Supplemental Material

BGS simulations

We conducted forward simulations using SliM (Messer 2013) to test if BGS could have an effect strong enough to mask positive selection. We did not intend to get a very precise estimate, since such an estimate would require knowledge of BGS patterns in the human genome far better than currently available. Rather we seek to obtain the order of magnitude of the decrease of diversity (is it closer to 1% or 10%?) that can be observed when comparing strongly and weakly constrained regions, as is the case in the near versus far test.

BGS varies in two ways along the human genome that make it very challenging to simulate realistically. It depends first on the amount of constrained sites, and second on what the distribution of fitness effects (DFE) of deleterious mutations looks like at these sites. The first issue when simulating BGS is that the average amount of constrained sites in human is not known precisely. In the windows bootstrapped for the near versus far test, 7% of positions belong to phastCons segments that are conserved between mammals and/or primates. This represents a low minimal estimate of the average amount of constrained sites. Indeed, this includes only segments that are conserved in all mammals or all primates, which may represent only a fraction of all, more species-specific constrained elements. Indeed, 40% of the human genome can be aligned with the mouse genome (Schwartz et al. 2003). In our simulations, we used conservative proportions of constrained positions ranging from 0 to 10%. In total, simulated regions span one megabase. The constrained positions fall within 250 bases segments distributed randomly between 250 and 750 kb of the simulated regions (to avoid edge effects). Each simulated region has 200 such segments, and we adjust the amount of constraint by randomly choosing the proportion of constrained sites in these segments between 0 and 1. To obtain accurate estimates of the decrease of diversity as a function of the amount of constrained sites, we simulated 40,000 regions with a recombination rate of 1 cM/Mb, and 40,000 regions with a recombination rate of 0.5 cM/Mb (supplemental Fig. 1).

The second issue is what distribution of fitness effects to use at constrained positions. Estimates of this distribution have been obtained for coding sequences (Keightley and Eyre-Walker 2007), but not for non-coding sequences. For simplicity we

used the DFE provided by Keightley and Eyre-Walker (Keightley and Eyre-Walker 2007) for all the simulated constrained sites.

The omnibus near versus far and functional versus non-functional tests reveals a strong departure from neutrality

The near versus far test (Fig. 1) and the functional versus non-functional tests (Fig. 2) look for different signals in the data and should be independent of one another. Indeed, all regions in the functional versus non-functional test are located near an amino acid substitution and thus they all are in the "near" category in the near versus far test. The fact that the near regions have lower diversity than the far regions should not affect the results of the test that looks within the near regions. In addition, we confirmed that the finite number of regions used in the bootstrap procedure does not generate spurious correlations between the two tests. To ensure that the near versus far and functional versus non-functional tests are independent, we verify that their results are uncorrelated. To do so we run 500 neutral population simulations using SliM (rescaling factor 20, see Methods). From these 500 simulations we calculate (as for Figs. 1 and 2) 500 pairs of π_{near}/π_{far} and $\pi_{func}/\pi_{non\text{-}func}$ and calculate the correlation between the two values. As expected there is no correlation between the two ratios (n=500, Spearman's ρ=5x10^{-4} P=0.99), which shows that the two tests are completely independent. The independence of the two tests enables us to run the randomization test separately for the *near/far* and for the *functional/non-functional* tests and combine the results. To do so we run 10,000 iterations of our randomization procedure separately for both tests and we couple *near/far* iteration *n* with f*unctional/non-functional* iteration *n*. This is made possible by the fact that the two tests are independent. In all human populations together the observed combined decreases are highly unexpected, as shown by the *P*-values of the combined randomization test on Fig. 3. Even in Africa where the signal of positive selection is consistently weaker in both the near versus far and the functional versus non-functional tests, the probability of both observed decreases by chance is less than 1%. In European and Asian populations the same probability is lower than 0.1%. Taken together these results strongly suggest that positive selection has significantly decreased neutral diversity in the human genome.

Testing the robustness of *iHS* and *XPEHH* to BGS

The strong sensitivity of average heterozygosity to BGS makes it very challenging to distinguish the specific effects of positive selection using this particular measure of diversity. Indeed, we could only detect a signature of selective sweeps when BGS is weak, that is in regions with a low density of conserved coding sequences (Results). Measures of diversity such as *iHS* and *XPEHH* should be much more robust to BGS than average heterozygosity. Indeed they are based on the frequency and length of haplotypes, and there is no *a priori* reason why BGS should affect haplotype length and frequency. In particular, BGS is not expected to create the long and frequent haplotypes that *iHS* and *XPEHH* are sensitive to.

