Signatures of long-term balancing selection in human genomes

Bárbara Domingues Bitarello[1,2,*]

Cesare de Filippo[2]
João Carlos Teixeira[2,3]
Joshua M. Schmidt[2]
Philip Kleinert[2,4]
Diogo Meyer[1, ¶ ,*]

Aida M. Andrés[2, ¶,*]

1, Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil

2, Evolutionary Genetics Department, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

3, Current address: Unit of Human Evolutionary Genetics, Institut Pasteur, Paris, France

4, Current address: Computational Molecular Biology Department, Max Planck Institute for Molecular Genetics, Berlin, Germany

¶, AA and DM co-supervised the study

*Corresponding author(s): barbara_domingues@eva.mpg.de (BDB), diogo@ib.usp.br (DM), aida_andres@eva.mpg.de (AMA).

Abstract

Balancing selection maintains advantageous diversity in populations through various mechanisms. While extensively explored from a theoretical perspective, an empirical understanding of its prevalence and targets lags behind our knowledge of positive selection. Here we describe the *Non-Central Deviation* (*NCD*), a simple yet powerful statistic to detect long-term balancing selection (LTBS) which quantifies how close frequencies are to expectations under LTBS, and provides the basis for a neutrality test. *NCD* can be applied to single loci or genomic data, to populations with or without known demographic history, and can be implemented considering only polymorphisms (*NCD1*) or also considering fixed differences (*NCD2*). Both statistics have very high power to detect LTBS in humans under different frequencies of the balanced allele(s), with *NCD2* having the highest power. Applied to genome-wide data from African and European human populations *NCD2* shows that, albeit not prevalent, LTBS affects a sizable portion of the genome: about 0.6% of analyzed genomic windows and 0.8% of analyzed positions. These windows overlap about 8% of the protein-coding genes, which interestingly have larger number of transcripts than expected by chance. Significant windows contain 1.6% of the SNPs in the genome, which disproportionally overlap sites within exons and that alter protein sequence, but not putatively regulatory sites. Our catalog of candidates includes known targets of LTBS, but a majority of them are novel. As expected, immune-related genes are among those with the strongest signatures, although most candidates are involved in other biological functions, suggesting that LTBS potentially influences diverse human phenotypes.

Introduction

Balancing selection refers to a class of selective mechanisms that maintains advantageous genetic diversity in populations. Decades of research have established HLA genes as a prime example of balancing selection (Meyer and Thomson 2001; Spurgin and Richardson 2010), with thousands of alleles segregating in humans (Robinson et al. 2013), extensive support for functional effects of these polymorphisms (e.g. Hedrick et al. 1991; Prugnolle et al. 2005), and various well-documented cases of association between selected alleles and disease susceptibility (e.g. Raychaudhuri et al. 2012; Howell 2014). The catalog of well-understood non-HLA targets of balancing selection in humans remains small, but includes genes associated to phenotypes such as auto-immune diseases (Ferrer-Admetlla et al. 2008; Sironi and Clerici 2010), resistance to malaria (Malaria Genomic Epidemiology Network 2015), HIV infection (Biasin et al. 2007) or susceptibility to polycystic ovary syndrome (Day et al. 2015). Thus, besides historically influencing individual fitness, balanced polymorphisms shape current phenotypic diversity and susceptibility to disease.

Balancing selection encompasses several mechanisms (reviewed in Andrés 2011; Key, Teixeira, et al. 2014; Fijarczyk and Babik 2015). These include heterozygote advantage (or overdominance), frequency-dependent selection (Clarke 1962; Charlesworth and Charlesworth 2010), selective pressures that fluctuate in time (Muehlenbachs et al. 2008; Bergland et al. 2014) or in space in panmitic populations (Charlesworth et al. 1997; Charlesworth 2006), and some cases of pleiotropy (Johnston et al. 2013). For overdominance, pleiotropy, and some instances of spatially variable selection, a stable equilibrium can be reached

(Charlesworth and Charlesworth 2010). For other mechanisms, the frequency of the selected allele can change in time without reaching a stable equilibrium. Regardless of the mechanism, long-term balancing selection (LTBS) has the potential to leave identifiable signatures in genomic data. These include a local site-frequency spectrum (SFS) with an excess of alleles at intermediate frequencies and, when selection started far in the past, an excess of polymorphisms relative to substitutions (reviewed in Key, Teixeira, et al. 2014). In some cases, very ancient balancing selection can maintain trans-species polymorphisms in sister species (Leffler et al. 2013; Teixeira et al. 2015). On the other hand, when balancing selection is very recent or transient (Sellis et al. 2011), signatures are difficult to distinguish from incomplete, recent selective sweeps (Key, Teixeira, et al. 2014).

While balancing selection has been extensively explored from a theoretical perspective, an empirical understanding of its prevalence lags behind our knowledge of positive selection. This stems from technical difficulties in detecting balancing selection, as well as the perception that it may be a rare selective process (Hedrick 2012). In fact, few methods have been developed to identify its targets, and only a handful of studies have sought to uncover them genome-wide in humans (Asthana et al. 2005; Bubb et al. 2006; Alonso et al. 2008; Andrés et al. 2009; Leffler et al. 2013; DeGiorgio et al. 2014; Rasmussen et al. 2014; Teixeira et al. 2015). Different approaches have been used to identify genes (Andrés et al. 2009) or genomic regions (DeGiorgio et al. 2014) with an excess of polymorphisms and intermediate frequency alleles, while other studies have identified trans-species polymorphisms between humans and their closest living relatives (chimpanzees and bonobos) (Leffler et al. 2013; Teixeira et al. 2015). Overall, these studies suggested that balancing selection may act on a small portion of the genome, although the limited extent of data available (e.g. exome data (DeGiorgio et al. 2014), small sample size (Andrés et al. 2009)), and stringency of the criteria (e.g. balanced polymorphisms predating human-chimpanzee divergence (Leffler et al. 2013; Teixeira et al. 2015)) may underlie the paucity of detected regions.

Here, we developed two statistics that summarize, directly and in a simple way, the degree to which allele frequencies of SNPs in a genomic region deviate from those expected under balancing selection. We use these statistics to test the null hypothesis of neutral evolution. We showed, through simulations, that one of our statistics outperforms existing methods under a realistic demographic scenario for human populations. We applied this statistic to genome-wide data from four human populations and used both outlier and simulation-based approaches to identify genomic regions bearing signatures of LTBS.

Results

The Non-Central Deviation (NCD) statistic

Background. Owing to linkage, the signature of long-term balancing selection extends to the genetic neighborhood of the selected variant(s); therefore, patterns of polymorphism and divergence in a genomic region can be used to infer whether it evolved under LTBS (Charlesworth 2006; Andrés 2011). LTBS leaves two distinctive signatures in linked variation, when compared with neutral expectations. The first is an

increase in the ratio of polymorphic to divergent sites: by reducing the probability of fixation of a variant, balancing selection increases the local time to the most recent common ancestor (Hudson and Kaplan 1988). The HKA test is commonly used to detect this signature (Hudson et al. 1987). The second signature is an excess of alleles segregating at intermediate frequencies. In humans, the folded SFS – the frequency distribution of minor allele frequencies (MAF) — is typically L-shaped, showing an excess of low-frequency alleles when compared to expectations under neutrality and demographic equilibrium. The abundance of rare alleles is further increased by recent population expansions (Coventry et al. 2010), purifying selection and recent selective sweeps (Fu and Akey 2013). Regions under LTBS, on the other hand, can show a markedly different SFS, with proportionally more alleles at intermediate frequency (fig. 1A-B). Such a deviation in the SFS is identified by classical neutrality tests such as Tajima's *D* (Taj*D*) and newer statistics such as MWU-high (Nielsen et al. 2009).

With heterozygote advantage, the frequency equilibrium (f_{eq}) depends on the relative fitness of each genotype (Charlesworth and Charlesworth 2010): when the two types of homozygotes have the same fitness (symmetric overdominance), $f_{eq} = 0.5$; when the fitness of the two homozygotes is different (asymmetric overdominance), $f_{eq} \neq 0.5$ (suppl. information S1). Under frequency-dependent selection and fluctuating selection, while an equilibrium may not be reached (suppl. information S1), f_{eq} can be thought of as the frequency of the balanced polymorphism at the time of sampling. Here we focus on overdominance, as comparing different mechanisms of balancing selection falls outside of the scope of the paper.

***NCD* statistic.** In the tradition of neutrality tests analyzing the SFS directly (e.g. Nielsen et al. 2005; Williamson et al. 2007; Nielsen et al. 2009), we propose and define the statistic "Non-Central Deviation" (*NCD*) which measures the degree to which the local SFS deviates from a pre-specified allele frequency (the *target frequency*, *tf*) in a genomic region. Under a model of balancing selection, *tf* can be thought of as the expected frequency of a balanced allele, with the *NCD* statistic quantifying how far the sampled SNP frequencies are from it. Because bi-allelic loci have complementary allele frequencies, and there is no prior expectation regarding whether ancestral or derived alleles should be maintained at higher frequency, we use the folded SFS (fig. 1B). *NCD* is defined as:

$$NCD(tf) = \sqrt{\frac{\sum_{i=1}^{n}(p_i - tf)^2}{n}} \quad (1)$$

where i is the i-th informative site in a locus, p_i is the MAF for the *i-th* informative site, n is the number of informative sites, and *tf* is the target frequency with respect to which the deviations of the observed alleles frequencies are computed. Thus, *NCD* is a type of standard deviation that quantifies the dispersion of allelic frequencies from *tf*, rather than from the mean of the distribution. Low *NCD* values reflect a low deviation of the SFS from a pre-defined *tf*, as expected under LTBS (fig. 1C; suppl. information S1). Of course, a priori *tf* is typically unknown, so we propose below a practical approach to deal with this uncertainty.

We propose two *NCD* implementations. *NCD1* uses only polymorphic sites, and *NCD2* also includes the number of fixed differences (FDs) relative to an outgroup species (i.e, all informative sites, ISs = SNPs + FDs, are used to compute the statistic). In *NCD2*, FDs are considered ISs with MAF = 0; thus, the greater the

number of FDs, the larger the *NCD2* and the weaker the support for LTBS. From equation 1 it follows that the maximum value for *NCD2*(*tf*) is the *tf* itself (for $tf \geq 0.25$, suppl. information S1), which occurs when there are no SNPs and the number of FDs ≥ 1. The maximum *NCD1* value approaches – but never reaches – *tf* when all SNPs are singletons. The minimum value for both *NCD*1 and *NCD*2 is 0, when all SNPs segregate at *tf* and, in the case of *NCD2*, the number of FDs = 0 (suppl. figs. S1 and S2).

Power of NCD to detect LTBS

We evaluated the sensitivity and specificity of *NCD1* and *NCD2* by benchmarking their performance using simulations. Specifically, we considered demographic scenarios inferred for African, European, and Asian human populations, and simulated sequences evolving both under neutrality and LTBS using an overdominance model. We explored the influence of parameters that can affect the power of *NCD* statistics: time since onset of balancing selection (*Tbs*), frequency equilibrium defined by selection coefficients (f_{eq}), demographic history of the sampled population, *tf* used in *NCD* calculation, length of the genomic region analyzed (*L*) and implementation (*NCD1* or *NCD2*). Box 1 summarizes nomenclature used throughout the text.

For simplicity, we averaged power estimates across *NCD* implementations (*NCD* being the average of *NCD1* and *NCD2*), African and European demographic models (Asian populations were not considered, see below and suppl. information S2), *L* and *Tbs* (Methods). These averages are helpful in that they reflect the general changes in power driven by individual parameters. Nevertheless, because they often include conditions for which power is low, they underestimate the power the test can reach under each condition. The complete set of power results is presented in suppl. table S1 and some key points are discussed below.

Time since the onset of balancing selection (*Tbs*) and sequence length. Signatures of LTBS are expected to be stronger for longer *Tbs*, because time to the most recent common ancestor is older and there will have been more time for linked mutations to accumulate and reach intermediate frequencies. We simulated sequences with variable *Tbs* (1, 3, 5 million years, mya). For simplicity, we only discuss cases where $tf = f_{eq}$, although this condition is relaxed in later sections. Power to detect LTBS with *Tbs* = 1 mya is low (*NCD(0.5)* = 0.32, averaged across populations and *L* values), and high for 3 (0.74) and 5 mya (0.83) (suppl. figs. S3-S8, suppl. table S1), suggesting that *NCD* statistics are well powered to detect LTBS starting at least 3 mya. We thus focus subsequent power analyses exclusively on this timescale (3 and 5 mya).

In the absence of epistasis, the long-term effects of recombination result in narrower signatures when *Tbs* is larger (Leffler et al. 2013; Teixeira et al. 2015). Accordingly, we find that, for example, power for *NCD(0.5)* (*Tbs* = 5 mya) is on average 10% higher for 3,000 bp loci than for 12,0000 bp loci (suppl. figs. S3-S8, suppl. table S1). In brief, our simulations show power is highest for windows of 3 kb centered on the selected site (suppl. information S2), and we report power results for this length henceforth.

