Ancient human genomes show that disease-causing mutations have increased in frequency over the past 40,000 years

Stéphane Aris-Brosou,[1,2,⋆]

[1]Department of Biology, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[2]Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON K1N 6N5, Canada

⋆Corresponding author: sarisbro@uottawa.ca

Abstract

The role played by natural selection in shaping present-day human populations has received extensive scrutiny [1, 2, 3], especially in the context of local adaptations [4]. However, most studies to date assume, either explicitly or not, that populations have been in their current locations long enough to adapt to local conditions [5], and that they were large enough to allow for the action of selection [6]. If these conditions were satisfied, not only should selection be effective at promoting local adaptations, but deleterious alleles should also be eliminated over time. To test this prediction, the genomes of 2,062 individuals, including 1,179 ancient humans, were reanalyzed to assess how frequencies of risk alleles and their homozygosity changed through space and time in Europe. While the overall deleterious homozygosity consistently decreased, risk alleles have steadily increased in frequency. Those that increased most are associated with diseases such as asthma, Crohn disease, Diabetes Mellitus and obesity, which are prevalent in present-day populations. These findings do not run against the existence of local adaptations, but highlight the limitations imposed by drift and range expansions on the strength of selection in purging the mutational load affecting human populations.

Keywords: negative selection, ancient DNA, SNP, GWAS, effective population size, CADD.

In order to trace the fate of deleterious alleles in human populations over time, the genotypes of 2,062 individuals, of which 1,179 were ancient humans, were assessed at about 1.2 million Single Nucleotide Polymorphisms (SNPs; the '1240k capture', [7]). The present study focuses on individuals sampled throughout Europe, from the Atlantic to the Ural Mountains, as this is where most ancient DNA has been recovered and genotyped throughout the world. To take into account the history of admixture and replacement events characterizing Europe [8, 7], this region was split into four quadrants along two orthogonal axes roughly going through the Alps (Figure 1). These four quadrants represent approximately four distinct areas that played significant roles in the three peopling waves that affected Europe, first from the Fertile Crescent (Q4 in Figure 1) to the rest of Europe during the Paleo-/Mesolithic period, between 45 and 10 thousands years ago (kya) [9, 10]. These early Europeans were then replaced after the Last Glacial Maximum, 10 kya, by Neolithic farmers [11] who reached the Iberian Peninsula (Q3), and Britain and Scandinavia (Q2) 7-6 kya [12]. A third major migration event swept into Europe, including central Europe (Q1), during the late Neolithic and the early Bronze Age, about 3.5 kyr ago [7, 13]. Under the constraint of analyzing approximately even numbers of individuals across these four quadrants, the timing of these migration events suggested splitting Europe into four different time periods, representing roughly the Mesolithic (t1: 45-10 kya), the Neolithic (t2: 10-3.5 kya), the Bronze Age (t3: 3.5-0 kya), and present-day populations (t4: 0-present) as in Figure 1.

To find which genomic positions among these 1.2 million SNPs are deleterious, a database of markers identified based on genome-wide association studies (DisGeNET) was used [14]. This database contains 83,582 SNPs, but not the associated risk alleles, which were retrieved separately from SNPedia [15] and SelfDecode (`www.selfdecode.com`). Overall, the average frequency of deleterious mutations significantly increased through

each time period in all four quadrants (Figure 2). One possible explanation for this non-adaptive pattern is that most of the SNPs tested are weakly deleterious, and increased in frequency as neutral mutations would have [16].

To test if weakly deleterious mutations are driving the increase in risk allele frequency, their relative pathogenicity was assessed by the Combined Annotation Dependent Depletion (CADD) scores [17], which take values from -6 (the least deleterious) to 100. The values observed here range from about -5 to 15, are sharply distributed around 0, and are skewed to the right (Figure 3). To cover the observed CADD scores, their range was split into five bins. In all four quadrants, the mutations that are expected to be the most weakly deleterious (up to CADD $= 2.5$) significantly increased in frequency (Figure 2A). As these mutations are the most numerous (Figure 3), they are most likely responsible for the overall increasing frequency pattern. However, the CADD analysis also reveals that the most deleterious mutations, with CADD > 10, did not significantly decrease in frequency (Figure 2A). If these mutations were recessive, the deleterious alleles would mostly be present in the heterozygous genotypes, and their homozygosity would be low.

To assess the effect of dominance on the general pattern of increasing frequency of risk alleles through time, the homozygosity of the risk alleles was calculated (Figure 2B). Overall, homozygosity decreased significantly through time, but again, this pattern is mostly driven by the least deleterious alleles (up to CADD $= 2.5$). The most deleterious ones (CADD > 10), on the other hand, saw their heterozygosity decrease recently (in t4) only in southwestern Europe (Q3). In all three other quadrants, their heterozygosity did not change significantly (Figure 2B), suggesting that these most deleterious mutations are only partially recessive.