In order to test the robustness of *iHS* and *XPEHH* to BGS, we use Slim to simulate 4 Mb sequences that include a 100 kb central region where deleterious mutations occur with a predefined strength of selection and rate. We simulate 4 Mb which is large to avoid edge effects in the calculations of *iHS* and *XPEHH*. For this reason we exclude the first and last megabases of the simulated sequences. The population size is 1000, the rescaled mutation rate is set at 10^{-7} and the rescaled recombination rate is uniform and set at 10 cM/Mb.

We test the effect of different BGS configurations on average heterozygosity, *iHS* and *XPEHH*. First, we test the effect of having 25% of the mutations in the 100kb central region with *Nes*=-200, which corresponds to strongly deleterious mutations (Supplemental Fig.1 first column). The same is done this time with 50% of strongly deleterious mutations (Supplemental Fig. 1 second column). We then test the effect of having 25% of weakly deleterious mutations with *Nes*=-5 (third column). The same is repeated with 50% of weakly deleterious mutations (fourth column). Finally we test the effect of deleterious mutations with gamma-distributed intensities of selection. Parameters of the gamma distribution fit current estimates of the distribution of fitness effects in human (Keightley and Eyre-Walker 2007). In this case we test 25% (fifth column) of mutations being deleterious, which results in a 20% decrease of heterozygosity that matches the average decrease observed near coding sequences in human (McVicker et al. 2009). For each condition we run 500 independent simulations to obtain averages and confidence intervals.

In every case tested, we find that BGS has virtually no effect on *XPEHH*, and a weak, conservative effect on *iHS* (Supplemental Fig. 2).

CLR analysis

The analysis of *iHS* amd *XPEHH* suggest that positive selection is more common near compared to far from amino acid substitutions. As a consequence, the allele frequency spectra of neutral polymorphism should more often show characteristic deviations consistent with positive selection near amino acid substitutions. We test this prediction using the composite likelihood ratio test (*CLR*) (Williamson et al. 2007). The *P*-values of the *CLR* test were retrieved from Williamson *et al*. (Williamson et al. 2007) for the East Asian and European populations. We do not consider the results of the *CLR* test for the African population because they were calculated by Williamson *et al*. (Williamson et al. 2007) using a sample of strongly admixed African-Americans individuals (Note that running the *CLR* test on the 1000 genomes phase 1 data would have been computationally prohibitive). We take the lowest *P*-value found in each window and then compare the average 10%, 5%, 2% and 1% lowest *P*-value windows near and far from amino acid substitutions as for *iHS* and *XPEHH*. We do detect more extreme values of *CLR P*-values near amino acid substitutions in Europeans, in the low recombination regions (<0.5 cM/Mbp) in East Asians, and in all regions in the combined analysis of the European and East Asian populations (Supplemental Fig. 3). These results again suggest that adaptation is more common near compared to far from amino acid substitutions in the human genome.

Correlations between *iHS*, ENCODE regulatory elements and coding sequences

We investigate the correlations between *iHS* (Voight et al. 2006) and ENCODE regulatory elements (ERE) and between *iHS* and coding sequences to test whether regulatory elements are the main source of adaptive mutations in the human genome. ERE density in our analysis is the density of elements predicted as DNASE1 hypersensitive sites and also as transcription factor binding sites identified by Chip-Seq by the ENCODE Consortium (Gerstein et al. 2012). Using the overlap of both prediction methods ensures that we work with higher confidence ERE. In this analysis we calculate *iHS* using the latest 1000 Genomes project data for three human population (Yoruba, British and Beijing Chinese phase 1 July 2012 release) instead of using publicly available results from published scans. Indeed in this correlation analysis we need to calculate the correlation between values of *iHS* and values of average heterozygosity that have to be calculated from the same set of variants. To calculate this correlation we use 1 Mb windows sliding every 50 kb along the human genome. For this analysis we

use 1 Mb instead of 500 kb windows because the correlation between coding density and *iHS* is slightly higher when using 1 Mb windows (windows of recombination rate lower than 0.5 cM/Mb; 1 Mb n=9,471; Yoruba Spearman's ρ=0.044 $P<2x10^{-16}$; British ρ=0.086 $P<2x10^{-16}$; Beijing Chinese ρ=0.137 $P<2x10^{-16}$; 500 kb n=12,515; Yoruba Spearman's ρ=0.012 P=0.14; British ρ=0.072 $P<2x10^{-16}$; Beijing Chinese ρ=0.107 $P<2x10^{-16}$). It is important that we use the window size that maximizes the correlation between *iHS* and coding density since we later evaluate the effect of controlling for coding density on the correlation between *iHS* and ERE. In addition windows less than 5 Mb from centromeres and telomeres or with more than 20% assembly gaps are removed from the correlation analysis.

