available and who had Hb levels within the normal range; thus, the lack of replication may well be due to the very low power of the replication sample. We note that, though variation in *EPAS1* and *EGLN1* has been consistently associated with Hb levels in Tibetans, no genome-wide significant association signal and no excess of low p-values were observed (see Figure S5 in Simonson et al [16,17]). While the signals we detected await replication, it is interesting to note that their effect sizes are as high as those found in Tibetans for SNPs in *EPAS1*, thus raising the interesting scenario that selection may have favored alleles with similar effect sizes on Hb levels even though the specific loci contributing to the trait are different.

Some interesting patterns are beginning to emerge with regard to the genetic contribution to variation in Hb levels and $O_2$ sat, the two phenotypes that have been most widely studied in highlander populations. No evidence of a genetic contribution to $O_2$ sat in Amhara, Oromo, and the combined Ethiopian sample could be detected (Figures S7, S8, S9, S10, S11, S12, S13, S14, S15). This is true also in the Tibetans, even though segregation analysis detected a major $O_2$ sat locus, which is also associated with reproductive success [15]. Therefore, the data so far suggest that while genetic factors contribute to variation in Hb levels, their importance in $O_2$ sat is lower. This is consistent with studies in Tibetans and Andeans showing a markedly lower heritability for $O_2$ sat compared to Hb levels; indeed, the $O_2$ sat heritability in Andeans was not significantly different from zero [67,68,69]. More data, especially at the genome-wide level, are needed to elucidate the contribution of genetic factors to these two phenotypes.

Second, to compare the genetic bases of Hb variation with the Tibetans, we tested for an association between SNP genotypes and Hb levels within the Ethiopians. Although we had appropriate power, none of the SNPs within 10 kb of the *EPAS1* and *EGLN1* genes or the genes in the hypoxia pathway, including SNPs previously associated with Hb variation (and signatures of natural selection) in Tibetans, associated with Hb. Therefore, we can rule out that the SNPs and loci contributing to Hb variation and showing selection signals in Tibetans affect the same trait in Ethiopians even though Tibetans and Amhara have lower Hb levels compared to all other highlanders. Alternatively, if the same variant affects Hb levels in both populations, their effect sizes in Amhara must be markedly lower than those reported for the Tibetans.

Third, by using approaches based on allele frequency divergence, we find that outlier SNPs in the HA Amhara sample are not in hypoxia response genes, but in loci known to play a role in immune defense. These include variation in the *HLA-DRA* locus and the null allele at the Duffy blood group locus; none of these variants is associated with Hb levels or $O_2$ sat. Interestingly, malaria and schistosomiasis were prevalent in the LA, but not in the HA Amhara communities sampled in this study (Text S1), reflecting important differences in pathogens between HA *versus* LA environments. Indeed, epidemiological studies in the areas near the LA sampling sites for the Amhara and Oromo reported malaria prevalence of 39.6% and as much as 25% of malaria morbidity is due to *P. vivax* [70,71,72]. Therefore, the immune defense variants with extreme frequency divergence represent excellent candidates as selection targets. These findings are important in several respects. First, they indicate that, because

HA and LA habitats differ by multiple environmental stresses, signals of allele frequency divergence cannot be unambiguously attributed to hypoxia without additional information about the gene function or the specific phenotypic effects of the alleles. Second, they further corroborate that the Amhara populations are indeed adapted to spatially-varying selective pressures, despite likely high levels of gene flow between HA and LA communities. Third, the fact that variants in hypoxia response genes are not outliers in these analyses suggests that adaptations to pathogens and to hypoxia have a different genetic architecture or that the intensity of pathogen-related selective pressures is stronger than those due to hypoxia. Overall, these findings highlight the opportunities and challenges of ecological genomic studies and point to the power of approaches that use environmental information combined with phenotypic data collected in the field.

Although selection has not created extreme HA *versus* LA frequency shifts in hypoxia genes, we find the SNPs within candidate genes for response to hypoxia *as a group* show an excess of allele frequency differentiation based on the PBS analysis performed in the Amhara (Table 2). This approach, however, requires specification of a set of three populations to be tested. When we used our MR approach, which uses information from all populations simultaneously, the response to hypoxia genes did not show a significant excess (Table 2). In contrast, both approaches detected a significant excess of allele frequency divergence in genes involved in cell cycle control, DNA repair, DNA damage, chromatin structure and modification, consistent with the known role of oxygen sensing in the regulation of cell proliferation [73]. Therefore, our findings raise the possibility that genetic variation in these pathways can contribute to adaptations to HA.

Fourth, though we did not observe a genome-wide excess of methylation differences between HA and LA samples, we found genome-wide significant signals in the Oromo in genes with a known function in pathogen defense or in the biology of hypoxia, i.e. VEGF signaling and hematopoiesis. Interestingly, no genome-wide significant differential methylation was observed in the Amhara; this may be due to the fact that DNA from different tissues was analyzed in the two ethnic groups. However, given the difference in the history of HA residence between Amhara and Oromo, it is also possible to speculate that epigenetic modifications play a role in the early phases of adaptations to new environments and that this role is replaced over time by adaptations at the genetic level.