Demography and sample size. Power is similar for samples simulated under African and European demographic histories for *NCD2* (table 1) and *NCD1* (suppl. information S2, suppl. table S1), but

considerably lower under the Asian one (suppl. table S1, suppl. figs. S3-S8), possibly due to lower N_e (suppl. information S2). While power estimates may be influenced by the particular demographic model used, we nevertheless focus on African and European populations, which by showing similar power allow fair comparisons between them. Sample size has a modest effect in power, which was similar for samples of 30 or 50 individuals and only lower for samples of 10 individuals (slight reduction for *NCD2*, marked reduction for *NCD1*; suppl. table S1, suppl. information S2).

Simulated and target frequencies. So far, we have only discussed cases where $tf = f_{eq}$, which is expected to favor the performance of *NCD*. Accordingly, under this condition *NCD* has high power: 0.91, 0.85, and 0.79 on average for f_{eq} = 0.5, 0.4, and 0.3, respectively (averaged across *Tbs* and populations, table 1). However, since in practice there is no prior knowledge about the f_{eq} of balanced polymorphisms, we evaluate the power of *NCD* when *tf* differs from the f_{eq}. When f_{eq} = 0.5, average power is high for *tf* = 0.5 or 0.4 (above 0.85), but lower for *tf* = 0.3 (0.50, table 1). Similar patterns are observed for other simulated f_{eq} (table 1) and sample sizes (suppl. table S1, suppl. information S2). Therefore, *NCD* statistics are overall well-powered both when the f_{eq} is the same as *tf*, but also in some instances of $f_{eq} \neq tf$. In any case, the closest *tf* is to f_{eq}, the higher the power, so when possible, it is desirable to perform tests across a range of *tf.*

***NCD* implementations and comparison to other methods.** Power for *NCD2* is greater than for *NCD1* for all *tf*: f_{eq} = 0.5 (average power of 0.94 for *NCD2(0.5)* vs. 0.88 for *NCD1(0.5),* averaged across populations and *Tbs*; table 1), f_{eq} = 0.4 (0.90 for *NCD2(0.4)* vs. 0.80 for *NCD1(0.4)*) and f_{eq} = 0.3 (0.86 for *NCD2(0.3)* vs. 0.73 for *NCD1(0.3)*) (table 1, fig. 2). This illustrates the gain in power by incorporating FDs in the NCD statistic, which is also more powerful than combining *NCD1* and HKA (suppl. table S1).

We compared the power of *NCD* to two statistics commonly used to detect balancing selection (Taj*D* and HKA), a composite statistic of *NCD1* and HKA (with the goal of quantifying the contribution of FDs to *NCD* power), and a pair of model-based composite likelihood-based measures, T_1 and T_2 (DeGiorgio et al. 2014). The T_2 statistic, similarly to *NCD2*, considers both the SFS and the ratio of SNPs to FDs. Power results are summarized in fig. 2. When f_{eq} = 0.5, *NCD2(0.5)* has the highest power: for example, in Africa (*Tbs* = 5 mya, and 3 kb) *NCD2(0.5)* power is 0.96 (the highest among other tests is 0.94, for T_2) but the difference in power is highest when f_{eq} departs from 0.5. For f_{eq} = 0.4, *NCD2(0.4)* power is 0.93 (compared to 0.90 for Taj*D* and T_2, and lower for the other tests). For f_{eq} = 0.3, *NCD2(0.3)* power is 0.93 (compared to 0.89 for T_2 and lower for the other tests). These patterns are consistent in the African and European simulations (fig. 2, suppl. fig. S10), where *NCD*2 has greater or comparable power to detect LTBS than other available methods. When focusing on the tests that use only polymorphic sites, *NCD1* has similar power to Taj*D* when f_{eq} = 0.5, and it outperforms it when f_{eq} departs from 0.5 (suppl. table S1). Altogether, the advantage of *NCD2* over classic neutrality tests is its high power, especially when f_{eq} departs from 0.5; the main advantage over T_2 is simplicity of implementation. We note also that *NCD* can be computed for particular loci, even in the absence of genome-wide data or a demographic model.

Recommendations based on power analyses. Overall, *NCD* performs very well in regions of 3 kb, and similarly in European and African demographic scenarios (table 1, fig. 2). We favor windows of a given length rather than number of informative sites because the density of SNPs is part of *NCD2*'s signature. Further, depending on sample size, fixing the number of SNPs or informative sites may result in shorter widows in regions under balancing selection than in regions under neutrality, which we consider undesirable for *NCD1* and *NCD2* analyses. However, the ideal window size will vary depending on species demographic properties and sampling. Here we use sliding windows to ensure overlap with the narrow signatures of balancing selection, but alternative approaches such as windows centered on each IS are also possible.

NCD2 is simple to implement and fast to run, yet it performs slightly or substantially better than all other methods tested (fig. 2, suppl. fig. S10), reaching very high power when $tf = f_{eq}$ (always > 0.89 for selection with an onset of 5 mya and always > 0.79 for 3 mya). While the f_{eq} of a putatively balanced allele is unknown, the simplicity of the *NCD* statistics makes it trivial to run for several *tf* values, allowing detection of balancing selection for a range of equilibrium frequencies. Because *NCD2* outperforms *NCD1*, we used it for our scan of human populations; *NCD1* is nevertheless a good choice when outgroup data is lacking.

Identifying signatures of LTBS

We aimed to identify regions of the human genome under LTBS. We chose *NCD2(0.5)*, *NCD2(0.4)* and *NCD2(0.3)*, which provide sets of candidate windows that are not fully overlapping (table 1). We calculated the statistics for 3 kb windows (1.5 kb step size) and tested for significance using two complementary approaches: one testing all windows with respect to neutral expectations, and one identifying outlier windows in the empirical genomic distribution. We analyzed genome-wide data from two African (YRI: Yoruba in Ibadan, Nigeria; LWK: Luhya in Webuye, Kenya) and two European populations (GBR: British, England and Scotland; TSI: Toscani, Italy) (Abecasis et al. 2012). We filtered the data for orthology with the chimpanzee genome (used as the outgroup) and implemented five additional filters to avoid technical artifacts (suppl. fig. S13). Finally, we excluded windows with less than 10 IS in any of the populations since these showed a high variance in *NCD2* due to noisy SFS that increases the rate of false positives (see empirical patterns in suppl. fig. S17 and neutral simulation patterns in suppl. fig. S11, for the choice of 10 IS as cutoff).

Simulation and empirical-based sets of windows. After all filters were implemented, we analyzed 1,657,989 windows (~ 81% of the autosomal genome; suppl. fig. S13), overlapping 18,633 protein-coding genes. For each analysis – *NCD2(0.5)*, *NCD2(0.4)* and *NCD2(0.3)* –, we defined a *p*-value for each window as the quantile of its *NCD2* value when compared to those from 10,000 neutral simulations under the inferred demographic history of each population and conditioned on the same number of IS. Depending on the population, between 6,226 and 6,854 (0.37-0.41%) of the scanned windows have a lower *NCD2(0.5)* value than any of the 10,000 neutral simulations ($p < 0.0001$), which was our criterion for significance. The proportions are similar for *NCD2(0.4)* (0.40-0.45%) and *NCD2(0.3)* (0.33-0.38%) (table 2). We refer to these sets, whose patterns cannot be explained by the neutral model, as the *significant* windows. In each population,

the union of significant windows considering all *tf* values spans, on average, 0.6% of the windows (table 2) and 0.77% of the base pairs. The coordinates for these windows are provided in suppl. table S5.

Due to our criterion, all significant windows had simulation-based $p < 0.0001$. In order to quantify how far the *NCD2* value of each window is from neutral expectations, we defined $Z_{tf\text{-}IS}$ (Equation 2, see Methods) as the number of standard deviations a window's *NCD2* value lies from the mean of the simulated distribution under neutrality. We defined as *outlier* windows for each *tf* those with the most extreme signatures of LTBS (in the 0.05% lower tail of the respective $Z_{tf\text{-}IS}$ distribution). This more conservative set contains 829 outlier windows for each population and *tf* value (table 2), which cover only ~ 0.09% of the base pairs analyzed and are largely included in the set of significant windows. Significant and outlier windows are collectively referred to as *candidate windows*.

Reliability of candidate windows. Significant windows are enriched both in polymorphic sites (fig. 3A-B) and intermediate-frequency alleles (fig. 3C-D), and the SFS shape reflects the *tf* for which they are significant (fig. 3C-D). Although expected, because these were the patterns used to identify these windows, this shows that significant windows are unusual in both signatures. Also, as expected with balancing selection, the significant windows are largely shared across populations (see below). The striking differences of significant windows with respect to the background distribution in fig. 3, combined with the fact that neutral simulations do not have *NCD2* values as low as those of the significant windows, precludes relaxation of selective constraint as a an alternative explanation to their signatures (Andrés et al. 2009).

To avoid technical artifacts among significant windows we filtered out regions that are prone to mapping errors (suppl. fig. S13). Also, we find that significant windows have similar coverage to the rest of the genome, i.e, they are not enriched in unannotated, polymorphic duplications (suppl. fig. S14). We also examined whether these signatures could be driven by two biological mechanisms other than LTBS: archaic introgression into modern humans and ectopic gene conversion (among paralogs). These mechanisms can increase the number of polymorphic sites and (in some cases) shift the SFS towards intermediate frequency alleles (suppl. information S5). We find introgression is an unlikely confounding mechanism, since candidate windows are depleted in SNPs introgressed from Neanderthals (suppl. fig. S16, suppl. table S4, suppl. information S5). Also, genes overlapped by significant windows are not predicted to be affected by ectopic gene conversion with neighboring paralogs to an unusually high degree, with the exception of olfactory receptor genes (suppl. fig. S15, suppl. information S5). Thus, candidate windows represent a catalog of strong candidate targets of LTBS in human populations.

Assigned *tf* values

Many windows were significant for more than one *tf*. For these cases, we used the $Z_{tf\text{-}IS}$ statistic (suppl. information S3) to identify which *tf* provides the strongest support for LTBS (i.e., for which *tf* the departure from neutral expectations was greatest). In this way, we could assign a *tf* to each significant window. On average ~53% of the candidate windows are assigned to *tf* = 0.3, 27% to *tf* = 0.4 and 20% to *tf* = 0.5 (suppl. table S3).

Non-random distribution across chromosomes. Candidate windows are not randomly distributed across the genome. Chromosome 6 is the most enriched for signatures of LTBS, contributing, for example, 10.2% of significant and 25% of outlier windows genome-wide for LWK while having only 6.4% of analyzed windows (suppl. fig. S12, with qualitatively similar results for the other populations). This is explained by the MHC region (fig. 4A), rich in genes with well-supported evidence for LTBS. Specifically, 10 HLA genes are among the strongest candidates for balancing selection in all four populations, most of which have prior evidence of balancing selection (suppl. table S4, suppl. information S4 and S6): *HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DQB2, HLA-DRB1, HLA-DRB5, HLA-G* (Tan et al. 2005; Liu et al. 2006; Meyer et al. 2006; Sanchez-Mazas 2007; Solberg et al. 2008; DeGiorgio et al. 2014; Teixeira et al. 2015).

Biological pathways influenced by LTBS

To gain insight on the biological pathways influenced by LTBS, we focused on protein-coding genes containing at least one candidate window (222-249 outlier and 1,404-1,616 significant genes per population), and investigated their annotations.

Regarding functional categories, significant genes are overrepresented in 28 GO categories, 24 of which are shared by at least two populations and 18 by four populations. Thirteen categories are immune-related according to a list of 386 immune-related keywords from ImmPort (Methods). The more stringent sets of outlier genes are enriched for 28 GO categories (21 shared by all four populations), 18 of which are immune-related. Furthermore, in both sets several of the remaining enriched categories are directly related to antigen presentation although not classified as immune-related (e.g. "ER to golgi transport vesicle membrane", "integral to membrane"). Among the non immune-related categories are "sarcolemma", "epidermis development", "keratin filament" and "negative regulation of blood coagulation" (suppl. table S2).

When classical HLA genes are removed from the analyses, only two categories remain enriched: "sarcolemma" (in YRI) and "epidermis development" (GBR), but the small set of genes per population hampers power. For the significant windows, "antigen processing and presentation of endogenous peptide antigen via MHC class I" remains significantly enriched (driven by *TAP1*, *TAP2, ERAP1* and *ERAP2*; suppl. table S2). Also, significant windows are still enriched in categories related to the extracellular space – "extracellular regions", "integral to membrane" – as in other studies (Andrés et al. 2009; DeGiorgio et al. 2014; Key, Teixeira, et al. 2014) – and "keratin filament". These categories are not immune-related *per se*, but they represent physical barriers to the invasion by pathogens. This indicates that LTBS maintains advantageous diversity in numerous defense-related genes other than classical HLA genes.

Overall, 33% of the outlier (and 31% of the significant) genes have at least one immune-related associated GO category, while only 24% of scanned genes do (see Methods). These results collectively suggest that immunity and defense are frequent targets of LTBS, although a large fraction of the candidates for LTBS have non-immune related functions or indirect connections with immunity hitherto unknown.

Functional annotation of SNPs in candidate windows

We next tested whether LTBS preferentially maintains SNPs at particular types of functional sites. To do so we investigated the overlap of candidate windows with different classes of functional annotations in the human genome, and tested the hypothesis of enrichment of certain classes of sites within our sets of candidate windows, when compared to sets of randomly sampled windows from the genome (suppl. table S5 and fig. 5).