In each window positive selection may result in extreme *iHS* values only for a small subset of variants. Thus for each window we select only the top 5% absolute *iHS* variants and measure the average absolute *iHS* of this top 5%. This is done separately for 50 individuals in the British population (GBR, first 50 individual in the July 2012 phase 1 VCF files), 50 individuals in the Chinese Beijing population (CHB, first 50 individual in the July 2012 phase 1 VCF files) and 50 individuals in the Yoruba population (YRI, first 50 individual in the July 2012 phase 1 VCF files). We use a limited number of individuals in each population so that the calculations of *iHS* are computationally feasible in a reasonable amount of time.

We first calculate the correlation between *iHS* (top 5% average) and average heterozygosity to confirm the robustness of *iHS* to BGS previously deduced from our forward simulations. In all the three populations, the top 5% average *iHS* correlates positively with average heterozygosity (windows with recombination rate lower than 0.5 cM/Mb; n=9,471; YRI ρ=0.073 $P<2x10^{-16}$; CHB ρ=0.117 $P<2x10^{-16}$; GBR ρ=0.144 $P<2x10^{-16}$). Thus *iHS* is conservative relative to BGS, as already observed in population simulations. We correct for this conservativeness by controlling for average heterozygosity when calculating the correlations between *iHS* and coding density and between *iHS* and ERE density (see above and Fig. 7). Therefore all the correlations in the analysis are calculated controlling for recombination rate and average heterozygosity, except for the correlation between *iHS* and average heterozygosity, where we control only for recombination rate.

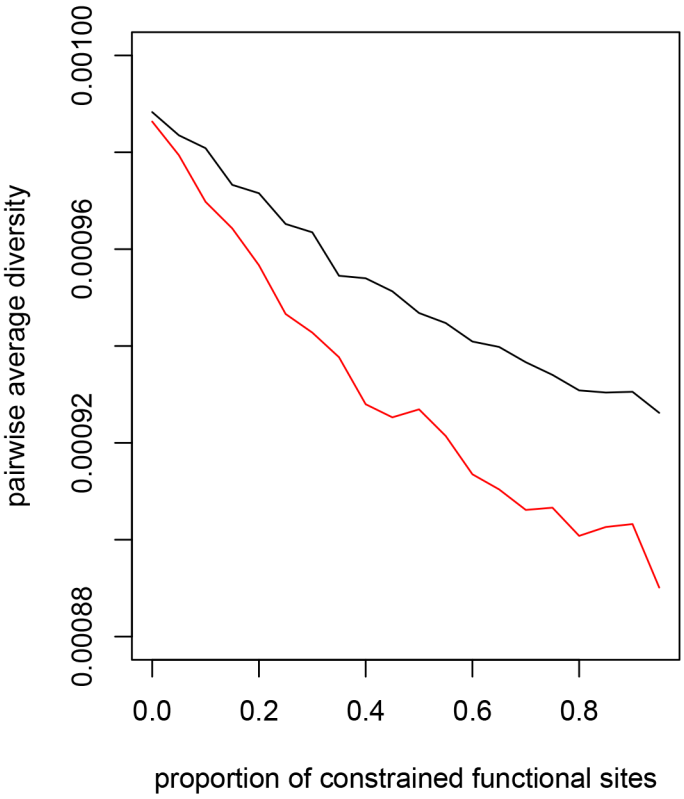
Defining genomic windows to measure diversity