To better understand which diseases increase in frequency through time, robust linear models were fitted to each SNP, irrespective of its CADD score. From the corresponding

volcano plot (Figure 4A), two groups of SNPs can be defined: those whose frequency significantly decrease (top left area of volcano plot) or increase (top right area) through time. In the first group, quadrant Q1 has both the most negative slopes and the smallest CADD scores (Figure 4B; Dunn test, FDR < 0.05), and the most affected phenotypes, with a CADD score greater than 1, include "age at menarche" and "drug dependence" (Table S1). In the second set, quadrant Q2 has the most positive slopes even if all quadrants has similar CADD scores (Figure 4C; Dunn test, FDR < 0.05), and the most affected phenotypes, with a CADD score greater than 1, include "asthma," "Crohn disease," "Diabetes Mellitus" and "obesity" (Table S1), diseases that tend to be frequent in present-day European populations. Enrichment analyses confirmed that SNPs involved in addictive behavior are overrepresented in the set of SNPs with decreasing frequency, while common diseases are overrepresented in the group with increasing frequencies. Note that in both sets of SNPs, the CADD scores are never greater that 10, suggesting that the most deleterious alleles either do not change in frequency, and / or do so at a very slow rate – which again is consistent with their not being eliminated through time.

The efficacy of selection at removing deleterious mutations depends on fitness effects, with the most deleterious variants being eliminated with the greatest ease. However, this efficacy can be affected by a number of factors such as the effective population size, which is arguably the most critical [18]. Indeed, ancient human demographics is marked during the Last Glacial Maximum (LGM) by a severe bottleneck, which left traces in present-day individuals of European ancestry [19] as well as in other worldwide populations [20]. The results presented here are therefore consistent with the population genetics context, which permitted the increase in frequency over time of deleterious variants.

Although the variance of allele frequencies in the oldest time period (t1: the Mesolithic) is very large, their median is also quite high (~ 0.25), begging the question of the origin of

these variants. Previous work showed that some of these disease-causing variants found in modern humans could have a Neanderthal origin, but it was also suggested that selection could be efficient at removing these variants [21], especially in East Asians, as selection is mostly acting on weakly deleterious mutations located in protein-coding genes [22]. This is not quite consistent with the results presented here, where frequencies of even the most weakly deleterious variants do not significantly change over time. This is potentially due to most of the SNPs used here being located outside of protein-coding genes.

An allelic frequency however represents a snapshot of a dynamic process, where variants are generated by mutations, and eliminated by selection (or lost by drift). The initial increase, between the Mesolithic and the Neolithic (t1 and t2, respectively), could be due to drift, but also to an increased mutational pressure. Little evidence however seems to support this possibility, except maybe for recent times, where humans tend to expose themselves to mutational agents, leading to a relaxation of selection against mildly deleterious mutations [23]. As a result, nonadaptive factors following the initial LGM bottleneck remain the most plausible explanation for the monotonic increase in frequency of deleterious alleles in the European human population.

Again, this is consistent with the demographic context, as this geographic region has been affected by at least three migration waves, associated with rapid demographic expansion(-s) [19]. Indeed, range expansions have been shown to enable new deleterious mutations to behave almost like neutral variants, hereby facilitating their increase in frequency [24]. This would suggest that (i) the deleterious mutations observed post-LGM are novel, *i.e.* do not have a Neanderthal origin, and (ii) increase in frequency because of a nonadaptive factor (repeated range expansions). One signature that is missing from typical range expansions is a frequency gradient along the expansion axis (from Q4 to the other quadrants; [25]); this absence of gradient may however be caused by multiple

expansions, whose individual directions are not necessarily overlapping. This may also be why an increase in the homozygosity of deleterious variants was not observed here. As a result, the existence of geographic regions of high fitness in which the effects of deleterious mutations would be partially masked, as predicted by range expansion models [26], seems unlikely in European humans.

The assessment of the phenotypic effects of variants was based on CADD scores, which correlate with a number of experimental data related to pathogenicity, disease severity, *etc.* [17], but that also show a number of discrepancies in select cases [27], or in a study of 12,391 cancer-related SNPs [28]. While this number represents 12% of the SNPs analyzed here, it is unlikely that the results presented here are affected by the predictive power of CADD scores. Indeed, previous studies used a different score, the Genomic Evolutionary Rate Profiling (GERP) Rejected Substitution (RS) score, to predict the phenotypic impact of mutations [29], and showed either the existence of a spatial increase of deleterious mutations from Africa to the rest of the world, consistent with the expansion axes out of Africa [20], or an increase of deleterious alleles at the front of a range expansion compared to the inner core of the source population [30]. The data presented here add a temporal dimension to these previous spatial results, which were found at a broader scale. It is therefore possible that the results presented here, limited to Europe, be generalizable to the rest of the non-Subsaharan human populations. As more ancient DNA from the rest of the world is accumulating [31, 32, 33], this hypothesis will soon be testable.