SNPs in outlier windows overlap disproportionally with protein-coding exons in all the populations ($p \leq 0.001$, one-tail test; fig. 5). The protein-coding enrichment is even stronger when considering only SNPs within genes, which both in outlier ($p < 0.001$) and significant windows ($p \leq 0.003$) are strongly enriched in protein-coding exons (fig. 5). Within the protein-coding exons, outlier windows in Africa ($p \leq 0.022$) and significant windows in all populations ($p \leq 0.037$) are enriched in non-synonymous SNPs (fig. 5). These observations show that our candidate targets of LTBS tend to be enriched in exonic and potentially functional (amino-acid altering) SNPs.

Conversely, outlier and significant windows have no excess of SNPs annotated as regulatory ($p \geq 0.458$ in all populations, fig. 5). When we explicitly compared protein-coding exons vs. regulatory sites by restricting our analysis to sites in these two categories, outlier windows have an excess of exonic SNPs ($p \leq 0.003$). The same is true for significant windows ($p \leq 0.016$; fig. 5). When only nonsynonymous and regulatory sites are considered, we see enrichment for LWK and YRI for the outlier windows ($p \leq 0.036$, fig. 5) but not for the significant windows ($p \geq 0.458$ for all populations, fig. 5), although the two analyses that consider nonsynonymous SNPs are likely underpowered due to low SNP counts (suppl. table S5). Finally, results using more detailed RegulomeDB annotations generally agree with the observation of lack of enrichment of regulatory sites in our candidate windows ($p \geq 0.121$ for a one tail test for enrichment for RegulomeDB1+3 for SNPs with MAF ≥ 0.2) (suppl. information S6, suppl. table S5).

Although perhaps limited by the quality of the annotation of regulatory sites and the low power associated to small SNP counts for nonsynonymous variants, we do not have strong evidence that LTBS in human populations has preferentially shaped variation at sites with a role in gene expression regulation. These results suggest that LTBS preferentially affects exons and non-synonymous mutations.

Overlap across populations

On average 86% of outlier windows in a given population are shared with another population (79% for significant windows), and 77% with another population within the same continent (66% for significant ones) (suppl. fig. S18). The sharing is similar when *tf* are considered separately (suppl. figs. S19 and S20). Consequently, there is also considerable overlap of candidate protein-coding genes across populations: e.g. in LWK ($tf = 0.5$), 76.6% of outlier genes are shared with any other population, and 66% are shared with YRI (89% and 77% for significant genes; fig. 4B). In fact, on average 44% of outlier genes for a given population are shared across *all* populations and 78.7% are shared by a same-continent population (50% and 77% for significant genes; suppl. fig. S21).

Candidate genes in more than one population. Instances where signatures of LTBS are not shared between populations may result from changes in selective pressure, which may be important during fast, local

adaptation (de Filippo et al. 2016). Still, loci with signatures of LTBS across human populations are more likely to represent stable selection. We considered as "shared" those candidate protein-coding genes (from the union of candidate windows for all *tf*) that are shared by all populations (suppl. table S4). For the rest, we considered as "African" those shared between YRI and LWK (but neither or only one European population), and "European" those shared between GBR and TSI (but neither or only one African population). We note that these designations do not imply that genes referred to as "African" or "European" are putative targets of LTBS for only one continent (partially because there are some power differences between Africa and Europe, table 1). The 79 African, 84 European and 102 shared outlier genes add up to 265 genes in total (~1.4% of all queried genes) and the 458 African, 400 European, and 736 shared significant genes add up to 1,594 (~8.5% of all queried genes; suppl. table S4). Several of them have been detected in other studies, but the vast majority are novel candidates (suppl. information S4).

Expression of candidate genes

Candidate genes are disproportionally expressed in a number of tissues: lung, adipose tissue, adrenal tissue, kidney, and prostate (suppl. table S2). Also, candidate genes have an unusually large number of different transcripts. For example, the set of outlier genes shared by at least two populations have on average 8.69 (compared to 7 in controls) and the respective set of significant genes 8.73 (7.1 in controls). These are both significantly higher than controls even after controlling for gene length ($p < 0.001$ in both cases for a one-tail test).

Genes with mono-allelic expression (MAE) – i.e, the random and mitotically stable choice of an active allele for a given locus – were enriched among the small set os genes previously reported to be under LTBS (Savova et al. 2016). Our observations are in agreement with these findings, with 64% and 62% of the sets of outlier and significant genes shared by at least two populations having MAE status (Savova et al. 2016), compared to only 41% for genes without signatures of LTBS ($p < 1.12^{-6}$ Fisher Exact Test, one-sided).

Discussion

Limitations of NCD statistics

LTBS produces signatures that are unexpected under neutrality and for most demographic scenarios, including those inferred for the human populations we studied. Still, all tests for balancing selection are affected by strong population substructure, which can result in an excess of polymorphism that, in certain cases (e.g. with similar contribution of the two subpopulations, or with positive selection) may be at intermediate frequencies. Some types of migration and introgression can also result in genome-wide or local excess of diversity and, under positive selection, also intermediate frequencies. Here we benefit from the extensive prior work on the 1000 Genomes populations showing absence of substructure in the populations we studied. Also, we show that Neanderthal introgression has a minimal effect on our inferences of LTBS. However, as with any neutrality test, applying *NCD* to other species requires consideration of the demographic history of the population. Certain selective regimes can

also generate signatures that are similar to those of LTBS. For example, incomplete or soft sweeps can produce an excess of intermediate frequency variants (Hermisson and Pennings 2017) that could be mistaken as evidence for balancing selection when only polymorphisms are used (e.g. with *NCD1*). Still, because none of these events should increase the time to the most recent common ancestor and density of polymorphisms, they do not confound *NCD2*. As mentioned above, recent balancing selection is difficult to distinguish from those two selective scenarios, which is one of the reasons why we focused on LTBS here.

Per definition *NCD* requires pre-defining a *tf*, which in practice is typically unknown; this can be easily addressed by running the test with different *tf* and combining results. Also, as with most neutrality tests, it is necessary to pre-define the length of the analyzed region. For *NCD* we favor a fixed length over a fixed number of IS, but this condition can be modified as needed. Despite these limitations, we show that *NCD* is a powerful, simple and fast method to identify the signatures of LTBS in polymorphism data, which we expect will be used in model and non-model organisms and that we used to uncover novel targets of LTBS in humans.

The targets of LTSB in the human genome

Using simulation-based and empirical outlier approaches, we uncovered windows with signatures of LTBS in humans. We showed that these windows are unlikely to be affected by technical artifacts or confounding biological processes other than LTBS, such as introgression from archaic hominins. On average, across populations, 0.6% of the windows in a population meet our criterion of significance (we never observe comparable or more extreme signatures of LTBS in 10,000 neutral simulations). These windows contain on average 0.77% of the base-pairs and 1.6% of the SNPs in the analyzed genome per population, and although they amount to a low proportion of the genome, on average 7.9% of the protein-coding genes in a population contain at least one significant window (considering UTRs, introns and protein coding exons). For the outlier windows (those within the 0.05% most extreme for *NCD2*), on average 1.2% of genes in each population show some evidence of selection. These proportions are similar to the ones found when requiring that a significant or outlier windows be shared by at least two same-continent populations (8.5% and 1.4%, respectively). We note that although these sets probably include some false positives, our method is only sensitive to very old balancing selection, so in practice we are likely underestimating the total influence of balancing selection in human genomes.

In both sets, we identified many previously known targets of LTBS, but also many new ones. For example, almost 70% of the outlier genes shared at least by same-continent populations (and 90% of the significant ones) are novel. Many of these candidate genes show strongest evidence for LTBS at *tf* values different from 0.5. This is expected, for instance, under asymmetric overdominance, and highlights the importance of considering selective regimes with different frequencies of the balanced polymorphism.

Functional properties of SNPs in candidate windows

In this study, we confirm cases where protein-coding regions are the likely target of selection, such as *HLA-B* and *HLA-C* (Hughes and Nei 1988), as well as cases where regulatory regions are probably targeted, such as

HLA-G, *UGT2B4, TRIM5* (Tan et al. 2005; R Cagliani et al. 2010; Sun et al. 2011). Overall, we found a strong enrichment of exonic, and a weaker enrichment of aminoacid-altering SNPs in the candidate windows, suggesting an abundance of potentially functional SNPs within selected regions.

While LTBS has been proposed to play an important role in maintaining genetic diversity that affects gene expression (Leffler et al. 2013; Savova et al. 2016), we find that regulatory SNPs are underrepresented within the candidate regions. This does not imply that there are no regulatory SNPs under balancing selection, but rather that with existing annotations (which are less precise for regulatory than protein-coding sites) they are not enriched within candidate targets. Overall, we show that LTBS plays an important role in maintaining diversity at the level of protein sequence. This is compatible with two scenarios: (a) direct selection on protein-coding sites or (b) accumulation of functional (including slightly deleterious) variants as a bi-product of balancing selection. Importantly, we show that significant windows are also extreme in their high density of polymorphisms and have a SFS that is markedly different from neutral expectations, suggesting that relaxation of purifying selection and background selection are unlikely to generate their signatures.

Overlap with previous studies

Whereas positive selection scans show a remarkably low overlap with respect to the genes they identify, with as few as 14% of protein-coding loci appearing in more than a single study (Akey 2009), 34% of our outlier genes (11% of significant ones) had evidence of LTBS in at least one previous study (Andrés et al. 2009; Leffler et al. 2013; DeGiorgio et al. 2014). Remarkably, 47% of the shared outliers across all four populations (17% of the shared significant ones) have been detected in at least one previous study, and the proportions are similar even when classical HLA genes are removed (39 and 16% overlap, respectively). This is a high degree of overlap, considering the differences in methods and datasets across studies.

Properties of candidate genes

Below we briefly discuss the outlier genes (suppl. table 6), highlighting the variety of biological functions and known genetic associations potentially shaped by LTBS in humans.

Gene expression. Candidate genes have on average a larger number of transcripts than other genes. As examples, *CLDN11*, *ROBO2*, *ESR1* or *PRKCE* (discussed below) have 10-15 different transcripts, a higher number than at least 75% of scanned genes. It is possible that these genes benefit from overall high levels of diversity, contributed both by genetic diversity and transcript diversity. Also in agreement with previous findings we find a significant excess of MAE genes among targets of balancing selection. This excess is not driven by HLA genes, which were filtered out in the study originally reporting MAE genes and supports the claim for a biological link between MAE and balancing selection (Savova et al. 2016). Heterozygosity in a MAE gene could lead to cell-to-cell heterogeneity within same-cell clusters, which could in turn be potentially advantageous (Savova et al. 2016; Sung et al. 2016)), particularly in the case of cell-surface proteins. Some of these MAE genes found in our study, and not previously detected in scans for balancing selection, are

involved in immunity/defense barriers (e.g. *IL1RL1*, *IL18R1, FAM114A1, EDARADD, SIRPA, TAS2R14)*, oxygen transport and hypoxia (e.g. *PRKCE*, *HBE1, HBG2*, *EGLN3),* or reproduction (e.g. *CLDN11).*

Oxygen transport and response to hypoxia. Among the outlier genes with MAE we find members of the beta-globin cluster (*HBE1* and *HBG2,* in the same window) that are involved in oxygen transport and have strong associations to hemoglobin levels and beta thalassemia (Danjou et al. 2015), and *EGLN3,* a regulator of the NF-kß pathway that is significantly upregulated under hypoxia in anti-inflammatory macrophages (Escribese et al. 2012) and also plays a role in skeletal muscle differentiation (Fu et al. 2007). The encoded protein hydroxylates the product of *EPAS1*, a gene shown to harbor variants responsible for human adaptation to high altitude in Tibet (Yi et al. 2013). Interestingly, in addition to having strong signatures of LTBS in all populations we analyzed, they also have evidence for recent positive selection in Andean (*HBE1*, *HBG2*) or Tibetan (*HBG2*) populations (Bigham et al. 2010; Rottgardt et al. 2010; Yi et al. 2010). It is plausible that these genes have been under LTBS, and have undergone a shift in selective pressures in high-altitude populations (as in de Filippo et al. 2016), but further analyses are required to confirm this possibility. Another outlier gene, *PRKCE,* is also strongly associated to hemoglobin levels and red blood cell traits.

Immunological function and defense barriers. It has long been argued that genes of immune function are prime candidates for balancing selection. As expected, we detect several classical HLA with known signatures of LTBS. However, many non-HLA candidates from our set of outlier genes have immunological functions. For example, we confirm signatures of LTBS in the *ABO* locus (suppl. information S4), a well-known case of LTBS in humans (Ségurel et al. 2012)*,* and *TRIM5*, a gene with important antiviral function (R Cagliani et al. 2010).

Among novel candidates of balancing selection, we find several genes involved in auto-immune disease. For example, *IL1RL1-IL18R1* have strong associations to celiac disease and atopic dermatitis, an auto-immune disease (Hirota et al. 2012). *HLA-DQB2* mediates superantigen activation of T cells (Lenormand et al. 2012) and is associated both to infectious (hepatitis B) and autoimmune diseases (e.g. Lee et al. 2012; Jiang et al. 2015). Two other significant genes for which there is prior evidence for LTBS, *ERAP1* and *ERAP2* (Andrés et al. 2010; Rachele Cagliani et al. 2010), are associated with ankylosing spondylitis and psoriasis (e.g. Strange et al. 2010; Evans et al. 2011; Robinson et al. 2015). Finally, there are several associations to autoimmune disease and susceptibility to infections in the classical HLA genes that we identify. In brief, our results are consistent with the hypothesis that auto-immune disease is linked to natural selection favoring effective immune response against pathogens (Corona et al. 2010; Sironi and Clerici 2010).