It has been proposed that epigenetic modifications are important in ecological adaptations [74] and methylation is known to play a crucial role in the cellular response to hypoxia [48]. A previous study showed that gene expression differences observed between Moroccan populations were not explained by methylation differences on the tested 1,505 CpG sites [75]. Our study is more comprehensive having interrogated 27,578 CpG sites in a larger sample size. Though our results do not unambiguously point to a major role for methylation in the adaptations to HA, they suggest that further studies using DNA extracted from different tissues and including additional epigenetic modifications, in addition to methylation, are warranted.

In conclusion, studies of genetic variation in indigenous populations with long-time residence at HA are giving rise to a composite picture regarding the genes contributing to and the genetic architecture of HA adaptations. Clearly, different loci contribute to Hb levels in Ethiopians and Tibetans. Moreover, while some aspects (i.e. phenotypic effect sizes) of the genetic architecture of Hb levels may be similar, others (i.e. the allele frequency shifts due to selection) are different. Additional examples of environmental pressures that acted on different human populations include malarial endemia, low UVB radiation levels, and an adult diet rich in dairy products. Detailed genome-wide studies of parallel adaptations to these selective pressures are needed to elucidate the impact of natural selection on the genetic architecture of complex adaptive traits.

## Materials and Methods

### Ethics statement

All participants in the study gave informed consent. The studies were approved by the Institutional Review Boards of Case Western Reserve University and of the University of Chicago and by the Ethiopian Science and Technology Council Ethics Review Board.

### Population samples

Samples were drawn from native residents of the Semien Mountains area of northern Ethiopia inhabited mainly by Christian Amhara and the Bale Mountains area of southern Ethiopia inhabited mainly by Muslim Oromo (also referred to as Galla, Boran and Gabbra). For each ethnic group, we sampled individuals at HA and LA. The LA samples were chosen to achieve the maximum altitude contrast, yet avoid confounding due to the presence of endemic malaria; all LA samples were from agropastoral people reporting the same ethnicity as the HA samples and no visits to altitudes above 2500 m in the past 6 months. For details on the sampled populations and their ecology, see Text S1.

DNA was extracted from blood samples provided by 192 Amhara individuals living at 3700 m in the Simien Mountains National Park or at 1200 m in the town of Zarima. Forty-seven of these individuals were sampled in 1995 and previously described [13,76], the remaining 145 individuals were sampled in a separate expedition in 2005. DNA was extracted from, saliva samples provided by 118 Oromo individuals and collected using Oragene DNA sample collection kits; 79 individuals lived at 4000 m in the Bale Mountains National Park while 39 individuals lived at 1560 m in the town of Melkibuta. The study participants were healthy (refer to Text S3 for details). Hb and $O_2$ sat of Hb were measured in all individuals. Hb was determined in duplicate using the cyanmethemoglobin technique (Hemocue Hemoglobinometer, Hemocue AB, Angelholm, Sweden), immediately after drawing a venous blood sample. $O_2$ sat was determined by pulse oximetry (Criticare Models 503 and SpO2) as the average of six readings taken 10 seconds apart.

### SNP genotyping and imputation

The samples were genotyped using Hap650Yv3 (n = 46), Human1M-duoV3 (n = 112), and Human Omni-Quad1 (n = 160) Illumina arrays at Southern California Genotyping Consortium. Nineteen samples with less than 93% genotype call rate were omitted from the analysis. We used the program RELPAIR 2.0. [77] to test for hidden relatedness in the samples by using 3 independent sets of 1800 autosomal and 200 X-linked SNPs; individuals with relationships closer than first cousins to any other individual in the sample were omitted from the analysis. After applying these filters, 260 unrelated samples remained: 102 HA Amhara, 60 LA Amhara, 63 HA Oromo and 35 LA Oromo. Principal component analysis (PCA) of the genotype data did not detect any major outlier in either population (see Text S2 and Figure S3). Due to the incomplete overlap between SNPs on the three genotyping arrays, genotypes for 1,819,369 HapMap3 SNPs were imputed by using the program IMPUTE2 [78] and the HapMap populations as a reference panel: Utah residents with

Northern and Western European ancestry from the CEPH collection (CEU), Han Chinese in Beijing, China (CHB), Japanese in Tokyo, Japan (JPT), Luhya in Webuye, Kenya (LWK), Maasai in Kinyawa, Kenya (MKK), Toscani in Italy (TSI) and Yoruba in Ibadan, Nigeria (YRI) samples. A total of 1,297,134 autosomal SNPs with minor allele frequency (MAF) higher than 0 and imputation accuracy higher than 90% were used in the downstream analyses (Figure S20).