Altogether, the results presented here add temporal evidence to a debate that has mostly focused on space (as a proxy for time), to understand the relative strength of selection and demography in shaping the mutational load in human populations [34, 35, 36, 24]. The crux of this debate actually hinges on the dominance of deleterious mutations, as theory predicts that load is greater under an additive model than under a recessive

model [37], while both experimental [38] and empirical [20] evidence suggests that the most deleterious mutations are more likely to be recessive, but not always [39]. The results presented here provide evidence that partially recessive variants with large effects can participate in the nonadaptive buildup of mutational load in human populations.

Materials and methods

Genotypic data

Complete genotypic data were retrieved for nine ancient DNA studies that mostly focused on European individuals [8, 7, 40, 41, 42, 43, 44, 45, 46] from the David Reich website (`reich.hms.harvard.edu`). The geno files were first converted into ped format with EIGENSTRAT v.6.1.4 [47], and the map files that contain information relative to SNP identifiers were created with a custom R script. The human reference genome, as well as those of the Neanderthal, Denisovan, and the Chimpanzee were removed. The date and sampling coordinates of each of the 3,409 unique individual were extracted from the Supplementary Information files of each publication. Individuals from outside of Europe were eliminated to keep only those north of the Arabian peninsula and south of the Arctic (latitudes $\in [20°, 80°]$), and west of Iceland / Cabo Verde and east of Mount Ural (longitudes $\in [-10°, 60°]$). This European region was further divided into four quadrants Q1 to Q4, placing the crosshair at 12°W and 47°N, so that each region has similar number of individuals in each time period t1 to t4 (present-day individuals: 126 in Q1; 67 in Q2; 166 in Q3; 524 in Q4: total present-day 883; ancient individuals: 410 in Q1; 315 in Q2; 134 in Q3; 320 in Q4: total ancient 1,179). These time periods correspond roughly to the mesolithic (t1: 45,000–10,000 BCE), the neolithic (t2: 10,000–3,500 BCE), the bronze age (t3: 3,500–0 BCE), and modern humans (t4: the 2000 CE).

Each individual was genotyped at 1,233,013 SNPs (the '1240k capture', [7]), and hence contained information for twice as many or 2,466,026 alleles. To avoid computational issues related to loading large files (with 2,466,026 rows for 3,406 individuals) into memory, each genome was saved to an individual file.

Phenotypic data

Disease data were retrieved from DisGeNET [14] at `www.disgenet.org/ds/DisGeNET/results/curated_variant_disease_associations.tsv.gz`, which contained 83,582 SNP identifiers involved in genetic diseases. Severity of each SNP was assessed by their Combined Annotation Dependent Depletion (CADD) score [17]. For this, a list of all possible SNVs in GRCh37/hg19 was downloaded from the CADD website (`krishna.gs.washington.edu/download/CADD/v1.3/whole_genome_SNVs.tsv.gz`), split into multiple files containing 100 million lines each (to avoid loading a single 250GB text file in memory), from CADD score was extracted at each SNPs in DisGeNET. At each SNP, the identity of the risk allele was extracted from SNPedia [15] and SelfDecode (`www.selfdecode.com`).

Matching genotypic and phenotypic data

In each quadrant and each time period, a list of corresponding individual was extracted, and used to estimate the frequency of the risk allele in each quadrant / time period combination. Significant frequency differences were assessed with the Dunn test, with Benjamini-Hochberg correction for multiple comparisons at a FDR ≤ 0.05 unless otherwise stated. Robust linear regressions were used to find SNPs with decreasing and increasing frequency through time. Enrichment analyses for these decreasing and increasing SNPs

were done with the clusterProfiler R package [48], based on lists of gene names containing the SNPs of interest, with respect to all gene names with SNPs in DisGeNET. All analyses were performed with R [49] ver. 3.2.3; all scripts are available at `github.com/sarisbro`.

Acknowledgements

I would like to thank first and foremost David Reich for making all the genotyping data produced in his lab freely available online. My gratitude also goes to Kelly Burkett for providing extensive comments. This work was partly conducted while being hosted by Yutaka Watanuki, at the University of Hokkaido in Hakodate, thanks to an Invitational Fellowship from the Japanese Society for the Promotion of Science, and was supported by the Natural Sciences Research Council of Canada. It would not have been possible without assistance from the Center for Advanced Computing and Compute Canada.

References

1. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C. & Clark, A. G. Recent and ongoing selection in the human genome. *Nat Rev Genet* **8**, 857–68 (2007).

2. Novembre, J. & Di Rienzo, A. Spatial patterns of variation due to natural selection in humans. *Nat Rev Genet* **10**, 745–55 (2009).

3. Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017).