Scaled neutral diversity is calculated within 500 kb windows sliding every 5 kb in the genome. A fixed physical size is chosen instead of a genetic size in order to make the windows used in the near versus far and the functional versus non-functional tests comparable. In fact an important problem with using windows with a fixed genetic size is that they can vary greatly in physical size. Depending on the recombination rate, a 0.1 cM window in the human genome can represent a physical size of 50 kb (if the recombination rate is 2 cM/Mb) or a megabase (if the recombination rate is 0.1 cM/Mb). Using a fixed genetic size can thus result in using windows with vastly different absolute content of functional elements. For instance background selection (BGS) is expected to decrease diversity more strongly in a 0.1 cM, one megabase window with 1% (10,000) of its positions in coding exons compared to a 0.1 cM, 50kb window also with 1% (500) coding exon positions. We therefore choose to use fixed physical distances and to match recombination rates as part of the bootstrap procedure used for our near versus far and functional versus non-functional tests (see Bootstrap procedure below). We use large windows of 500kb to prevent other additional issues with using smaller windows. First, bigger windows tend to exhibit less variable, closer to genomic average parameters such as GC content, CDS, UTR content, and others compared to smaller windows. This is crucial for the bootstrap procedure used in the near versus far test and in the functional versus non-functional test. Because large windows tend to be closer to the genomic average compared to smaller windows, in both tests it is much easier to find control windows that match the tested window in terms of recombination, GC content and diverse functional contents (see Bootstrap procedure below). For instance, using 500 kb windows for the near versus far test (conserved coding density<0.5%, windows further than 1 cM windows excluded, recombination rate lower than 1 cM/Mb), after 10 bootstraps we can match on average 3,260 near windows with far windows out of the 13,678 near windows in the genome. In other words, 24% of the windows of interest can be controlled for. Using 100kb instead of 500kb windows, we found that only 1.8% of the near windows can be used.

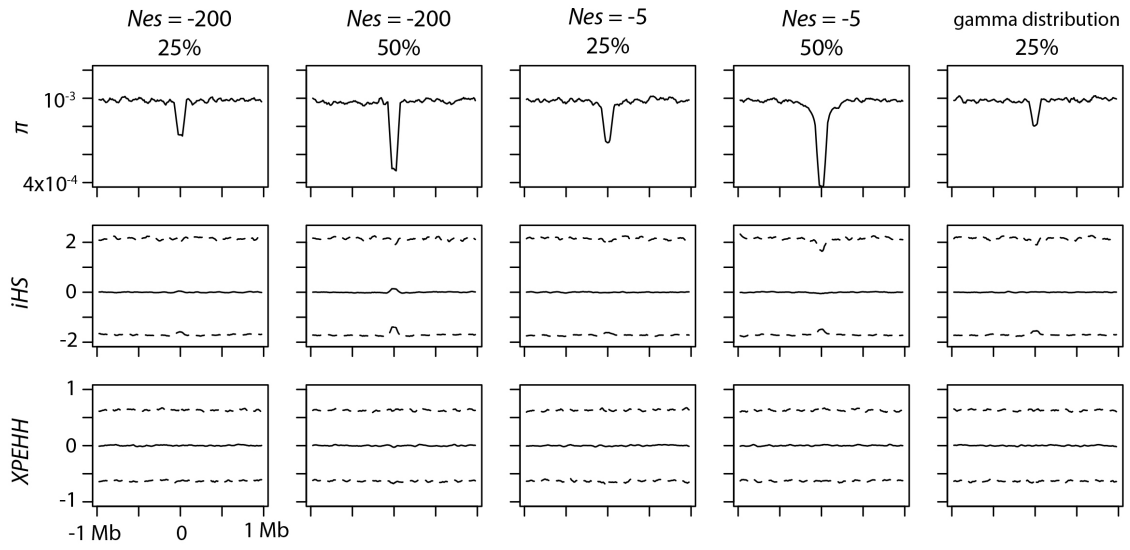
Second, an important issue with small windows is that functional elements outside of the windows but at their immediate proximity may influence diversity inside windows more strongly than if larger windows are used. Consider for example a 100 kb window within a region of low recombination rate. This window has a 1% CDS density, but is surrounded by regions with a 5% CDS density. In such a case it is likely that BGS due to the

surrounding CDS affects neutral diversity within the window even more than the CDS within the window itself. Using larger windows does not remove this edge effect, but it does improve it to some extent by reducing the effect of outside compared to inside functional elements on diversity. The effect of nearby functional elements on diversity can be estimated by measuring the partial correlation between neutral diversity and the amount of functional elements surrounding windows at a close genetic distance, controlling for the amount of functional elements within windows and recombination rate. For CDS, we measure that using 100kb windows sliding every 50kb, the partial Spearman's correlation coefficient is -0.14 between diversity in the African population and the absolute amount of CDS surrounding the windows up to a genetic distance of 0.1 cM (n=44,986, $P<2x10^{-16}$). The same partial correlation coefficient is reduced to -0.06 when using 500 kb windows sliding every 50kb (n=44,858, $P<2x10^{-16}$). The smaller influence of nearby functional elements on bigger windows reflects the fact that on average any position in bigger windows is further from the surrounding functional elements compared to smaller windows. Within 500kb windows, the average physical distance of a position to the closest window boundary is 125 kb, whereas for 100 kb the average distance is only 25 kb.