## Genotype-phenotype association

Phenotype-genotype associations were tested at each SNP with MAF>0.1 by linear regression using the whole-genome association analysis toolset in PLINK [79]. To determine whether there was an excess of low p-values taking into account the large number of tests performed, the distribution of observed p-values from the linear regression tests were compared to a null distribution obtained by permuting 100 times the phenotype (and corresponding covariate variables – see below) value across individuals and running the same linear model. The results were visualized by means of quantile-quantile (QQ) plots and the 95% confidence interval (CI) was estimated by permutations. Gender, body mass index (BMI), altitude, ethnic group and year of sampling were used as covariates in the linear regression. As an alternative approach, we grouped the samples based on their gender, altitude, ethnic group and year of sampling and we quantile normalized the observed phenotypes within each group; these phenotypes were then pooled across groups before testing for an association with genotype by linear regression. This latter approach preserves information about the individual ranks without introducing a bias due to the effects of covariates. However, it does not preserve information about the phenotype values, thus leading to a reduction in power. The results of these two approaches correlated highly ($r^2>0.92$); therefore, only the results for the linear regression with the covariates are shown.

## Population genetics analyses

The population branch statistic ($PBS_{A\_BC}$) was used to summarize the amount of allele frequency change in the history of population A since its divergence from two related populations, B and C [18]. In a complementary approach based on multiple regression (MR), we used the allele frequencies for the LA Oromo and world-wide population samples (i.e., Human Genome Diversity Project (HGDP) panel populations [80] and 4 HapMap Phase III populations -LWK, MKK, TSI and Gujarati Indians in Houston, Texas (GIH) - (www.hapmap.org)) to detect SNPs whose HA Amhara frequencies deviate most from estimated frequencies. Briefly, denoting the observed allele frequencies in the $p$ populations by $X_1$, $X_2$, …, $X_p$ and the expected allele frequency within the HA Amhara by $y$, we can predict the allele frequency for the $i^{th}$ SNP using the linear model $y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + … + \beta_p X_{ip} + \varepsilon_i$. Using the data from all genotyped SNPs that overlap among the $p$ populations and applying multivariate linear regression, we found $b_0$, $b_1$, $b_2$, …, $b_p$, which are estimates of parameters $\beta_0$, $\beta_1$, $\beta_2$, …, $\beta_p$ and estimated the magnitude of the residual for the $i^{th}$ SNP as $|\varepsilon_i| = |y_i - (b_0 + b_1 X_1 + b_2 X_2 + … + b_p X_p)|$. We refer to this residual as the MR score.

To test for an enrichment of allele frequency differentiation in candidate hypoxia genes, we focused on the 28 genes belonging to the "Response to hypoxia" category in the Gene Ontology database (GO:0001666). We then compared the proportion of SNPs within 10 kb of each gene in this category to that of SNPs within 10 kb of all other genes in the tail of the PBS and MR score distributions. Given the arbitrary nature of choosing a single cutoff for the tail of the distribution, we set three cutoffs (5%, 1% and

0.5%). In other words, we looked at the top 5%, 1% and 0.5% of all PBS and MR values and asked whether there is an enrichment of hypoxia gene SNPs relative to all other genic SNPs for each tail cut-off. A value of 1 represents no excess and a value greater than 1 represents enrichment in the tail of the distribution. SNPs are likely to cluster along the genome due to linkage disequilibrium, thus reducing the number of independent signals contributing to an observed enrichment. To account for this possibility, we found the confidence interval for the enrichment using a bootstrap approach described in Hancock et al [81]. An enrichment of hypoxia SNPs was considered significant (with a one-tailed test) if at least 95% of the 1000 bootstrap replicates were enriched (i.e., had a ratio above 1).

## Methylation

Methylation levels were measured at 27,578 CpG sites in 17 HA and 17 LA Amhara and Oromo DNA samples, for a total of 68 individuals, using six Infinium HumanMethylation27 arrays at Southern California Genotyping Consortium. Two LA Amhara sample data were discarded due to low data quality. In each ethnic group, the HA methylation level of each CpG site was compared to corresponding LA levels using the following linear model correcting for age, gender, and the methylation array: lm (% methylation ∼ altitude + gender + age + [array]). Two comparisons were performed: (1) HA versus LA Amhara and (2) HA versus LA Oromo. As for the genotype-phenotype association, excess of differential methylation between HA and LA was tested by comparing the observed p-value distribution to the null distribution obtained by permuting 100 times the altitude label across individuals and running the same linear model. Permutation based 95% CI was estimated.