4. Fan, S., Hansen, M. E. B., Lo, Y. & Tishkoff, S. A. Going global by adapting local: A review of recent human adaptation. *Science* **354**, 54–59 (2016).

5. Pickrell, J. K. & Reich, D. Toward a new history and geography of human genes informed by ancient dna. *Trends Genet* **30**, 377–89 (2014).

6. Gravel, S. When is selection effective? *Genetics* **203**, 451–62 (2016).

7. Haak, W. *et al.* Massive migration from the steppe was a source for indo-european languages in europe. *Nature* **522**, 207–11 (2015).

8. Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day europeans. *Nature* **513**, 409–13 (2014).

9. Higham, T. *et al.* The timing and spatiotemporal patterning of neanderthal disappearance. *Nature* **512**, 306–9 (2014).

10. Benazzi, S. *et al.* Early dispersal of modern humans in europe and implications for neanderthal behaviour. *Nature* **479**, 525–8 (2011).

11. Günther, T. & Jakobsson, M. Genes mirror migrations and cultures in prehistoric europe-a population genomic perspective. *Curr Opin Genet Dev* **41**, 115–123 (2016).

12. Günther, T. *et al.* Ancient genomes link early farmers from atapuerca in spain to modern-day basques. *Proc Natl Acad Sci U S A* **112**, 11917–22 (2015).

13. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–72 (2015).

14. Piñero, J. *et al.* Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* **45**, D833–D839 (2017).

15. Cariaso, M. & Lennon, G. Snpedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res* **40**, D1308–12 (2012).

16. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–8 (1973).

17. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310–5 (2014).

18. Lynch, M. *The origins of genome architecture* (Sinauer Associates, Sunderland, Mass., 2007). URL `http://www.loc.gov/catdir/toc/ecip077/2007000012.html`.

19. Prüfer, K. *et al.* The complete genome sequence of a neanderthal from the altai mountains. *Nature* **505**, 43–9 (2014).

20. Henn, B. M. *et al.* Distance from sub-saharan africa predicts mutational load in diverse human genomes. *Proc Natl Acad Sci U S A* **113**, E440–9 (2016).

21. Sankararaman, S. *et al.* The genomic landscape of neanderthal ancestry in present-day humans. *Nature* **507**, 354–7 (2014).

22. Juric, I., Aeschbacher, S. & Coop, G. The strength of selection against neanderthal introgression. *PLoS Genet* **12**, e1006340 (2016).

23. Lynch, M. Mutation and human exceptionalism: Our future genetic load. *Genetics* **202**, 869–75 (2016).

24. Peischl, S., Dupanloup, I., Bosshard, L. & Excoffier, L. Genetic surfing in human populations: from genes to genomes. *Curr Opin Genet Dev* **41**, 53–61 (2016).

25. Peischl, S., Dupanloup, I., Kirkpatrick, M. & Excoffier, L. On the accumulation of deleterious mutations during range expansions. *Mol Ecol* **22**, 5972–82 (2013).

26. Peischl, S. & Excoffier, L. Expansion load: recessive mutations and the role of standing genetic variation. *Mol Ecol* **24**, 2084–94 (2015).

27. van der Velde, K. J. *et al.* Evaluation of cadd scores in curated mismatch repair gene variants yields a model for clinical validation and prioritization. *Hum Mutat* **36**, 712–9 (2015).

28. Mather, C. A. *et al.* Cadd score has limited clinical validity for the identification of pathogenic variants in noncoding regions in a hereditary cancer panel. *Genet Med* **18**, 1269–1275 (2016).

29. Cooper, G. M. *et al.* Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat Methods* **7**, 250–1 (2010).

30. Peischl, S. *et al.* Relaxed selection during a recent human expansion. *Genetics* **208**, 763–777 (2018).

31. Malaspinas, A.-S. *et al.* A genomic history of aboriginal australia. *Nature* **538**, 207–214 (2016).

32. Skoglund, P. *et al.* Reconstructing prehistoric african population structure. *Cell* **171**, 59–71.e21 (2017).

33. Lipson, M. *et al.* Population turnover in remote oceania shortly after initial settlement. *Curr Biol* **28**, 1157–1165.e7 (2018).

34. Lohmueller, K. E. *et al.* Proportionally more deleterious genetic variation in european than in african populations. *Nature* **451**, 994–7 (2008).

35. Do, R. *et al.* No evidence that selection has been less effective at removing deleterious mutations in europeans than in africans. *Nat Genet* **47**, 126–31 (2015).

36. Simons, Y. B., Turchin, M. C., Pritchard, J. K. & Sella, G. The deleterious mutation load is insensitive to recent population history. *Nat Genet* **46**, 220–4 (2014).

37. Henn, B. M., Botigué, L. R., Bustamante, C. D., Clark, A. G. & Gravel, S. Estimating the mutation load in human genomes. *Nature Reviews Genetics* **16**, 333 (2015).