Another important aspect of defense is the avoidance of poisonous substances. As suggested previously by studies on polymorphism in PTC receptors (Wooding et al. 2004; Wooding et al. 2006), avoidance of bitterness might have been adaptive throughout human evolutionary history because several potentially harmful substances are bitter. The *TAS2R14* gene encodes for a bitter taste receptor, and in humans it has strong associations to taste perception of quinine and caffeine (Ledda et al. 2014), is considered a promiscuous receptor (Meyerhof et al. 2010; Thalmann et al. 2013; Karaman et al. 2016), and is one of the

few bitter taste receptors that binds a vast array of compounds, and for which no common structure has been found (Behrens et al. 2004; Meyerhof et al. 2010). This entails diversity in the antigen binding portions of the receptors, which may be enhanced by balancing selection. Indeed, an elevated *dN/dS* ratio was reported for a cluster of bitter taste receptors which includes *TAS2R14* (Kosiol et al. 2008). To our knowledge, our study is the first in detecting signatures of LTBS in this gene.

Cognition. Interestingly, several candidate genes are involved in cognitive abilities, or their variation is associated with diversity in related phenotypes. The *KL* (life extension factor klotho) is a gene that has been associated to human longevity (Arking et al. 2002) and for which signatures of LTBS have been previously reported (DeGiorgio et al. 2014). In mice, decreased levels of klotho shorten lifespan (reviewed in Welberg 2014). In humans, heterozygotes for the KL-VS variant show higher levels of serum klotho and enhanced cognition, independent of sex and age, than wild-type homozygotes. On the other hand, KL-VS homozygotes show decreased lifespan and reduced cognition (Dubal et al. 2014). If higher cognition is advantageous, overdominance for this phenotype can explain the signatures of balancing selection we observe (although klotho's effect in lifespan could also influence).

PDGFD encodes a growth factor that plays an essential role in wound healing and angiogenesis. A human-mice comparison revealed that the PDGFD-induced signaling is crucial for human (but not mouse) proliferation of the neocortex due to neural stem-cell proliferation (Lui et al. 2014), a trait that underlies human cognition capacities (Rakic 2009). This gene has strong associations to coronary artery disease and myocardial infarction, which are related to aging.

Also, among our outliers, a gene with a cognitive-related genetic association is *ROBO2,* a transmembrane receptor involved in axon guidance. Associations with vocabulary growth have been reported for variants in its vicinity (St Pourcain et al. 2014). *ROBO2* has signatures of ancient selective sweeps in modern humans after the split with Neanderthals and Denisova (Peyrégne et al. forthcoming) on a portion of the gene (chr3:77027850-77034264) almost 40kb apart from the one for which we identified a signature of LTBS (chr3:76985072-76988072). The occurrence of both these signatures highlights the complex evolutionary relevance of this gene.

Associations of candidate genes with cognition are also exemplified by case-control and cohort studies linking polymorphisms in the estrogen receptor alpha (ER-α) gene, *ESR1*, to dementia and cognitive decline. Links between ER-α variants and anxiety and depression in women have been proposed but lack confirmation (reviewed in Sundermann et al. 2010). Interestingly, three other of our candidate genes (*PDLIM1,GRIP1, SMYD3)* interact with ER-α at the protein level (Szklarczyk et al. 2015), and two have strong association with suicide risk (*PDLIM1,GRIP1*) (Perlis et al. 2010; Mullins et al. 2014).

In genes like *KL*, where heterozygotes show higher cognitive abilities than homozygotes, cognition may be a driving selective force. This is a possible scenario in other genes, too. Still, given the complexity of brain development and function, it is also possible that cognitive effects of this variation are a byproduct of diversity maintained for other phenotypes. For example, MHC proteins and other immune effectors are believed to affect connectivity and function of the brain (reviewed in (Shatz 2009; Needleman and McAllister

2012), with certain alleles being clearly associated with autism disorder (Careaga et al. 2010; Needleman and McAllister 2012; Torres et al. 2012).

Reproduction. Among candidate genes, there is an enrichment for preferential expression in the prostate. There are also a number of outlier genes involved in the formation of the sperm. For example, *CLDN11* encodes a tight-junction protein expressed in several tissues and crucial for spermatogenesis. Knockout mice for the murine homologue show both neurological and reproductive impairment, i.e, mutations have pleiotropic effects (Gow et al. 1999; Wu et al. 2012). In humans, variants in the gene are strongly associated to prostate cancer.

ESR1, mentioned above, encodes an estrogen-activated transcription factor and leads to abnormal secondary sexual characteristics in females when defective (Quaynor et al. 2013). ER-α interacts directly with the product of *BRCA1* and has strong associations to breast cancer (Michailidou et al. 2013), breast size (Eriksson et al. 2012) and menarche (age at onset). In males, it is involved in gonadal development and differentiation, and lack of estrogen and/or ER-α in males can lead to poor sperm viability (reviewed in Lazari et al. 2009). Strikingly, this gene also has SNPs strongly associated to a diverse array of phenotypes, including height, bone mineral density (spine and hip), and sudden cardiac arrest (Rivadeneira et al. 2009; Aouizerat et al. 2011; Wood et al. 2014). Two other genes among our candidates are also part of the estrogen signaling pathway: *PLCB4* and *ADCY5* (which is strongly associated to birth weight). Estrogens are not only involved in reproductive functions (both in male and females), but also in several processes of neural (see above), muscular or immune nature, and the ER-α-estrogen complex can act directly on promoter regions of other genes, or interact with transcription factors of genes without estrogen-sensitive promotor regions (Heldring et al. 2007). In this case, balancing selection could be explained by the high level of pleiotropy (if different alleles are beneficial for different functions), including the function in male and female reproduction (if different alleles are beneficial in males than females).

Conclusions

We present two new summary statistics, *NCD1* and *NCD2*, which are both simple and fast to implement on large datasets to identify genomic regions with signatures of LTBS. They have a high degree of sensitivity for different equilibrium frequencies of the balanced polymorphism and, unlike classical statistics such as Tajima's *D* or the Mann-Whitney U (Andrés et al. 2009; Nielsen et al. 2009), allow an exploration of the most likely frequencies at which balancing selection maintains the polymorphisms. This property is shared with the likelihood-based T_1 and T_2 tests (DeGiorgio et al. 2014). We show that the NCD statistics are well-powered to detect LTBS within a complex demographic scenario, such as that of human populations. They can be applied to either single loci or the whole-genome, in species with or without detailed demographic information, and both in the presence and absence of an appropriate outgroup.

More than 85% of our outlier windows are shared across populations, raising the possibility that long-term selective pressures have been maintained after human populations colonized new areas of the globe. Still, about 15% of outlier windows show signatures exclusively in one sampled population and a few of these

show opposing signatures of selective regimes between human groups; they are of particular relevance to understand how recent human demography might impact loci evolving under LTBS for millions of years or subsequent local adaptations through selective pressure shifts (e.g. de Filippo et al. 2016).

Our analyses indicate that, in humans, LTBS may be shaping variation in less than 2 % of variable genomic positions, but that these on average overlap with about 8% of all protein-coding genes. Although immune-related genes represent a substantial proportion of them, almost 70% of the candidate genes shared by at least same-continent populations cannot be ascribed to immune-related functions, suggesting that diverse biological functions, and the corresponding phenotypes, contain advantageous genetic diversity.

Methods

Simulations and power analyses

NCD performance was evaluated by simulations with MSMS (Ewing and Hermisson 2010) following the demographic model and parameter values described in Gravel et al. (2011) for African, European, and East Asian human populations (fig. 2). To obtain the neutral distribution for the NCD statistics, we simulated sequence data under the following demographic model: generation time of 25 years, mutation rate of 2.5×10^{-8} per site, recombination rate of 1×10^{-8}, and a human-chimpanzee split at 6.5 mya was added to the model. For simulations with selection, a balanced polymorphism was added to the center of the simulated sequence and modeled to achieve a pre-specified frequency equilibrium (f_{eq} = 0.3, 0.4, 0.5) following an overdominant model (suppl. information S2). Simulations with and without selection were run for different sequence lengths (3, 6, 12 kb) and times of onset of balancing selection (1, 3, 5 mya). For each combination of parameters, 1,000 simulations, with and without selection, were used to compare the relationship between true (TPR, the power of the statistic) and false (FPR) positive rates for the NCD statistics, represented by ROC curves. For performance comparisons, we used FPR = 0.05. When comparing performance under a given condition, power was averaged across *NCD* implementations, demographic scenarios, *L*, and *Tbs*. When comparing *NCD* performance to other methods (Taj*D,* HKA, and a combined *NCD1*+HKA test), we simulated under *NCD* optimal conditions: *L* = 3 kb and *Tbs* = 5 mya (suppl. table S1, where other conditions are also shown). T_1 and T_2 require longer genomic regions to identify the signature of balancing selection (DeGiorgio et al. 2014), so power is reported based on windows of 100 informative sites (~ 14 kb for YRI and CEU) up and downstream of the target site, following BALLET's original publication (DeGiorgio et al. 2014). We simulated 15 kb windows and calculated T_1 and T_2 with BALLET (DeGiorgio et al. 2014) for windows of 100 IS and selected the highest T_1 or T_2 value from each simulation to obtain their power for the same set of parameters used for the other simulations. Our power for T_1 and T_2 is extremely similar to the one originally reported by DeGiorgio et al. (2014), and our power for Taj*D* is substantially higher than in DeGiorgio et al. (2014) (probably due to the choice of window size and different models). Reported power values are for a sample size of 50 diploid individuals, but sample sizes of 30 and 10 individuals were also explored (suppl. information S2).

Human population genetic data

We analyzed genome-wide data from the 1000 Genomes (1000G) Project phase I (Abecasis et al. 2012), excluding SNPs only detected in the high coverage exome sequencing in order to avoid SNP density differences between coding and non-coding regions. We queried genomes of individuals from two African (YRI, LWK) and two European populations (GBR, TSI). We did not consider Asian populations due to lower NCD performance for these populations according to our simulations (suppl. table S1, suppl. figs. S7-8). To equalize sample size, we randomly sampled 50 unrelated individuals from each population (as in (Key, Peter, et al. 2014)). We dedicated extensive efforts to obtain an unbiased dataset by extensive filtering in order to avoid the inclusion of errors that may bias results. We kept positions that passed mappability (50mer CRG, 2 mismatches (Thomas Derrien et al. 2012)), segmental duplication (Cheng et al. 2005; Alkan et al. 2009) and tandem repeats filters (Benson 1999), as well as the requirement of orthology to chimp (suppl. fig. S13) because *NCD2* requires divergence information (Equation 1). Further, we excluded 3 kb windows: with less than 10 IS in any population (~2% of scanned windows) and less than 500 bp of positions with orthology in chimp (1.6%); the two criteria combined resulted in the exclusion of 2.2% of scanned windows.

Identifying signatures of LTBS

After applying all filters and requiring the presence of at least one informative site, *NCD2* was computed for 1,695,655 windows per population. Because in simulations 3kb windows yielded the highest power for *NCD2* (table 1; suppl. figs. S3-S6), we queried the 1000G data with sliding windows of 3 kb (1.5 kb step size). Windows were defined in physical distance since the presence of LTBS may affect the population-based estimates of recombination rate. For each window in each population we calculated *NCD2* for three *tf* values (0.3-0.5). In neutral simulations these three measures are, as expected, highly correlated (suppl. fig. S9).

Filtering and correction for number of informative sites. Genome-wide studies of natural selection typically place a threshold on the minimum number of IS necessary (e.g. at least 10 IS in (Andrés et al. 2009), or 100 IS in (DeGiorgio et al. 2014)). We observe considerable variance in the number of IS per 3 kb window in the 1000G data; also, *NCD2* has high variance when the number of IS is low in neutral simulations (suppl. figs. S11 and S17). We thus excluded windows with less than 10 IS in a given population because, for higher values of IS, *NCD2* stabilizes. We then analyzed the 1,657,989 windows that remained in all populations, covering 2,145,937,383 base pairs (suppl. fig. S13). Neutral simulations with different mutation rates were performed in order to retrieve 10,000 simulations for each value of IS (suppl. fig. S17, and Methods). *NCD*2 (*tf* = 0.3, 0.4, 0.5) was calculated for all simulations, allowing the assignment of significant windows and the calculation of $Z_{tf\text{-}IS}$ (Equation 2 below).