## Supporting Information

**Figure S1** $F_{ST}$-based neighbor-joining tree showing the relationships of the Ethiopians to the worldwide populations. Oromo and Amhara cluster closely together in the tree and occupy an intermediate position between African (with the exception of the Mozabites) and non-African populations. An ancestral population fixed for ancestral alleles at all SNPs was used.
(TIF)

**Figure S2** Worldwide STRUCTURE plot. Each vertical line represents an individual, and the colors comprising each line correspond to the inferred proportion of ancestry from seven ancestral populations using 57652 random autosomal SNPs.
(TIF)

**Figure S3** Scatter plot of the first two coordinates obtained by principal component analysis. A set of 13,000 random autosomal SNPs were used. HA Amhara individuals are represented in red, LA Amhara in blue, HA Oromo in orange and LA Oromo in green.
(TIF)

**Figure S4** Ethiopian STRUCTURE plot. Each vertical line represents an individual, and the colors comprising each line correspond to the inferred proportion of ancestry from three ancestral populations using 57652 random autosomal SNPs. Samples were ordered from left to right as follows: HA Amhara, LA Amhara, HA Oromo and LA Oromo.
(TIF)

**Figure S5** Population differentiation among Ethiopian subgroups. The QQplots represent difference in allele frequency as summarized by $F_{ST}$ between HA and LA Amhara (A), Oromo (B)

and Ethiopia (C). The observed $F_{ST}$ distribution is ranked from smallest to largest and plotted against the expected $F_{ST}$ in black. The expected $F_{ST}$ distribution was obtained by permuting the subgroup labels mimicking random mating. The grey area indicates the 95% confidence interval of the expected distribution (see Methods).
(TIF)

**Figure S6** Altitude differences in anthropometric and phenotypic characteristics are summarized in the left two (gray) panels on terms of effect size d. Ethnic differences are summarized in the right two (green) panels the same way. D is dimensionless and is calculated as the difference between two sample means divided by their pooled standard deviation. Comparison based on d values allows contrasting the altitude and ethnic-group differences in phenotypes independent of the units of measurement. By convention effect sizes of 0.8 or more are considered to be 'large' [82].
(TIF)

**Figure S7** Amhara Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S8** HA Amhara Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S9** LA Amhara Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S10** Oromo Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S11** HA Oromo Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S12** LA Oromo Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the

observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S13** Ethiopia Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S14** HA Ethiopia Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S15** LA Ethiopia Hb level and $O_2$ sat GWAS results. The QQplot compares the observed $-\log10$ association p-value distribution (y-axis) with an expected distribution (x-axis) in black (see Methods) for Hb (A) and $O_2$ sat (C). The grey area represents the 95% confidence interval. The Manhattan plot shows the observed $-\log10$ association p-value of SNPs for Hb (B) and $O_2$ sat (D).
(TIF)

**Figure S16** Power plots. The effect of $\beta$ and MAF on the power of association tests based on the Oromo (A–C) and Amhara (D–F) sample sizes (corrected for the number of SNPs tested within 10 kb from gene) is illustrated for *EPAS1* (A and D, 72 SNPs), *EGLN1* (B and E, 38 SNPs) and any gene within the Response to Hypoxia gene ontology category (C and F, 1309 SNPs).
(TIF)

**Figure S17** Manhattan plots of PBS and MR transformed rank. Manhattan plots with SNPs with transformed rank less than 0.1 are shown for Amhara PBS (*versus* Maasai and Luyha) (A), for high altitude Amhara PBS (*versus* low altitude Amhara and low altitude Oromo) (B), and high altitude Amhara MR (C). The transformed rank is the rank of the SNP in the corresponding distribution divided by the total number of SNPs. The horizontal line represents 0.001 transformed rank.
(TIF)

**Figure S18** Multiple linear regression coefficients. This plot shows the regression coefficients of each of the 61 populations used to predict the expected allele frequencies in the HA Amhara in the multiple linear regression analysis. Populations have been grouped in Africa, Middle East, Europe, Southwest Asia, East Asia, America and Oceania. In each group, populations have been ordered from larger to smaller coefficients.
(TIF)

**Figure S19** Quantile-quantile (QQ) plots of methylation differentiation p-value. Each QQplot compares the observed methylation differentiation p-value distribution (y-axis) with the corresponding null distribution (x-axis) – which was created by reshuffling 100 times the altitude status among the compared samples and running the same lineal model. The grey area represents the 95% confidence interval. No evidence of genome-wide HA *versus* LA epigenetic adaptation was observed within Oromo (A) and Amhara (B).
(TIF)

**Figure S20** Frequency distribution of imputation accuracy. SNPs with accuracy values less than 0.9 were removed from further analyses.
(TIF)

**Table S1** Sample description for Amhara and Oromo high and low altitude (HA and LA) males and females (mean ± SEM).
(PDF)

**Table S2** Study phenotypes for Amhara and Oromo high and low altitude (HA and LA) males and females (mean ± SEM).
(PDF)