38. Agrawal, A. F. & Whitlock, M. C. Inferences about the distribution of dominance drawn from yeast gene knockout data. *Genetics* **187**, 553–66 (2011).

39. Joseph, S. B., Peck, K. M. & Burch, C. L. Dominance effects of deleterious and beneficial mutations in a single gene of the rna virus 6. *PLoS One* **9**, e97717 (2014).

40. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient eurasians. *Nature* **528**, 499–503 (2015).

41. Fu, Q. *et al.* The genetic history of ice age europe. *Nature* **534**, 200–5 (2016).

42. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient near east. *Nature* **536**, 419–24 (2016).

43. Unterländer, M. *et al.* Ancestry and demography and descendants of iron age nomads of the eurasian steppe. *Nat Commun* **8**, 14615 (2017).

44. Lipson, M. *et al.* Parallel palaeogenomic transects reveal complex genetic history of early european farmers. *Nature* **551**, 368–372 (2017).

45. Mathieson, I. *et al.* The genomic history of southeastern europe. *Nature* **555**, 197–203 (2018).

46. Olalde, I. *et al.* The beaker phenomenon and the genomic transformation of northwest europe. *Nature* **555**, 190–196 (2018).

47. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–9 (2006).

48. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012).

49. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2015). URL `https://www.R-project.org/`.

Figures

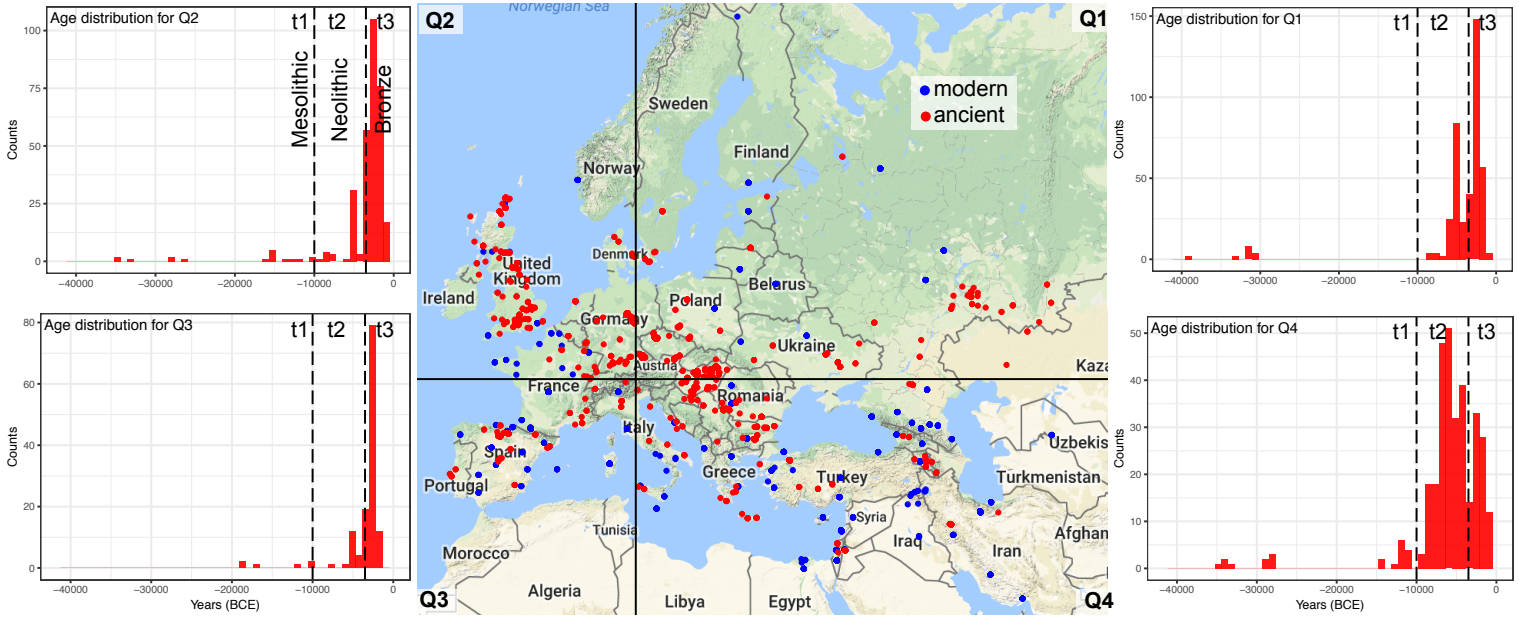

Figure 1. Spatiotemporal distribution of the human genomes analyzed in this study. Center: Sampling location of modern (blue) and ancient (red) individuals throughout Europe. The maps is divided into four quadrants representing northeastern (Q1), northwestern (Q2), southwestern (Q3), and southeastern (Q4) Europe. By each quadrant, the age distributions of the ancient humans are shown; the four time periods considered correspond roughly to the mesolithic (t1: 45,000–10,000 BCE), the neolithic (t2: 10,000–3,500 BCE), the bronze age (t3: 3,500–0 BCE), and modern humans (t4: the 2000 CE). Map background: Google (2018).