Significant and outlier windows. We defined two sets of windows with signatures of LTBS: the *significant* (based on neutral simulations) and *outlier* windows (based on the empirical distribution of $Z_{tf\text{-}IS}$, see below). When referring to both sets, we use the term *candidate windows*. *Significant windows* were defined as those fulfilling the criterion whereby the observed *NCD2* value is lower than all values obtained from 10,000

simulations with the same number of IS. Thus, all significant windows have the same *p*-value ($p < 0.0001$). In order to rank the windows and define outliers, we used a standardized distance measure between the observed *NCD2* (for a queried window) and the mean of the *NCD2* values for the 10,000 simulations with the matching number of IS:

$$Z_{tf-IS} = \frac{NCD2_{tf} - \overline{NCD2}_{tf-IS}}{sd_{tf-IS}} \quad (2)$$

, where $Z_{tf\text{-}IS}$ is the standardized *NCD2*, conditional on the value of IS, $NCD2_{tf}$ is the *NCD2* value with a given *tf* for the *n-th* empirical window, $\overline{NCD2}_{tf-IS}$ is the mean *NCD2* for 10,000 neutral simulations for the corresponding value of IS, and $sd_{tf\text{-}IS}$ is the standard deviation for 10,000 *NCD2* values from simulations with matching IS. $Z_{tf\text{-}IS}$ allows the ranking of windows for a given *tf*, while taking into account the residual effect of IS number on $NCD2_{tf}$, as well as a comparison between the rankings of a window considering different *tf* values. An empirical *p*-value was attributed to each window based on the $Z_{tf\text{-}IS}$ values for each *tf*. Windows with empirical *p*-value < 0.0005 (829 windows) were defined as the *outlier windows*. Outlier windows are essentially a subset of significant windows (except for 5 windows in LWK, 1 window in YRI, 3 windows in GBR, and 4 windows in TSI). Significant and outlier windows for multiple *tf* values had an *assigned tf value*, defined as the one that minimizes the empirical *p*-value for a given window (suppl. information S3).

Coverage as a proxy for undetected short duplication. To test whether the signatures of LTBS are driven by undetected short duplications, which can produce mapping and SNP call errors, we analyzed an alternative modern human genome-wide dataset, sequenced to an average coverage of 20x-30x per individual (Meyer et al. 2012; Prüfer et al. 2013). We used an independent data set because read coverage data is low and cryptic in the 1000G, and putative duplications affecting the SFS must be at appreciable frequency and should be present in other data sets. We considered 2 genomes from each of the following populations: Yoruba, San, French, Sardinian, Dai, and Han Chinese. For each sample, we retrieved positions above the 97.5% quantile of the coverage distribution for that sample ("high coverage" positions). For each window with signatures of LTBS, we calculated the proportion of the 3kb window having high coverage in at least two samples and plotted the distributions for different *NCD*2 $Z_{tf\text{-}IS}$ *p*-values. Extreme NCD windows are not enriched in high-coverage regions; in fact, they are depleted of them in some cases (suppl. fig. S14) (Mann-Whitney U two-tail test; $p < 0.02$ for $tf = 0.5$ and $tf = 0.4$ for GBR and TSI).

Enrichment Analyses

Gene ontology (GO) and tissue-specific expression. We analyzed protein-coding genes overlapped by one or more candidate windows. GO and tissue of expression enrichment analyses were performed using GOWINDA (Kofler and Schlötterer 2012), which corrects for gene length-related biases and/or gene clustering (suppl. information S6). GO accession terms were downloaded from the GO Consortium (http://geneontology.org). We ran analyses in *mode:gene* (which assumes that all SNPs in a gene are

completely linked) and performed 100,000 simulations for FDR (false discovery rate) estimation. Significant GO and tissue-specific categories were defined for a FDR ≤ 0.05. A minimum of three genes in the enriched category was required.

For tissue-specific expression analysis we used Illumina BodyMap 2.0 (T. Derrien et al. 2012) expression data for 16 tissues, and considered genes significantly highly expressed in a particular tissue when compared to the remaining 15 tissues using the DESeq package (Anders and Huber 2010), as done in Sankararaman et al. (2014). GO and tissue-specific expression analyses were performed for each population and set of genes: outliers or significant; different *tf* values (or union of *tf*); with or without classical HLA genes (suppl. information S6).

Archaic introgression and ectopic gene conversion. We evaluated two potentially confounding biological factors: ectopic gene conversion and archaic introgression. We verified the proportion of European SNPs in candidate windows that are potentially of archaic origin, and whether candidate genes tend to have elevated number of paralogs in the same chromosome. Details in suppl. information S5.

SNP annotations and re-sampling procedure. Functional annotations for SNPs were obtained from ENSEMBL-based annotations on the 1000G data (http://www.ensembl.org/info/genome/variation/predicted_data.html). Specifically, we categorized SNPs as: intergenic, genic, exonic, regulatory, synonymous, and non-synonymous. Details on which annotations were allocated to each of these broad categories are presented in suppl. information S6. Within each category, each SNP was only considered when variable in the population under analysis (suppl. information S6). For each candidate window, we sum the number of SNPs with each score, and then sum across candidate windows. To compare with non-candidate windows, we performed 1,000 re-samplings of the number of candidate windows (which were merged in case of overlap) from the set of background windows (all windows scanned). For each re-sampled set, we summed the number of SNPs in a particular category and then computed the ratios in suppl. table S5 and fig. 5. We therefore obtained ratios for each re-sampling set, to which we compared the values from candidate windows to obtain empirical *p*-values. Because we considered the sum of scores across windows, and counted each SNP only once, results should be insensitive to window length (as overlapping candidate windows were merged). As before, we performed these analyses for each population and sets of windows: outliers or significant, considering the union of all *tf*.

Splicing variants, monoallelic expression (MAE), immune-related genes. Candidate genes have a larger number of transcripts than controls. To test if this is a significant difference, while controlling for gene length, we: 1) divided candidate genes into quantiles of gene length; 2) sampled a set of genes (265 or 1594) matching candidate genes in quantile bin; 3) calculated the mean and median number of transcripts for each set; 4) repeated this process 1,000 times and calculated the empirical one-tailed p-value (based on where the candidate sets' values fall in the distributions based on re-sampling). The number of transcripts per gene was

obtained from Ensembl Biomart. For this analysis and to reduce complexity we used the set of candidate genes (outliers, or siginificant) that are shared by same continent populations.

To test for enrichment for genes with MAE, we quantified the number of outlier and significant genes with MAE and the number that have bi-allelic expression as described in (Savova et al. 2016). We compared these proportions to those observed for all scanned genes (one-tailed Fisher's test.) The same procedure was adopted to test for enrichment of immune-related genes among our sets: we used a list of 386 keywords from the Comprehensive List of Immune Related Genes (https://immport.niaid.nih.gov/immportWeb/queryref/immportgene/immportGeneList.do) and queried how many of the outlier protein-coding genes (402 genes in total across populations and *tf*, of which 378 had at least one associated GO term) had at least one immune-related associated GO category.

All statistical analyses and figures were performed in R (Development Core Team 2009) (scripts available on https://github.com/bbitarello/NCV_dir_package and NCD code available in https://github.com/bbitarello/NCD-Statistics). Gene Cards (www.genecards.org) and Enrichr (Kuleshov et al. 2016) were used to obtain basic functional information about genes and STRING v10 (Szklarczyk et al. 2015) was used to obtain information for interactions between genes for the discussion. The GWAS catalog (Welter et al. 2014) was used to search for associations included in the discussion (we only report "strong associations", i.e, when there is at least one SNP with $p < 10^{-8}$).

Acknowledgements

We would like to dedicate this manuscript to Scott Williamson, in memoriam, for playing a fundamental role in the conception of NCD. We also thank Warren Kretszchmar for analyses on the properties of related statistics not included here, and Eric Green for his support of that work. We thank Michael DeGiorgio for assistance with BALLET, Felix Key help with 1000 Genomes data sets, Michael Dannemann for assistance in the implementation of expression analyses, Stéphane Peyrégne for comments on the manuscript, and David Reher, members of the Evolutionary Genetics Group (São Paulo), Alex Cagan and Svante Päabo for helpful comments. This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (grant numbers 11/12500-2 and 12/19563-2 to BDB and 12/18010-0 to DM) and the Max Planck Society (AMA).

References

Abecasis GR, Auton A, Brooks LD, DePristo M a, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.

Akey JM. 2009. Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res. 19:711–722.

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, et al. 2009. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat. Genet. 41:1061–1067.

Alonso S, Lopez S, Izagirre N, de la Rua C. 2008. Overdominance in the Human Genome and Olfactory Receptor Activity. Mol. Biol. Evol. 25:997–1001.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol. 11:R106.

Andrés AM. 2011. Balancing Selection in the Human Genome. eLS:1–8.

Andrés AM, Dennis MY, Kretzschmar WW, Cannons JL, Lee-Lin S-Q, Hurle B, Schwartzberg PL, Williamson SH, Bustamante CD, Nielsen R, et al. 2010. Balancing Selection Maintains a Form of ERAP2 that Undergoes Nonsense-Mediated Decay and Affects Antigen Presentation. PLoS Genet. 6:e1001157.

Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD, et al. 2009. Targets of balancing selection in the human genome. Mol. Biol. Evol. 26:2755–2764.

Aouizerat BE, Vittinghoff E, Musone SL, Pawlikowska L, Kwok P-Y, Olgin JE, Tseng ZH. 2011. GWAS for discovery and replication of genetic loci associated with sudden cardiac arrest in patients with coronary artery disease. BMC Cardiovasc. Disord. 11:29.

Arking DE, Krebsova A, Macek M, Macek M, Arking A, Mian IS, Fried L, Hamosh A, Dey S, McIntosh I, et al. 2002. Association of human aging with a functional variant of klotho. Proc. Natl. Acad. Sci. 99:856–861.

Asthana S, Schmidt S, Sunyaev SR. 2005. A limited role for balancing selection. Trends Genet. 21:30–32.

Behrens M, Brockhoff A, Kuhn C, Bufe B, Winnig M, Meyerhof W. 2004. The human taste receptor hTAS2R14 responds to a variety of different bitter compounds. Biochem. Biophys. Res. Commun. 319:479–485.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27:573–580.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. PLoS Genet. 10:e1004775.

Biasin M, Piacentini L, Caputo S Lo, Kanari Y, Magri G, Trabattoni D, Naddeo V, Lopalco L, Clivio A, Cesana E, et al. 2007. Apolipoprotein B mRNA—Editing Enzyme, Catalytic Polypeptide—Like 3G: A Possible Role in the Resistance to HIV of HIV-Exposed Seronegative Individuals. J. Infect. Dis. 195:960–964.

Bigham A, Bauchet M, Pinto D, Mao X, Akey JM, Mei R, Scherer SW, Julian CG, Wilson MJ, López Herráez D, et al. 2010. Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data.Begun DJ, editor. PLoS Genet. 6:e1001116.

Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, Palmieri a, Subramanian S, Zhou Y, Kaul R, et al. 2006. Scan of human genome reveals no new Loci under ancient balancing selection. Genetics 173:2165–2177.

Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, Pozzoli U, Bonaglia MC, Bresolin N, Clerici M, Sironi M. 2010. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. Hum. Genet. 128:577–588.

Cagliani R, Riva S, Biasin M, Fumagalli M, Pozzoli U, Lo Caputo S, Mazzotta F, Piacentini L, Bresolin N, Clerici M, et al. 2010. Genetic diversity at endoplasmic reticulum aminopeptidases is maintained by balancing selection and is associated with natural resistance to HIV-1 infection. Hum. Mol. Genet. 19:4705–4714.

Careaga M, Water J, Ashwood P. 2010. Immune dysfunction in autism: A pathway to treatment. Neurotherapeutics 7:283–292.

Charlesworth B, Charlesworth D. 2010. Elements of Evolutionary Genetics. 1st ed. Roberts and Company Publishers

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided population. Genet. Res. 70:155–174.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2:379–384.

Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Pääbo S, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88–93.

Clarke B. 1962. Balanced polymorphism and the diversity of sympatric species. In: Nichols D, editor. Taxonomy and Geography. Oxford: Systematics Association.

Corona E, Dudley JT, Butte AJ. 2010. Extreme evolutionary disparities seen in positive selection across seven complex diseases. PLoS One 5:1–10.

Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, Hixson JE, Rea TJ, Muzny DM, Lewis LR, et al. 2010. Deep resequencing reveals excess rare recent variants consistent with explosive population growth. Nat. Commun. 1:131.

Danjou F, Zoledziewska M, Sidore C, Steri M, Busonero F, Maschio A, Mulas A, Perseu L, Barella S, Porcu E, et al. 2015. Genome-wide association analyses based on whole-genome sequencing in Sardinia provide insights into regulation of hemoglobin levels. Nat. Genet. 47:1264–1271.

Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, Bjonnes A, Broer L, Dunger DB, Halldorsson B V., et al. 2015. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nat. Commun. 6:8464.

DeGiorgio M, Lohmueller KE, Nielsen R. 2014. A model-based approach for identifying signatures of ancient balancing selection in genetic data. PLoS Genet. 10:e1004561.

Derrien T, Estellé J, Marco Sola S, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast Computation and Applications of Genome Mappability.Ouzounis CA, editor. PLoS One 7:e30377.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. Genome Res. 22:1775–1789.

Development Core Team R. 2009. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing

Dubal DB, Yokoyama JS, Zhu L, Broestl L, Worden K, Wang D, Sturm VE, Kim D, Klein E, Yu GQ, et al. 2014. Life Extension Factor Klotho Enhances Cognition. Cell Rep. 7:1065–1076.

Eriksson N, Benton GM, Do CB, Kiefer AK, Mountain JL, Hinds DA, Francke U, Tung JY. 2012. Genetic variants associated with breast size also influence breast cancer risk. BMC Med. Genet. 13:53.

Escribese M, Sierra-Filardi E, Nieto C, Samaniego R, Sánchez-Torres C, Masuyama T, Canderon-Gómez E, Vega M, Salas A, Sánchez-Mateos P, et al. 2012. The prolyl hydroxylase PHD3 identifies proinflammatory macrophages and its expression is regulated by activin A. J. Immunol. 189:1946–1954.

Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, Opperman U, Dilthey A, Pirinen M, Stone MA, et al. 2011. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat. Genet. 43:761–767.

Ewing G, Hermisson J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. Bioinformatics 26:2064–2065.

Ferrer-Admetlla A, Bosch E, Sikora M, Marques-Bonet T, Ramirez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, et al. 2008. Balancing Selection Is the Main Force Shaping the Evolution of Innate Immunity Genes. J. Immunol. 181:1315–1322.

Fijarczyk A, Babik W. 2015. Detecting balancing selection in genomes: Limits and prospects. Mol. Ecol.:n/a-n/a.

de Filippo C, Key FM, Ghirotto S, Benazzo A, Meneu JR, Weihmann A, Parra G, Green ED, Andrés AM. 2016. Recent Selection Changes in Human Genes under Long-Term Balancing Selection. Mol. Biol. Evol.:msw023.

Fu J, Menzies K, Freeman RS, Taubman MB. 2007. EGLN3 Prolyl Hydroxylase Regulates Skeletal Muscle Differentiation and Myogenin Protein Stability. J. Biol. Chem. 282:12410–12418.

Fu W, Akey JM. 2013. Selection and Adaptation in the Human Genome. Annu. Rev. Genomics Hum. Genet. 14:467–489.

Gow A, Southwood CM, Li JS, Pariali M, Riordan GP, Brodie SE, Danias J, Bronstein JM, Kachar B, Lazzarini RA. 1999. CNS myelin and sertoli cell tight junction strands are absent in Osp/claudin-11 null mice. Cell 99:649–659.

Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, Yu F, Gibbs RA, Bustamante CD. 2011. Demographic history and rare allele sharing among human populations. Proc. Natl. Acad. Sci. U. S. A. 108:11983–11988.

Hedrick PW. 2012. What is the evidence for heterozygote advantage selection? Trends Ecol. Evol. Evol. 27:698–704.

Hedrick PW, Whittam TS, Parham P. 1991. Heterozygosity at individual amino acid sites: extremely high levels for HLA-A and -B genes. Proc. Natl. Acad. Sci. 88:5897–5901.

Heldring N, Pike A, Andersson S, Matthews J, Cheng G, Treuter E, Warner M, Hartman J, Tujague M, Stro A. 2007. Estrogen Receptors : How Do They Signal and What Are Their Targets. Physiol. Rev. 87:905–931.

Hermisson J, Pennings PS. 2017. Soft sweeps and beyond: understanding the patterns and probabilities of

selection footprints under rapid adaptation.Kelley J, editor. Methods Ecol. Evol. 8:700–716.
Hirota T, Takahashi A, Kubo M, Tsunoda T, Tomita K, Sakashita M, Yamada T, Fujieda S, Tanaka S, Doi S, et al. 2012. Genome-wide association study identifies eight new susceptibility loci for atopic dermatitis in the Japanese population. Nat. Genet. 44:1222–1226.
Howell WM. 2014. HLA and disease: Guilt by association. Int. J. Immunogenet. 41:1–12.
Hudson RR, Kaplan NL. 1988. The coalescent process in models with selection and recombination. Genetics 120:831–840.
Hudson RR, Kreitman M, Aguade M. 1987. A Test of Neutral Molecular Evolution Based on Nucleotide Data. Genetics 116:153–159.
Hughes AL, Nei M. 1988. Pattern of nucleotide substitution at major histocompatibility class I loci reveals overdominant selection. Lett. to Nat. 335:167–170.
Jiang D-K, Ma X-P, Yu H, Cao G, Ding D-L, Chen H, Huang H-X, Gao Y-Z, Wu X-P, Long X-D, et al. 2015. Genetic variants in five novel loci including CFB and CD40 predispose to chronic hepatitis B. Hepatology 62:118–128.
Johnston SE, Gratten J, Berenos C, Pilkington JG, Clutton-Brock TH, Pemberton JM, Slate J. 2013. Life history trade-offs at a single locus maintain sexually selected genetic variation. Nature 502:93–95.
Karaman R, Nowak S, Di Pizio A, Kitaneh H, Abu-Jaish A, Meyerhof W, Niv MY, Behrens M. 2016. Probing the Binding Pocket of the Broadly Tuned Human Bitter Taste Receptor TAS2R14 by Chemical Modification of Cognate Agonists. Chem. Biol. Drug Des. 88:66–75.
Key FM, Peter B, Dennis MY, Huerta-Sánchez E, Tang W, Prokunina-Olsson L, Nielsen R, Andrés AM. 2014. Selection on a Variant Associated with Improved Viral Clearance Drives Local, Adaptive Pseudogenization of Interferon Lambda 4 (IFNL4). PLoS Genet. 10:e1004681.
Key FM, Teixeira JC, de Filippo C, Andrés AM. 2014. Advantageous diversity maintained by balancing selection in humans. Curr. Opin. Genet. Dev. 29:45–51.
Kofler R, Schlötterer C. 2012. Gowinda: Unbiased analysis of gene set enrichment for genome-wide association studies. Bioinformatics 28:2084–2085.
Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of positive selection in six Mammalian genomes. PLoS Genet. 4:1–17.
Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44:W90–W97.
Lazari MFM, Lucas TFG, Yasuhara F, Gomes GRO, Siu ER, Royer C, Fernandes SAF, Porto CS. 2009. Estrogen receptors and function in the male reproductive system. Arq. Bras. Endocrinol. Metabol. 53:923–933.
Ledda M, Kutalik Z, Souza Destito MC, Souza MM, Cirillo CA, Zamboni A, Martin N, Morya E, Sameshima K, Beckmann JS, et al. 2014. GWAS of human bitter taste perception identifies new loci and reveals additional complexity of bitter taste genetics. Hum. Mol. Genet. 23:259–267.
Lee YH, Bae S-C, Choi SJ, Ji JD, Song GG. 2012. Genome-wide pathway analysis of genome-wide

association studies on systemic lupus erythematosus and rheumatoid arthritis. Mol. Biol. Rep. 39:10627–10635.

Leffler EM, Pfeifer S, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, Mcvean G, et al. 2013. Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. Science (80-.). 339:1578–1582.

Lenormand C, Bausinger H, Gross F, Signorino-Gelo F, Koch S, Peressin M, Fricker D, Cazenave JP, Bieber T, Hanau D, et al. 2012. HLA-DQA2 and HLA-DQB2 genes are specifically expressed in human Langerhans cells and encode a new HLA class II molecule. J Immunol 188:3903–3911.

Liu X, Fu Y, Liu Z, Lin B, Xie Y, Liu Y, Xu Y, Lin J, Fan X, Dong M, et al. 2006. An Ancient Balanced Polymorphism in a Regulatory Region of Human Major Histocompatibility Complex Is Retained in Chinese Minorities but Lost Worldwide. Am. J. Hum. Genet. 78:393–400.

Lui JH, Nowakowski TJ, Pollen AA, Javaherian A, Kriegstein AR, Oldham MC. 2014. Radial glia require PDGFD–PDGFRβ signalling in human but not mouse neocortex. Nature 515:264–268.

Malaria Genomic Epidemiology Network. 2015. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature 526:253–257.

Meyer D, Single RM, Mack SJ, Erlich HA, Thomson G. 2006. Signatures of demographic history and natural selection in the human major histocompatibility complex Loci. Genetics 173:2121–2142.

Meyer D, Thomson G. 2001. How selection shapes variation of the human major histocompatibility complex: a review. Ann. Hum. Genet. 65:1–26.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prufer K, de Filippo C, et al. 2012. A High-Coverage Genome Sequence from an Archaic Denisovan Individual. Science (80-.). 338:222–226.

Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E, Bufe B, Appendino G, Behrens M. 2010. The Molecular Receptive Ranges of Human TAS2R Bitter Taste Receptors. Chem. Senses 35:157–170.

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, et al. 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat. Genet. 45:353–361, 361-2.

Muehlenbachs A, Fried M, Lachowitzer J, Mutabingwa TK, Duffy PE. 2008. Natural selection of FLT1 alleles and their association with malaria resistance in utero. Proc. Natl. Acad. Sci. 105:14488–14491.

Mullins N, Perroud N, Uher R, Butler AW, Cohen-Woods S, Rivera M, Malki K, Euesden J, Power RA, Tansey KE, et al. 2014. Genetic relationships between suicide attempts, suicidal ideation and major psychiatric disorders: a genome-wide association and polygenic scoring study. Am. J. Med. Genet. B. Neuropsychiatr. Genet. 165B:428–437.

Needleman LA, McAllister AK. 2012. The major histocompatibility complex and autism spectrum disorder. Dev. Neurobiol. 72:1288–1301.

Nielsen R, Bustamante C, Clark A, Glanowski S, Sackton T, Hubisz MJ, Fledel-Alon A, Tanenbaum D, Civello D, White T, et al. 2005. A Scan for Positively Selected Genes in the Genomes of Humans and Chimpanzees. PLoS Biol. 3:e170.

Nielsen R, Hubisz MJ, Hellmann I, Torgerson D, Andrés AM, Albrechtsen A, Gutenkunst R, Adams MD, Cargill M, Boyko A, et al. 2009. Darwinian and demographic forces affecting human protein coding genes. Genome Res. 19:838–849.

Perlis RH, Huang J, Purcell S, Fava M, Rush AJ, Sullivan PF, Hamilton SP, McMahon FJ, Schulze T, Potash JB, et al. 2010. Genome-Wide Association Study of Suicide Attempts in Mood Disorder Patients. Am. J. Psychiatry 167:1499–1507.

Peyrégne S, Dannemann M, Prüfer K. Detecting ancient positive selection in humans using extended lineage sorting. Genome Res.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, et al. 2013. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505:43–49.

Prugnolle F, Manica A, Charpentier M, Guégan JF, Guernier V, Balloux F. 2005. Pathogen-driven selection and worldwide HLA class I diversity. Curr. Biol. 15:1022–1027.

Quaynor SD, Stradtman EW, Kim H-G, Shen Y, Chorich LP, Schreihofer DA, Layman LC. 2013. Delayed puberty and estrogen resistance in a woman with estrogen receptor α variant. N. Engl. J. Med. 369:164–171.

Rakic P. 2009. Evolution of the neocortex: a perspective from developmental biology. Nat. Rev. Neurosci. 10:724–735.

Rasmussen MD, Hubisz MJ, Gronau I, Siepel A. 2014. Genome-Wide Inference of Ancestral Recombination Graphs.Coop G, editor. PLoS Genet. 10:e1004342.

Raychaudhuri S, Sandor C, Stahl E a, Freudenberg J, Lee H-S, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, et al. 2012. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat. Genet. 44:291–296.

Rivadeneira F, Styrkársdottir U, Estrada K, Halldórsson B V, Hsu Y-H, Richards JB, Zillikens MC, Kavvoura FK, Amin N, Aulchenko YS, et al. 2009. Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies. Nat. Genet. 41:1199–1206.

Robinson J, Halliwell JA, McWilliam H, Lopez R, Parham P, Marsh SGE. 2013. The IMGT/HLA database. Nucleic Acids Res. 41:D1222-7.

Robinson PC, Costello M-E, Leo P, Bradbury LA, Hollis K, Cortes A, Lee S, Joo K Bin, Shim S-C, Weisman M, et al. 2015. ERAP2 is associated with ankylosing spondylitis in HLA-B27 -positive and HLA-B27-negative patients. Ann. Rheum. Dis. 74:1627–1629.

Rottgardt I, Rothhammer F, Dittmar M. 2010. Native highland and lowland populations differ in γ-globin gene promoter polymorphisms related to altered fetal hemoglobin levels and delayed fetal to adult globin switch after birth. Anthropol. Sci. 118:41–48.

Sanchez-Mazas A. 2007. An apportionment of human HLA diversity. Tissue Antigens 69:198–202.

Sankararaman S, Mallick S, Dannemann M, Prüfer K, Kelso J, Pääbo S, Patterson N, Reich D. 2014. The genomic landscape of Neanderthal ancestry in present-day humans. Nature 507:354–357.

Savova V, Chun S, Sohail M, McCole RB, Witwicki R, Gai L, Lenz TL, Wu C, Sunyaev SR, Gimelbrant AA.

2016. Genes with monoallelic expression contribute disproportionately to genetic diversity in humans. Nat. Genet. 48:231–237.

Ségurel L, Thompson EE, Flutre T, Lovstad J, Venkat A, Margulis SW, Moyse J, Ross S, Gamble K, Sella G, et al. 2012. The ABO blood group is a trans-species polymorphism in primates. Proc. Natl. Acad. Sci. 109:18493–18498.

Sellis D, Callahan BJ, Petrov D a., Messer PW. 2011. Heterozygote advantage as a natural consequence of adaptation in diploids. Proc. Natl. Acad. Sci. 108:20666–20671.

Shatz CJ. 2009. MHC Class I: An Unexpected Role in Neuronal Plasticity. Neuron 64:40–45.

Sironi M, Clerici M. 2010. The hygiene hypothesis: an evolutionary perspective. Microbes Infect. 12:421–427.

Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. 2008. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: A meta-analytic review of 497 population studies. Hum. Immunol. 69:443–464.

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. Biol. Sci. 277:979–988.