**Table S3** 20 SNPs with lowest Hb association p-values within Amhara.
(PDF)

**Table S4** 20 SNPs with lowest oxygen saturation association p-values within Amhara.
(PDF)

**Table S5** 20 SNPs with lowest hemoglobin association p-values within high altitude Amhara.
(PDF)

**Table S6** 20 SNPs with lowest oxygen saturation association p-values within high altitude Amhara.
(PDF)

**Table S7** 20 SNPs with lowest hemoglobin association p-values within low altitude Amhara.
(PDF)

**Table S8** 20 SNPs with lowest oxygen saturation association p-values within low altitude Amhara.
(PDF)

**Table S9** 20 SNPs with lowest hemoglobin association p-values within Oromo.
(PDF)

**Table S10** 20 SNPs with lowest oxygen saturation association p-values within Oromo.
(PDF)

**Table S11** 20 SNPs with lowest hemoglobin association p-values within high altitude Oromo.
(PDF)

**Table S12** 20 SNPs with lowest oxygen saturation p-values within high altitude Oromo.
(PDF)

**Table S13** 20 SNPs with lowest hemoglobin p-values within low altitude Oromo.
(PDF)

**Table S14** 20 SNPs with lowest oxygen saturation p-values within low altitude Oromo.
(PDF)

**Table S15** 20 SNPs with lowest hemoglobin p-values within total Ethiopian sample.
(PDF)

**Table S16** 20 SNPs with lowest oxygen saturation p-values within total Ethiopian sample.
(PDF)

**Table S17** 20 SNPs with lowest hemoglobin p-values within the total high altitude Ethiopian sample.
(PDF)

**Table S18** 20 SNPs with lowest oxygen saturation p-values within the total high altitude Ethiopian sample.
(PDF)

**Table S19** 20 SNPs with lowest hemoglobin p-values within the total low altitude Ethiopian sample.
(PDF)

**Table S20** 20 SNPs with lowest oxygen saturation p-values within the total low altitude Ethiopian sample.
(PDF)

**Table S21** Power calculations within the Ethiopian samples.
(PDF)

**Table S22** Proportions of Response to Hypoxia genic SNPs relative to the proportion of all other genic SNPs in the tails of the PBS distributions for each population trio tested. * and ** denote support from ≥95% and 99% of bootstrap replicates, respectively.
(PDF)

**Table S23** List of the 20 SNPs with the largest Amhara PBS (*versus* Maasai and Luya).
(PDF)

**Table S24** 20 SNPs with largest high altitude Amhara PBS (*versus* low altitude Amhara and low altitude Oromo).
(PDF)

**Table S25** SNPs with oxygen saturation ($O_2$ Sat) association p-value<0.05 among the top 20 Amhara PBS SNPs.
(PDF)

**Table S26** 20 SNPs with largest high altitude Amhara PBS or MR and Hb or $O_2$ sat association pvalue<0.05 within HA Amhara.
(PDF)

**Table S27** 20 SNPs with largest high altitude Amhara MR score.
(PDF)

**Table S28** 40 CpG sites with highest high versus low methylation difference within Oromo. TSS denotes transcription start site.
(PDF)

**Table S29** 40 CpG sites with highest high versus low methylation difference within Amhara. TSS denotes transcription start site.
(PDF)

**Text S1** Sampled populations and their ecology.
(DOCX)

**Text S2** The genetic structure of the Amhara and Oromo populations.
(DOCX)

**Text S3** Phenotypic variation in Amhara and Oromo.
(DOCX)

**Text S4** Comparing the genetic architecture of Hb levels between Ethiopian and Tibetans.
(DOCX)

## Acknowledgments

and local arrangements. The Frankfurt Zoological Society generously allowed use of the Wolf Research Camp in the Bale Mountains. Daniel Tessema, laboratory technician, and Gezahegne Fentahun, M.D., of Addis Ababa University worked in both field sites. The authors also thank members of the Di Rienzo lab, John Marioni, Jordana T. Bell, and Lucy Godley for helpful discussions during the course of this project as well as Sarah Tishkoff and Laura Scheinfeldt for sharing data on genotype–phenotype association in their samples.