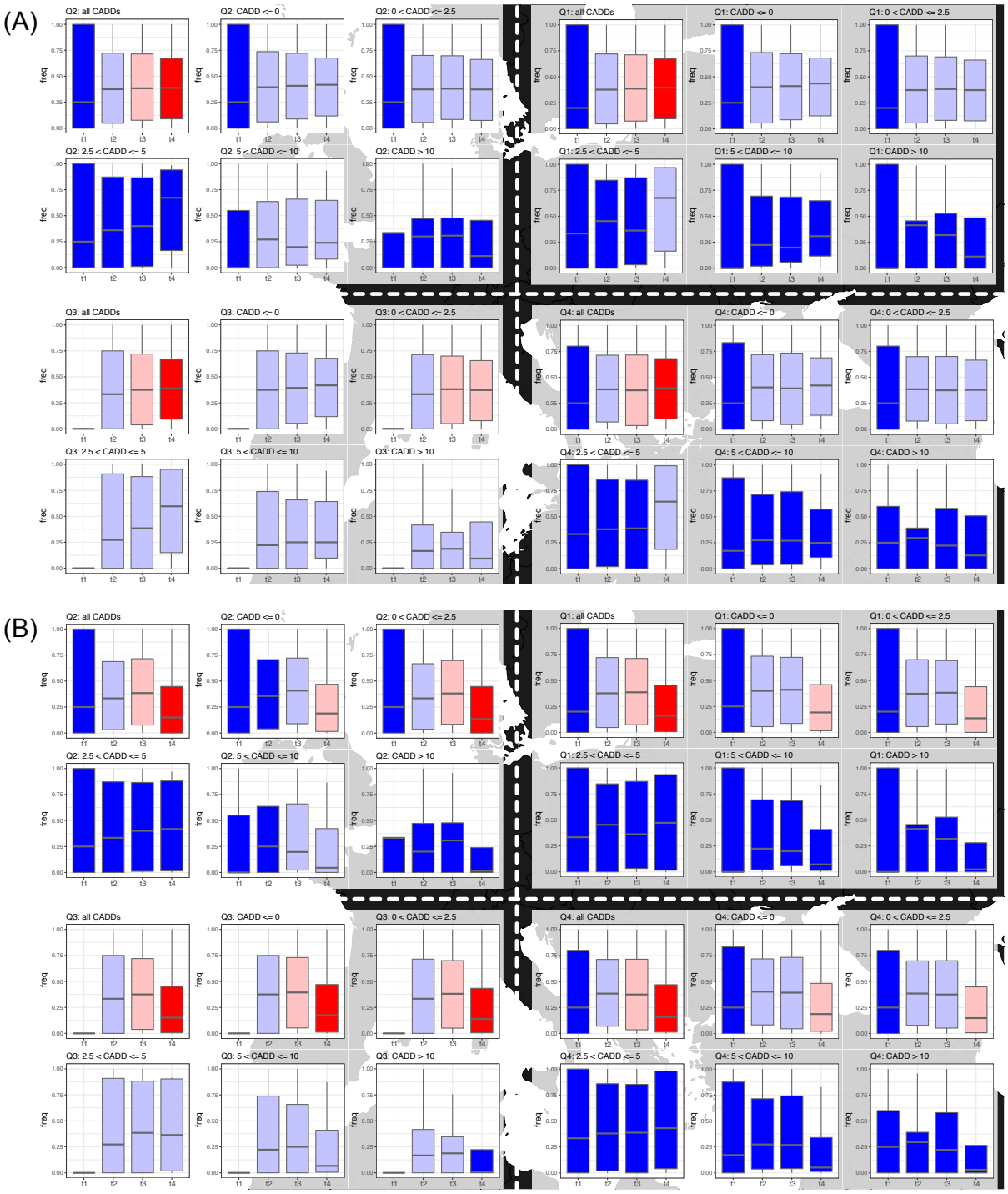

Figure 2. Evolution of SNP frequencies through space and time. (A) Frequencies of risk alleles. (B) Frequencies of homozygous genotypes at risk alleles. The top left panel in each quadrant show the frequency distributions at all SNPs over the four time periods, while the next five panels show the frequency distributions at SNPs in each severity category, ranging from benign (CADD < 0) to severe (CADD ≥ 10) effects. Color differences in each panel indicate significance of the Dunn test at the 1% level with Benjamini-Hochberg correction for multiple tests – colors are ordered from significantly lower (cool colors) to significantly higher (warm colors) frequencies of the risk alleles.