St Pourcain B, Cents RAM, Whitehouse AJO, Haworth CMA, Davis OSP, O'Reilly PF, Roulstone S, Wren Y, Ang QW, Velders FP, et al. 2014. Common variation near ROBO2 is associated with expressive vocabulary in infancy. Nat. Commun. 5:4831.

Strange A, Capon F, Spencer CCA, Knight J, Weale ME, Allen MH, Barton A, Band G, Bellenguez C, Bergboer JGM, et al. 2010. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between HLA-C and ERAP1. Nat. Genet. 42:985–990.

Sun C, Huo D, Southard C, Nemesure B, Hennis A, Cristina Leske M, Wu S-Y, Witonsky DB, Olopade OI, Di Rienzo A. 2011. A signature of balancing selection in the region upstream to the human UGT2B4 gene and implications for breast cancer risk. Hum. Genet. 130:767–775.

Sundermann EE, Maki PM, Bishop JR. 2010. A review of estrogen receptor α gene (ESR1) polymorphisms, mood, and cognition. Menopause 17:874–886.

Sung MK, Jang J, Lee KS, Ghim C, Choi JK. 2016. Selected heterozygosity at cis -regulatory sequences increases the expression homogeneity of a cell population in humans. Genome Biol.:1–15.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43:D447–D452.

Tan Z, Shon AM, Ober C. 2005. Evidence of balancing selection at the HLA-G promoter region. Hum. Mol. Genet. 14:3619–3628.

Teixeira JC, de Filippo C, Weihmann A, Meneu JR, Racimo F, Dannemann M, Nickel B, Fischer A, Halbwax M, Andre C, et al. 2015. Long-Term Balancing Selection in LAD1 Maintains a Missense Trans-Species Polymorphism in Humans, Chimpanzees, and Bonobos. Mol. Biol. Evol. 32:1186–1196.

Thalmann S, Behrens M, Meyerhof W. 2013. Major haplotypes of the human bitter taste receptor TAS2R41 encode functional receptors for chloramphenicol. Biochem. Biophys. Res. Commun. 435:267–273.

Torres AR, Westover JB, Rosenspire AJ. 2012. HLA Immune Function Genes in Autism. Autism Res. Treat. 2012:1–13.

Welberg L. 2014. Cognition: Klotho spins cognitive fate. Nat. Rev. Neurosci. 15:425–425.

Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorff L, et al. 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 42:D1001–D1006.

Williamson SH, Hubisz MJ, Clark AG, Payseur BA, Bustamante CD, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet. 3:e90.

Wood AR, Esko T, Yang J, Vedantam S, Pers TH, Gustafsson S, Chu AY, Estrada K, Luan J, Kutalik Z, et al. 2014. Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46:1173–1186.

Wooding S, Bufe B, Grassi C, Howard MT, Stone AC, Vazquez M, Dunn DM, Meyerhof W, Weiss RB, Bamshad MJ. 2006. Independent evolution of bitter-taste sensitivity in humans and chimpanzees. Nature 440:930–934.

Wooding S, Kim U, Bamshad MJ, Larsen J, Jorde LB, Drayna D. 2004. Natural Selection and Molecular Evolution in PTC, a Bitter-Taste Receptor Gene. Am. J. Hum. Genet. 74:637–646.

Wu X, Peppi M, Vengalil MJ, Maheras KJ, Southwood CM, Bradley M, Gow A. 2012. Transgene-Mediated Rescue of Spermatogenesis in Cldn11-Null Mice1. Biol. Reprod. 86.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2010. Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. Science (80-.). 329:75–78.

Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, Xu X, Jiang H, Vinckenbosch N, Korneliussen TS, et al. 2013. Sequencing of Fifty Human Exomes Reveals Adaptation to High Altitude. Science (80-.). 329:75–78.

Tables

Table 1. Power for simulations under the African and European demographic models

		Africa						Europe					
		NCD2			***NCD1***			***NCD2***			***NCD1***		
Tbs	f_{eq}	***tf***			***tf***			***tf***			***tf***		
		0.5	**0.4**	**0.3**	**0.5**	**0.4**	**0.3**	**0.5**	**0.4**	**0.3**	**0.5**	**0.4**	**0.3**
5	0.5	0.96	0.94	0.84	0.93	0.91	0.39	0.97	0.95	0.84	0.92	0.85	0.20
5	0.4	0.94	0.94	0.89	0.89	0.89	0.67	0.95	0.94	0.91	0.85	0.84	0.60
5	0.3	0.90	0.91	0.93	0.72	0.80	0.84	0.84	0.86	0.89	0.47	0.57	0.74
3	0.5	0.91	0.88	0.68	0.86	0.80	0.24	0.93	0.89	0.68	0.81	0.69	0.15
3	0.4	0.88	0.86	0.76	0.78	0.78	0.56	0.89	0.88	0.79	0.74	0.71	0.46
3	0.3	0.75	0.77	0.81	0.56	0.64	0.71	0.73	0.76	0.80	0.39	0.48	0.63

Power at false positive rate (FPR) = 5%. Simulations with L = 3 kb. *Tbs*, time in mya since onset of balancing selection; f_{eq}, equilibrium frequency in the simulations. Power on additional conditions is presented on suppl. table S1.

Table 2. Candidate windows and protein-coding genes across populations

Population	LWK				YRI				GBR				TSI			
tf	0.3	0.4	0.5	U	0.3	0.4	0.5	U	0.3	0.4	0.5	U	0.3	0.4	0.5	U
Significant windows	7,364	6,841	6,226	10,072	8,098	7,420	6,854	10,997	6,173	6,858	6,519	9,521	5,998	6,661	6,403	9,2
Outlier windows	829	829	829	1,248	829	829	829	1,230	829	829	829	1,251	829	829	829	1,2
Significant genes	1,182	1,099	1,026	1,457	1,326	1,187	1,110	1,616	1,026	1,094	1,051	1,414	1,039	1,091	1,056	1,4
Outlier genes	151	158	167	249	142	151	157	232	140	157	159	222	143	157	166	23

Note. – Significant and outlier genes and windows, see main text. U, union of windows considering three *tf* values. Total number of queried windows per population is 1,657,989. Union of all candidate genes is 2,348 (significant) and 402 (outlier).

Figure Captions

Fig. 1. A schematic representation of site frequency spectra (SFS) under neutrality and selection, which motivates the NCD statistic. (**A**) Unfolded SFS (ranging from 0 to 1) of derived allele frequencies (DAF) for loci under neutrality (grey) or containing one site under balancing selection with frequency equilibrium (f_{eq}) of 0.5 (blue), 0.4 (orange) and 0.3 (pink). (**B**) Folded SFS (ranging from 0 to 0.5) for minor allele frequencies (MAF). Colors as in **A**. (**C**) Distribution of *NCD* (Non-Central Deviation) expected under neutrality (grey) and under selection assuming $tf = f_{eq}$. Colors as in **A.** *x*-axis shows minimum and maximum values that *NCD* can have for a given *tf* .

Fig. 2. Power to detect balancing selection for *NCD(0.5)* and other tests. The ROC curves summarize the power (in function of the false positive rate) to detect LTBS for simulations where the balanced polymorphism was modeled to achieve f_{eq} of **(A)** 0.3, **(B)** 0.4, and **(C)** 0.5. Plotted values are for the African demography, *Tbs* = 5 mya. L = 3 kb, except for T_1 and T_2 where L = 100 ISs, following (DeGiorgio et al. 2014) (see Methods). For *NCD* calculations, $tf = f_{eq}$. European demography yields similar results (suppl. fig. S10).

Fig. 3. Polymorphism-to-divergence and SFS. (A,B) SNP/(FD+1) for LWK **(A)** and GBR **(B)** populations. SNP/(FD+1) measures the proportion of polymorphic-to-divergent sites for the union of significant windows for all *tf* (purple, green) compared to all scanned windows (gray). **(C-D)** SFS in LWK **(C)** and GBR **(D)** of all scanned windows in chr1 (gray), significant windows for *NCD2(0.5)* (blue), *NCD2(0.4)* (orange), *NCD2(0.3)* (pink). DAF, derived allele frequency.

Fig. 4. Manhattan plot and population sharing. (A) Manhattan plot of all scanned windows, for one analysis (*NCD2(0.5)* and LWK). *y*-axis, *p*-value (log-scale) based on $Z_{tf\text{-}IS}$. *x*-axis, ordered location of analyzed windows on the genome. Each point is a scanned (grey and black), significant (blue) or outlier (pink) window. Names of outlier protein-coding genes are provided, sorted by name. We note that significant windows were defined based of simulations, not on $Z_{tf\text{-}IS}$. ($Z_{tf\text{-}IS}$ is used to rank even those with $p < 0.0001$) **(B)** Venn diagram showing the overlap in signatures of the 167 outlier genes annotated in **(A)** with other populations.

Fig. 5. Enrichment of classes of sites amongst candidate windows. Dashed lines mark the $p = 0.975$ (bottom) and $p = 0.025$ (top) thresholds for the one-tailed *p*-values (hypothesis: enrichment). NSyn, nonsynonymous; all, Genic plus Intergenic plus Regulatory. The annotation is based on Ensembl variant predictor (suppl. information S6). $p < 0.001$ was treated as 0.001 to avoid infinite values.