Finally, using larger windows makes the measures of diversity less noisy, especially given the fact that on average only a third of the positions within each window are used to measure scaled neutral diversity (as a reminder, those positions that occur out of CDS, UTR and CNEs, out of repeats and that are aligned with a nucleotide in macaque).

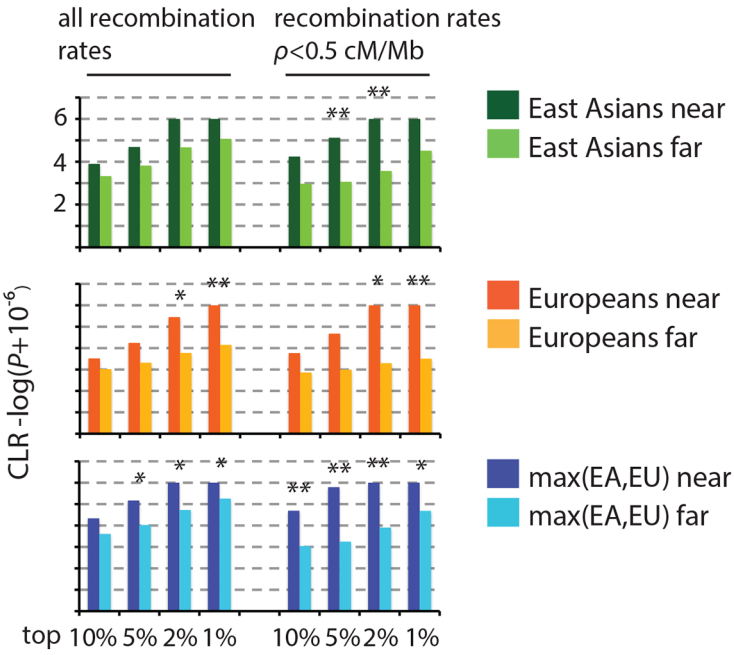


Supplemental Figure 1. Results of BGS simulations
Black curve: simulation recombination rate 1 cM/Mb. Red Curve: simulation recombination rate 0.5 cM/Mb. Curves were obtained by calculating average decreases with a smoothing range of 0.05 (proportion of constrained sites, x axis).