## Author Contributions

## References

1. Rytkonen KT, Storz JF (2011) Evolutionary origins of oxygen sensing in animals. EMBO Rep 12: 3–4.
2. Moore LG, Shriver M, Bemis L, Hickler B, Wilson M, et al. (2004) Maternal adaptation to high-altitude pregnancy: an experiment of nature–a review. Placenta 25 Suppl A: S60–71.
3. Brutsaert TD (2008) Do high-altitude natives have enhanced exercise performance at altitude? Appl Physiol Nutr Metab 33: 582–592.
4. Lowe GD, Lee AJ, Rumley A, Price JF, Fowkes FG (1997) Blood viscosity and risk of cardiovascular events: the Edinburgh Artery Study. Br J Haematol 96: 168–173.
5. Danesh J, Collins R, Peto R, Lowe GD (2000) Haematocrit, viscosity, erythrocyte sedimentation rate: meta-analyses of prospective studies of coronary heart disease. Eur Heart J 21: 515–520.
6. Ciuffetti G, Schillaci G, Lombardini R, Pirro M, Vaudo G, et al. (2005) Prognostic impact of low-shear whole blood viscosity in hypertensive men. Eur J Clin Invest 35: 93–98.
7. Gonzales GF, Steenland K, Tapia V (2009) Maternal hemoglobin level and fetal outcome at low and high altitudes. Am J Physiol Regul Integr Comp Physiol 297: R1477–1485.
8. Aldenderfer M (2011) Peopling the Tibetan plateau: insights from archaeology. High Alt Med Biol 12: 141–147.
9. Aldenderfer MS (2003) Moving Up in the World; Archaeologists seek to understand how and when people came to occupy the Andean and Tibetan plateaus. Am Sci: 542–549.
10. Pleurdeau D (2006) Human Technical Behavior in the African Middle Stone Age: The Lithic Assemblange of Porc-Epic Cave (Dire Dawa, Ethiopia). Afr Archaeol Rev 177–197.
11. Beall CM (2006) Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. Integr Comp Biol 46: 18–24.
12. Beall CM (2007) Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proc Natl Acad Sci U S A 104 Suppl 1: 8655–8660.
13. Beall CM, Decker MJ, Brittenham GM, Kushner I, Gebremedhin A, et al. (2002) An Ethiopian pattern of human adaptation to high-altitude hypoxia. Proc Natl Acad Sci U S A 99: 17215–17218.
14. Hoit BD, Dalton ND, Gebremedhin A, Janocha A, Zimmerman PA, et al. (2011) Elevated pulmonary artery pressure among Amhara highlanders in Ethiopia. Am J Hum Biol 23: 168–176.
15. Beall CM, Song K, Elston RC, Goldstein MC (2004) Higher offspring survival among Tibetan women with high oxygen saturation genotypes residing at 4,000 m. Proc Natl Acad Sci U S A 101: 14300–14304.
16. Simonson TS, Yang Y, Huff CD, Yun H, Qin G, et al. (2010) Genetic evidence for high-altitude adaptation in Tibet. Science 329: 72–75.
17. Beall CM, Cavalleri GL, Deng L, Elston RC, Gao Y, et al. (2010) Natural selection on EPAS1 (HIF2alpha) associated with low hemoglobin concentration in Tibetan highlanders. Proc Natl Acad Sci U S A 107: 11459–11464.
18. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZX, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329: 75–78.
19. Storz JF, Scott GR, Cheviron ZA (2010) Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. J Exp Biol 213: 4125–4136.
20. Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, et al. (2010) Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genet 6: e1001116. doi:10.1371/journal.pgen.1001116
21. Scheinfeldt LB, Soi S, Thompson S, Ranciaro A, Meskel DW, et al. (2012) Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biol 13: R1.
22. Hassen M (1990) The Oromo of Ethiopia: a history, 1570–1860. Great Britain: Cambridge University Press.
23. Lewis HS (1966) The Origins of the Galla and Somali. The Journal of African History 7: 27–46.
24. Aldenderfer MS (2003) Moving Up in the World; Archaeologists seek to understand how and when people came to occupy the Andean and Tibetan plateaus. American Scientist 91: 542–549.
25. Pleurdeau D (2006) Human Technical Behavior in the African Middle Stone Age: The Lithic Assemblange of Porc-Epic Cave (Dire Dawa, Ethiopia). African Archaeological Review 22: 177–197.
26. Kim EJ, Yoo YG, Yang WK, Lim YS, Na TY, et al. (2008) Transcriptional activation of HIF-1 by RORalpha and its role in hypoxia signaling. Arterioscler Thromb Vasc Biol 28: 1796–1802.
27. Jiang Y, Zhang W, Kondo K, Klco JM, St Martin TB, et al. (2003) Gene expression profiling in a renal cell carcinoma cell line: dissecting VHL and hypoxia-dependent pathways. Mol Cancer Res 1: 453–462.