Figures

Fig. 1

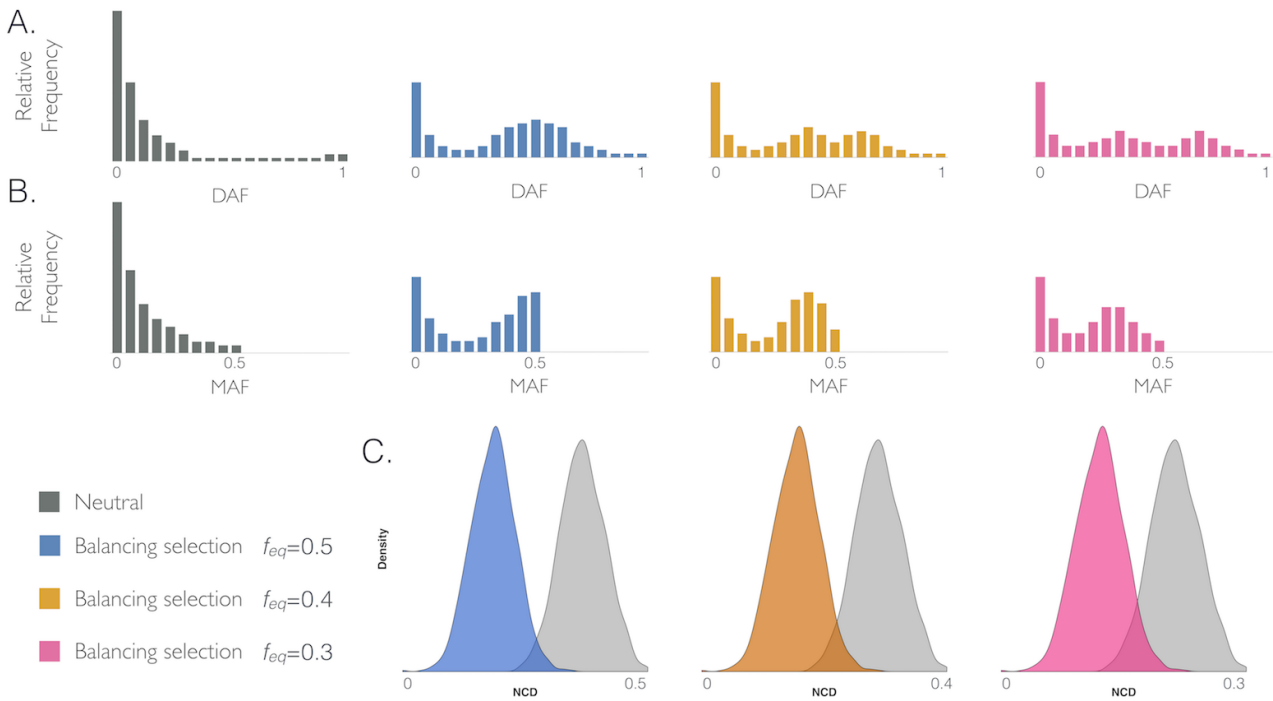


Fig. 2

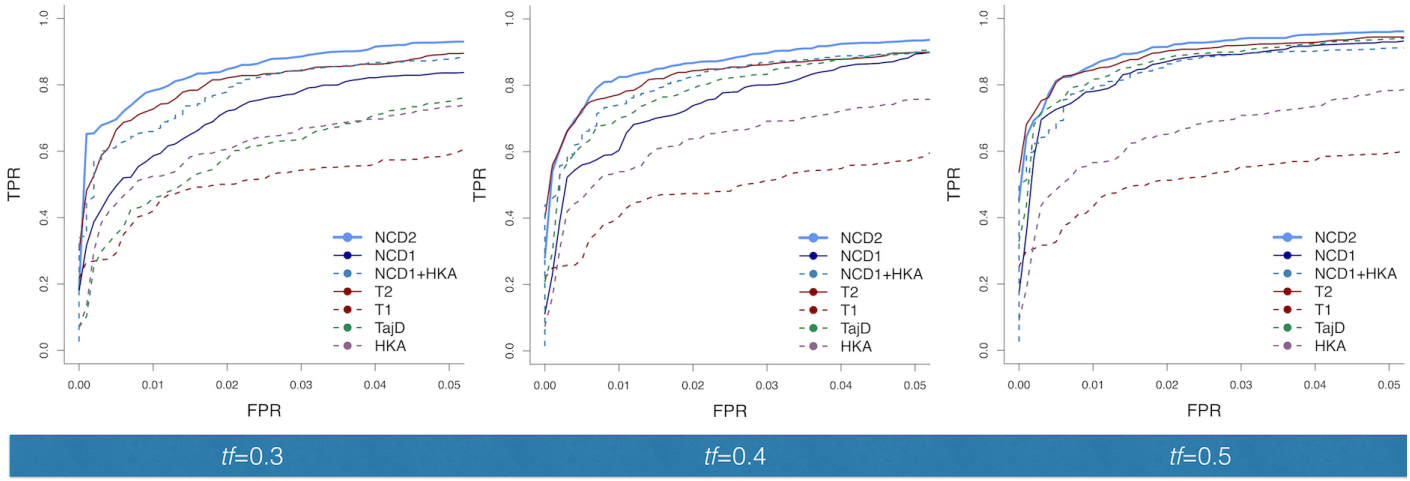

TPR
FPR
NCD2
NCD1
NCD1+HKA
T2
T1
TajD
HKA
tf=0.3
tf=0.4
tf=0.5

Fig. 3

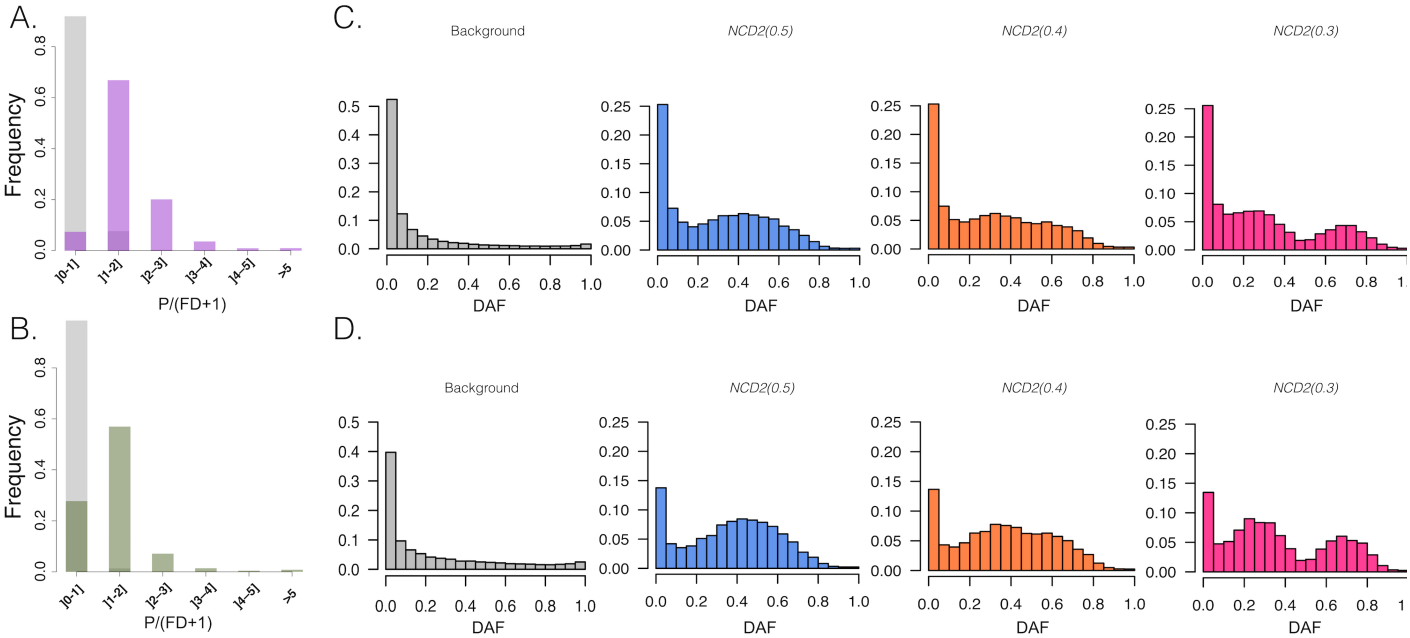

A.
B.
C.
D.
Frequency
P/(FD+1)
[0-1]
[1-2]
[2-3]
[3-4]
[4-5]
>5
Background
NCD2(0.5)
NCD2(0.4)
NCD2(0.3)
DAF

Fig. 4

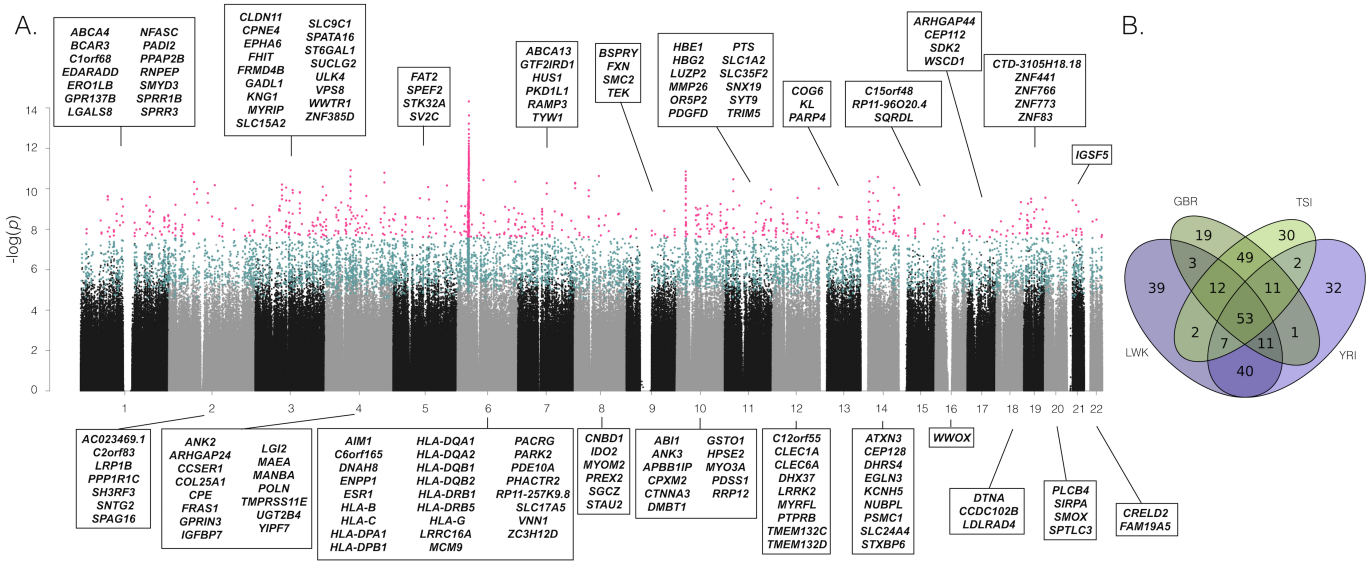
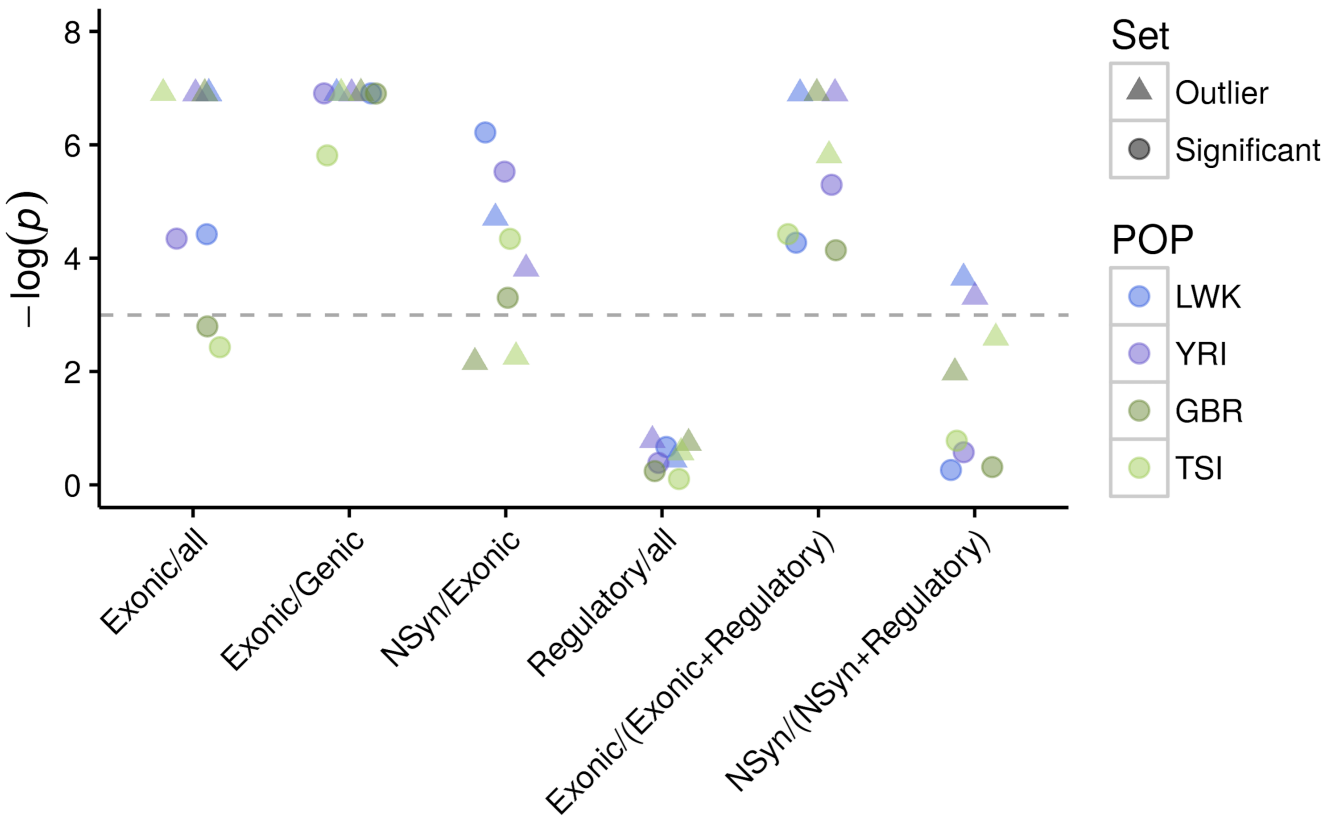

A.
B.
-log(p)
ABCA4 BCAR3 C1orf68 EDARADD ERO1LB GPR137B LGALS8
NFASC PADI2 PPAP2B RNPEP SMYD3 SPRR1B SPRR3
CLDN11 CPNE4 EPHA6 FHIT FRMD4B GADL1 KNG1 MYRIP SLC15A2
SLC9C1 SPATA16 ST6GAL1 SUCLG2 ULK4 VPS8 WWTR1 ZNF385D
FAT2 SPEF2 STK32A SV2C
ABCA13 GTF2IRD1 HUS1 PKD1L1 RAMP3 TYW1
BSPRY FXN SMC2 TEK
HBE1 HBG2 LUZP2 MMP26 OR5P2 PDGFD
PTS SLC1A2 SLC35F2 SNX19 SYT9 TRIM5
COG6 KL PARP4
C15orf48 RP11-96O20.4 SQRDL
ARHGAP44 CEP112 SDK2 WSCD1
CTD-3105H18.18 ZNF441 ZNF766 ZNF773 ZNF83
IGSF5
AC023469.1 C2orf83 LRP1B PPP1R1C SH3RF3 SNTG2 SPAG16
ANK2 ARHGAP24 CCSER1 COL25A1 CPE FRAS1 GPRIN3 IGFBP7
LGI2 MAEA MANBA POLN TMPRSS11E UGT2B4 YIPF7
AIM1 C6orf165 DNAH8 ENPP1 ESR1 HLA-B HLA-C HLA-DPA1 HLA-DPB1
HLA-DQA1 HLA-DQA2 HLA-DQB1 HLA-DQB2 HLA-DRB1 HLA-DRB5 HLA-G LRRC16A MCM9
PACRG PARK2 PDE10A PHACTR2 RP11-257K9.8 SLC17A5 VNN1 ZC3H12D
CNBD1 IDO2 MYOM2 PREX2 SGCZ STAU2
ABI1 ANK3 APBB1IP CPXM2 CTNNA3 DMBT1
GSTO1 HPSE2 MYO3A PDSS1 RRP12
C12orf55 CLEC1A CLEC6A DHX37 LRRK2 MYRFL PTPRB TMEM132C TMEM132D
ATXN3 CEP128 DHRS4 EGLN3 KCNH5 NUBPL PSMC1 SLC24A4 STXBP6
WWOX
DTNA CCDC102B LDLRAD4
PLCB4 SIRPA SMOX SPTLC3
CRELD2 FAM19A5
GBR
TSI
LWK
YRI
19 30 3 49 2 39 12 11 32 53 2 1 7 11 40

Fig. 5


−log(p)
Exonic/all
Exonic/Genic
NSyn/Exonic
Regulatory/all
Exonic/(Exonic+Regulatory)
NSyn/(NSyn+Regulatory)
Set
Outlier
Significant
POP
LWK
YRI
GBR
TSI