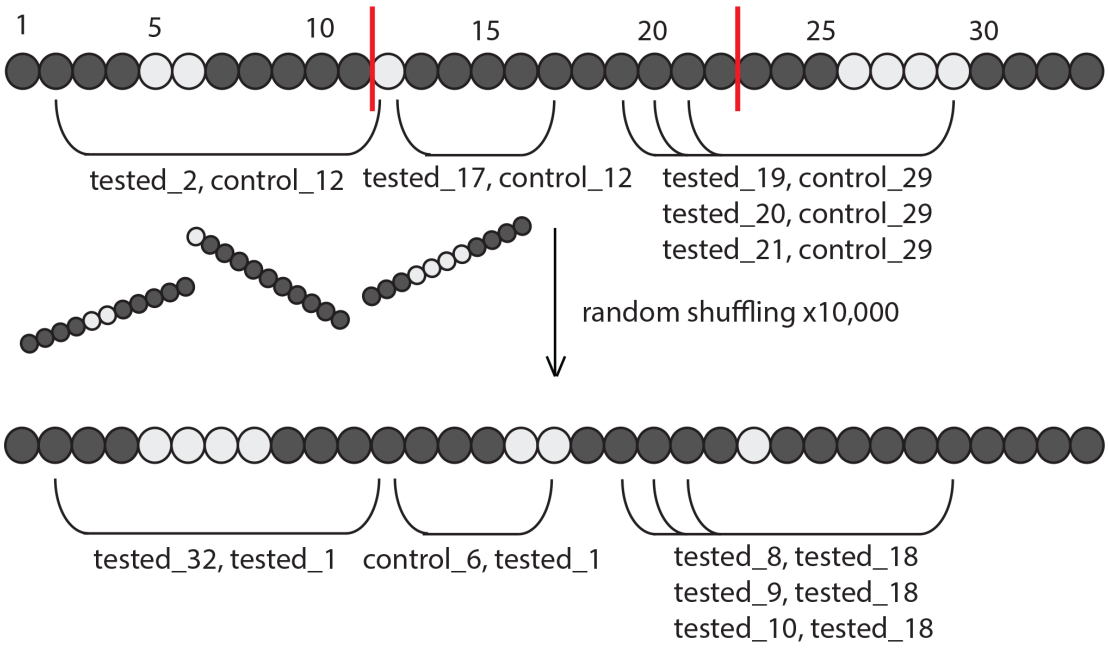


Supplemental Figure 2. Robustness of *iHS* and *XPEHH* to various rates and strengths of BGS

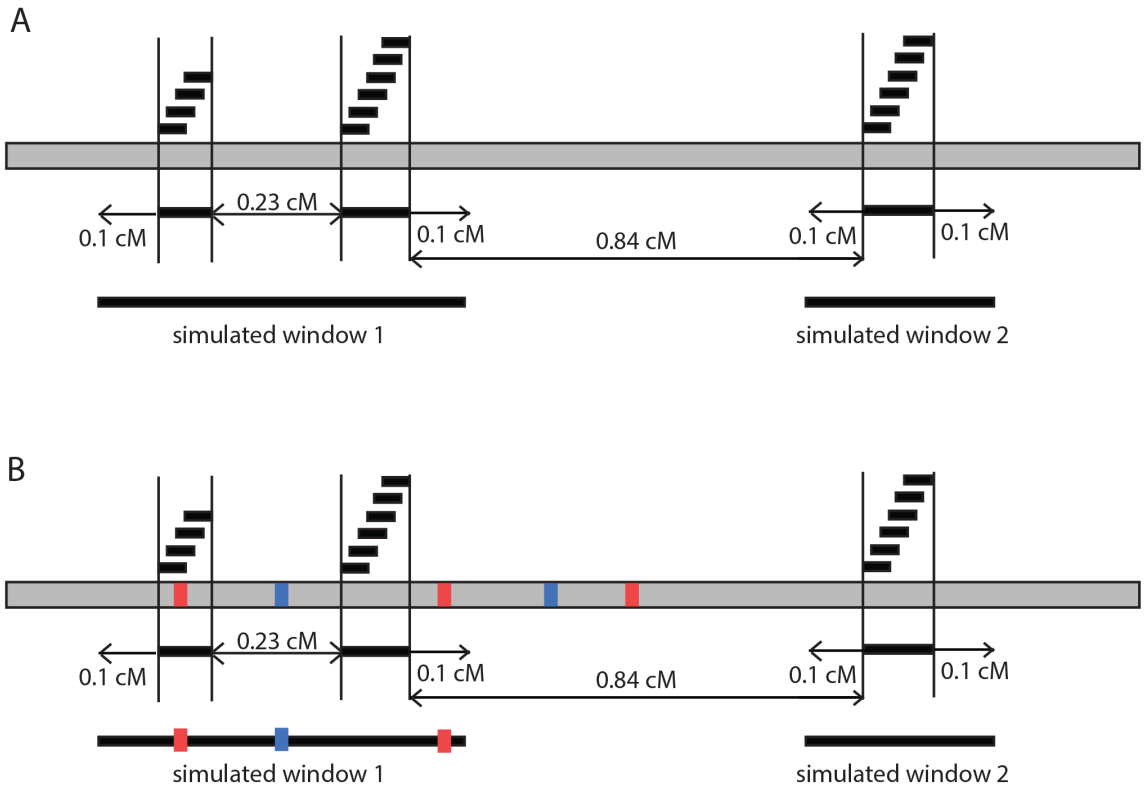
We tested the effect of BGS on *iHS* and *XPEHH* (Results and Supplemental Material). Upper row: average heterozygosity. Middle row: *iHS*. Lower row: *XPEHH*. The full line represents the average *iHS* or *XPEHH* along the simulated region. The dashed lines represent the limits of the *iHS* or *XPEHH* 95% confidence intervals.