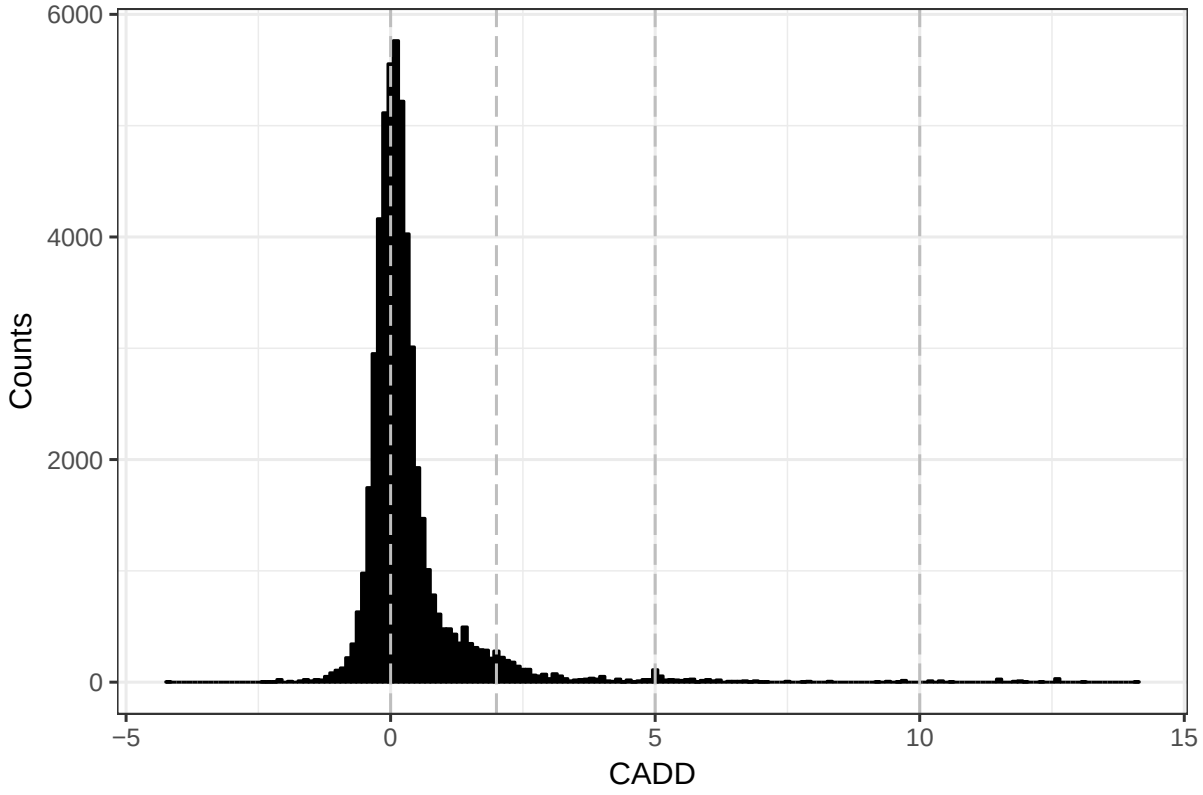

Figure 3. Distribution of CADD scores for the SNPs in DisGeNET. CADD scores are shown for the 83,582 SNPs contained in DisGeNET 4.0 database. Vertical lines split SNPs into five categories ranging from inferred to have benign effects (CADD < 0) to severe effect (CADD ≥ 10).