28. O'Connell K, Posthumus M, Collins M (2011) COL6A1 gene and Ironman triathlon performance. Int J Sports Med 32: 896–901.
29. Niu X, Zhang T, Liao L, Zhou L, Lindner DJ, et al. (2012) The von Hippel-Lindau tumor suppressor protein regulates gene expression and tumor growth through histone demethylase JARID1C. Oncogene 31: 776–786.
30. Kitajima Y, Ide T, Ohtsuka T, Miyazaki K (2008) Induction of hepatocyte growth factor activator gene expression under hypoxia activates the hepatocyte growth factor/c-Met system via hypoxia inducible factor-1 in pancreatic cancer. Cancer Sci 99: 1341–1347.
31. Tacchini L, Dansi P, Matteucci E, Desiderio MA (2001) Hepatocyte growth factor signalling stimulates hypoxia inducible factor-1 (HIF-1) activity in HepG2 hepatoma cells. Carcinogenesis 22: 1363–1371.
32. He F, Wu LX, Shu KX, Liu FY, Yang LJ, et al. (2008) HGF protects cultured cortical neurons against hypoxia/reoxygenation induced cell injury via ERK1/2 and PI-3K/Akt pathways. Colloids Surf B Biointerfaces 61: 290–297.
33. Kimura K, Teranishi S, Kawamoto K, Nishida T (2010) Protection of human corneal epithelial cells from hypoxia-induced disruption of barrier function by hepatocyte growth factor. Exp Eye Res 90: 337–343.
34. Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. PLoS Biol 4: e72. e72. doi:10.1371/journal.pbio.0040072
35. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, et al. (2007) Genome-wide detection and characterization of positive selection in human populations. Nature 449: 913–918.
36. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, et al. (2009) Signals of recent positive selection in a worldwide sample of human populations. Genome Res 19: 826–837.
37. Tournamille C, Colin Y, Cartron JP, Le Van Kim C (1995) Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. Nat Genet 10: 224–228.
38. Reich D, Nalls MA, Kao WH, Akylbekova EL, Tandon A, et al. (2009) Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS Genet 5: e1000360. doi:10.1371/journal.pgen.1000360
39. Yuan WC, Lee YR, Huang SF, Lin YM, Chen TY, et al. (2011) A Cullin3-KLHL20 Ubiquitin ligase-dependent pathway targets PML to potentiate HIF-1 signaling and prostate cancer progression. Cancer Cell 20: 214–228.
40. Lombardi MS, van den Tweel E, Kavelaars A, Groenendaal F, van Bel F, et al. (2004) Hypoxia/ischemia modulates G protein-coupled receptor kinase 2 and beta-arrestin-1 levels in the neonatal rat brain. Stroke 35: 981–986.
41. Elvidge GP, Glenny L, Appelhoff RJ, Ratcliffe PJ, Ragoussis J, et al. (2006) Concordant regulation of gene expression by hypoxia and 2-oxoglutarate-dependent dioxygenase inhibition: the role of HIF-1alpha, HIF-2alpha, and other pathways. J Biol Chem 281: 15215–15226.
42. Manalo DJ, Rowan A, Lavoie T, Natarajan L, Kelly BD, et al. (2005) Transcriptional regulation of vascular endothelial cell responses to hypoxia by HIF-1. Blood 105: 659–669.
43. Kayyali US, Pennella CM, Trujillo C, Villa O, Gaestel M, et al. (2002) Cytoskeletal changes in hypoxic pulmonary endothelial cells are dependent on MAPK-activated protein kinase MK2. J Biol Chem 277: 42596–42602.
44. Chakkalakal JV, Stocksley MA, Harrison MA, Angus LM, Deschenes-Furry J, et al. (2003) Expression of utrophin A mRNA correlates with the oxidative capacity of skeletal muscle fiber types and is regulated by calcineurin/NFAT signaling. Proc Natl Acad Sci U S A 100: 7791–7796.
45. Stepto NK, Coffey VG, Carey AL, Ponnampalam AP, Canny BJ, et al. (2009) Global gene expression in skeletal muscle from well-trained strength and endurance athletes. Med Sci Sports Exerc 41: 546–565.
46. Gramolini AO, Belanger G, Thompson JM, Chakkalakal JV, Jasmin BJ (2001) Increased expression of utrophin in a slow versus a fast muscle involves posttranscriptional events. Am J Physiol Cell Physiol 281: C1300–1309.
47. Cervera AM, Apostolova N, Crespo FL, Mata M, McCreath KJ (2008) Cells silenced for SDHB expression display characteristic features of the tumor phenotype. Cancer Res 68: 4058–4067.
48. Watson JA, Watson CJ, McCann A, Baugh J (2010) Epigenetics, the epicenter of the hypoxic response. Epigenetics 5: 293–296.
49. Norman JM, Mashiba M, McNamara LA, Onafuwa-Nuga A, Chiari-Fort E, et al. (2011) The antiviral factor APOBEC3G enhances the recognition of HIV-infected primary T cells by natural killer cells. Nat Immunol 12: 975–983.