Supplemental Figure 3. CLR analysis for the near vs. far test
For the *CLR* test we take the lowest *P*-value found in each window and then compare the average 10%, 5%, 2% and 1% lowest *P*-value windows near and far from amino acid substitutions. * randomization test $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$. Left side of histograms: all regions irrespective of recombination rates. Rigth side: only regions with recombination rates lower than 0.5 cM/Mb. The max(EA,EU) histograms show the results obtained when retaining for each window the maximum signal of the East Asian and European populations.



Supplemental Figure 4. Randomization test scheme.



Supplemental Figure 5. Population simulation scheme.

Supplemental Table 1. Amount of sequences used in the bootstrap procedure*

Recombination rate	<0.5 cM/Mb	<1 cM/Mb	<1.5 cM/Mb	<2 cM/Mb	all
Near vs far test	na	614-105	838-185	971-262	1149-463
near vs far test CCDS<0.5%	na	136-57	204-105	242-149	290-236
functional vs non-functional test	na	523-320	662-471	733-587	823-768
functional vs non-functional test CCDS<0.5%	na	103-101	129-134	141-150	157-179
near vs far test *XPEHH iHS CLR*	665-87	na	Na	na	1562-619

*This table provides the amount of sequence used in the different tests conducted. In the grey-shaded cell 614-105 means the following: for the near vs test using only windows with a recombination rate lower than 1 cM/Mb, 614 megabases of sequences near amino acid changes are used, and 105 megabases of sequences far from amino acid changes are used.

Supplemental Table 2. Robustness of the ERE and CDS versus *iHS* correlations relative to window overlap

population	Sparsing (kb)	ERE vs iHS	ERE vs iHS controlling for CDS	CDS vs iHS	CDS vs iHS controlling for ERE	n
CHB	50	0.17	0.1	0.14	0.01	9449
	250	0.17	0.11	0.12	-0.01	1884
	500	0.18	0.12	0.14	0	943
	1000	0.21, 5e-6	0.12, 1e-2	0.17, 2e-4	0.01, 8e-1	466
GBR	50	0.14	0.11	0.09	-0.03	9451
	250	0.15	0.13	0.08	-0.04	1896
	500	0.15	0.11	0.1	-0.01	942
	1000	0.17, 1e-4	0.13, 4e-3	0.12, 1e-2	f -0.02, 6e-1	478
YRI	50	0.07	0.06	0.04	-0.01	9313
	250	0.08	0.04	0.06	0.01	1866
	500	0.08	0.04	0.07	0.02	927
	1000	0.1, 3e-2	0.06, 2e-1	0.09, 6e-2	0.02, 6e-1	466

This table shows the correlation coefficients also depicted in Fig.7 as a function of the sparsing of genomic windows. For a sparsing of 1000 kb we also show the *P*-values of the correlation test (greyed cells, right side number) since the windows are expected to be largely independent from each other under these conditions.

Supplemental Table 3. Thresholds used for tested/control window matching in the bootstrap procedure*

	near vs far	functional vs non-functional	near vs far *XPEHH iHS CLR*
Recombination rate	50%-150%	50%-150%	10%-160%
GC content	97.2%-1.028%	95.3%-105%	
CDS	90%-∞	75%-1.45%	70%-500%
CCDS	10%-125%	70%-200%	
surrounding CDS	40%-∞	45%-∞	
UTR	70%-160%	70%-160%	
TFD	80%-130%	65%-140%	

*This table provides the thresholds used for choosing control windows as a function of the values of different variables of the tested windows (first column). The grey-shaded cell means that in the near versus far test recombination rate within a control window has to be comprised between 50% and 150% of the recombination rate within the tested window.

Supplemental Table 4. Robustness of the bootstrap test results to changing thresholds

Conditions changed	average TT_{near}/TT_{far}	number of bootstrapped windows
UTR(70,200)	0.925	3561
UTR(70,200), REC(50,200)	0.91	4037
REC(50,200)	0.905	3492
REC(10,275)	0.916	4110
REC(10,275), GC(96,104)	0.92	5032
REC(10,275), GC(96,104), CCDS(10,150)	0.93	6555

This table gives the results of the near vs. test (low CCDS density, recombination rates lower than 1 cM/Mb) for different thresholds of the controlled factors in the bootstrap procedure. The factors that are included in the table and their combination are those that we could vary while still comparing near and far windows conservatively. In the first column, REC means recombination, UTR is for UTRs, GC is for GC content. The new thresholds compared to Supplemental Table 3 are in parentheses.