50. Wang X, Ao Z, Chen L, Kobinger G, Peng J, et al. (2012) The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. J Virol 86: 3777–3786.

51. Izmailova E, Bertley FM, Huang Q, Makori N, Miller CJ, et al. (2003) HIV-1 Tat reprograms immature dendritic cells to express chemoattractants for activated T cells and macrophages. Nat Med 9: 191–197.

52. Kebede D, Aklilu M, Sanders E (2000) The HIV epidemic and the state of its surveillance in Ethiopia. Ethiop Med J 38: 283–302.

53. Hladik W, Shabbir I, Jelaludin A, Woldu A, Tsehaynesh M, et al. (2006) HIV/AIDS in Ethiopia: where is the epidemic heading? Sex Transm Infect 82 Suppl 1: i32–35.

54. Joshi B, Ordonez-Ercan D, Dasgupta P, Chellappan S (2005) Induction of human metallothionein 1G promoter by VEGF and heavy metals: differential involvement of E2F and metal transcription factors. Oncogene 24: 2204–2217.

55. Kieusseian A, Chagraoui J, Kerdudo C, Mangeot PE, Gage PJ, et al. (2006) Expression of Pitx2 in stromal cells is required for normal hematopoiesis. Blood 107: 492–500.

56. Zhang HZ, Degar BA, Rogoulina S, Resor C, Booth CJ, et al. (2006) Hematopoiesis following disruption of the Pitx2 homeodomain gene. Exp Hematol 34: 167–178.

57. Kuhlicke J, Frick JS, Morote-Garcia JC, Rosenberger P, Eltzschig HK (2007) Hypoxia inducible factor (HIF)-1 coordinates induction of Toll-like receptors TLR2 and TLR6 during hypoxia. PLoS ONE 2: e1364. doi:10.1371/journal.pone.0001364

58. Millen J, MacLean MR, Houslay MD (2006) Hypoxia-induced remodelling of PDE4 isoform expression and cAMP handling in human pulmonary artery smooth muscle cells. Eur J Cell Biol 85: 679–691.

59. Liu YL, Song XR, Wang XW, Wei L, Li DP (2007) [Prolonged hypoxia induces GST-pi expression and drug resistance in lung adenocarcinoma cell line SPCA1]. Zhonghua Jie He He Hu Xi Za Zhi 30: 108–111.

60. Millar DS, Ow KK, Paul CL, Russell PJ, Molloy PL, et al. (1999) Detailed methylation analysis of the glutathione S-transferase pi (GSTP1) gene in prostate cancer. Oncogene 18: 1313–1324.

61. Hassen M (1990) The Oromo of Ethiopia: a history, 1570–1860. : Great Britain: Cambridge University Press.

62. Lewis H (1966) The Origins of the Galla and Somali. The Journal of African History 7: 27–46.

63. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, et al. (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293: 455–462.

64. Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, et al. (2007) Convergent adaptation of human lactase persistence in Africa and Europe. Nat Genet 39: 31–40.

65. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, et al. (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74: 1111–1120.

66. Enattah NS, Jensen TG, Nielsen M, Lewinski R, Kuokkanen M, et al. (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. Am J Hum Genet 82: 57–72.

67. Beall CM, Strohl KP, Blangero J, Williams-Blangero S, Decker MJ, et al. (1997) Quantitative genetic analysis of arterial oxygen saturation in Tibetan highlanders. Hum Biol 69: 597–604.

68. Beall CM, Brittenham GM, Strohl KP, Blangero J, Williams-Blangero S, et al. (1998) Hemoglobin concentration of high-altitude Tibetans and Bolivian Aymara. Am J Phys Anthropol 106: 385–400.

69. Beall CM, Almasy LA, Blangero J, Williams-Blangero S, Brittenham GM, et al. (1999) Percent of oxygen saturation of arterial hemoglobin among Bolivian Aymara at 3,900–4,000 m. Am J Phys Anthropol 108: 41–51.

70. Alemu A, Muluye D, Mihret M, Adugna M, Gebeyaw M (2012) Ten year trend analysis of malaria prevalence in Kola Diba, North Gondar, Northwest Ethiopia. Parasit Vectors 5: 173.

71. Endeshaw T, Graves PM, Ayele B, Mosher AW, Gebre T, et al. (2012) Performance of local light microscopy and the ParaScreen Pan/Pf rapid diagnostic test to detect malaria in health centers in Northwest Ethiopia. PLoS ONE 7: e33014. doi:10.1371/journal.pone.0033014

72. Woyessa A, Deressa W, Ali A, Lindtjorn B (2012) Prevalence of malaria infection in Butajira area, south-central Ethiopia. Malar J 11: 84.

73. Semenza GL (2011) Hypoxia. Cross talk between oxygen sensing and the cell cycle machinery. Am J Physiol Cell Physiol 301: C550–552.

74. Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. Ecol Lett 11: 106–115.

75. Idaghdour Y, Storey JD, Jadallah SJ, Gibson G (2008) A genome-wide gene expression signature of environmental geography in leukocytes of Moroccan Amazighs. PLoS Genet 4: e1000052. doi:10.1371/journal.pgen.1000052

76. Beall CM, Gebremedhin A, Brittenham GM, Decker MJ, Shamebo M (1997) Blood pressure variation among Ethiopians on the Simien Plateau. Ann Hum Biol 24: 333–342.

77. Epstein MP, Duren WL, Boehnke M (2000) Improved inference of relationship for pairs of individuals. Am J Hum Genet 67: 1219–1231.

78. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5: e1000529. doi:10.1371/journal.pgen.1000529

79. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575.

80. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. Science 319: 1100–1104.

81. Hancock AM, Witonsky DB, Alkorta-Aranburu G, Beall CM, Gebremedhin A, et al. (2011) Adaptations to climate-mediated selective pressures in humans. PLoS Genet 7: e1001375. doi:10.1371/journal.pgen.1001375

82. Cohen J (1988) Statistical power analysis for the behavioral sciences, 2nd edition. Hillsdale, NJ: Lawrence Earlbaum Associates.