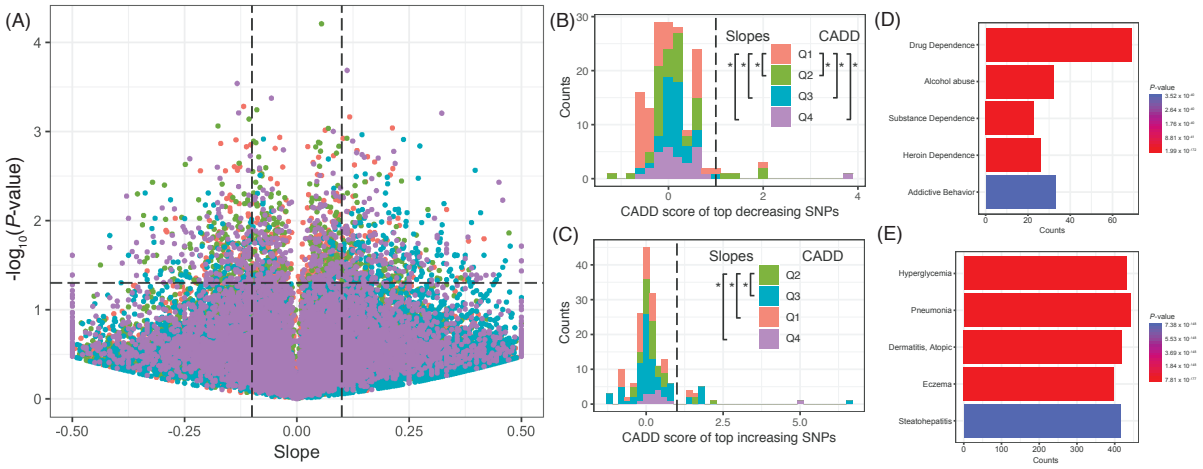

Figure 4. Volcano plot of the regressions of SNP frequencies through time. (A) The plot shows the P-values, transformed on a $-\log_{10}$ scale, as a function of the slope estimates of the robust regressions, with the four quadrants color coded (Q1: salmon, Q2: green, Q3: blue, Q4: magenta). (B) Distribution of the CADD scores for the SNPs whose frequency decreases through time ($P \leq 0.05$, slope ≥ 0.1). The Dunn test was used to assess differences between slopes (left of inset legend) and CADD scores (right of inset legend) at an FDR ≤ 0.05. (C) Same as in (B) for the SNPs whose frequency increases through time. Top five enriched gene names among the top decreasing (D) and top increasing (E) genes identified from the volcano plot with CADD scores ≥ 1.

Supplementary table

Supplementary table 1. Sets of SNPs and their associated phenotype that are significantly decreasing (top) and increasing (bottom) through time. For each SNP, the table shows the quadrant ID, the SNP frequency at times t1 to t4, the SNP identifier, the disease identifier and name from DisGeNET, the CADD score and the slope and *P*-value from the robust regression of SNP frequency *vs.* time.

Decreasing	Quad ID	t1	t2	t3	t4	SNP ID	Disease ID	Disease name	CADD	Slopes	*P*-values
	Q4	1.0000	0.7532	0.7059	0.5602	rs1052030	C1568247	Usher Syndrome Type I	3.8589	-0.1367	0.0355
	Q2	1.0000	0.8462	0.7303	0.6970	rs807601	C0201976	Creatinine measurement serum	2.0105	-0.1025	0.0337
	Q2	1.0000	0.8462	0.7303	0.6970	rs807601	C1561549	Glomerular filtration rate finding	2.0105	-0.1025	0.0337
	Q1	1.0000	0.6582	0.5660	0.3577	rs951366	C1314691	Age at menarche	1.9794	-0.2019	0.0263
	Q2	1.0000	0.8182	0.6923	0.5909	rs12995333	C1510472	Drug Dependence	1.4576	-0.1353	0.0090
	Q2	1.0000	0.7750	0.5955	0.5149	rs913588	C1314691	Age at menarche	1.1493	-0.1635	0.0195

Increasing	Quad ID	t1	t2	t3	t4	SNP ID	Disease ID	Disease name	CADD	Slopes	*P*-values
	Q3	0.0000	0.0833	0.2745	0.3855	rs2304725	C0004096	Asthma	6.6726	0.1348	0.0106
	Q4	0.5000	0.7034	0.7727	0.8922	rs3816183	C0848558	Hypospadias	5.0550	0.1246	0.0216
	Q2	0.0000	0.0882	0.2875	0.2955	rs604723	C0428886	Mean blood pressure	2.2455	0.1086	0.0487
	Q3	0.0000	0.3333	0.8889	0.9458	rs11209026	C0009324	Ulcerative Colitis	1.7046	0.3393	0.0370
	Q3	0.0000	0.3333	0.8889	0.9458	rs11209026	C0010346	Crohn Disease	1.7046	0.3393	0.0370
	Q3	0.0000	0.3333	0.8889	0.9458	rs11209026	C0021390	Inflammatory Bowel Diseases	1.7046	0.3393	0.0370
	Q3	0.0000	0.3333	0.8889	0.9458	rs11209026	C0033860	Psoriasis	1.7046	0.3393	0.0370
	Q3	0.0000	0.3333	0.8889	0.9458	rs11209026	C0038013	Ankylosing spondylitis	1.7046	0.3393	0.0370
	Q2	0.2000	0.4318	0.6000	0.6642	rs507506	C2700366	Adiponectin Measurement	1.5824	0.1561	0.0280
	Q1	0.0000	0.0513	0.2018	0.3306	rs12521868	C0010346	Crohn Disease	1.5272	0.1142	0.0167
	Q2	0.0000	0.0625	0.2080	0.3769	rs12521868	C0010346	Crohn Disease	1.5272	0.1276	0.0180
	Q1	0.0000	0.1800	0.2568	0.3387	rs2617170	C0004943	Behcet Syndrome	1.4717	0.1093	0.0241
	Q3	0.0000	0.2500	0.3571	0.5934	rs1421085	C0011860	Diabetes Mellitus Non-Insulin-Dept	1.3791	0.1887	0.0104
	Q3	0.0000	0.2500	0.3571	0.5934	rs1421085	C0028754	Obesity	1.3791	0.1887	0.0104
	Q1	0.3333	0.5000	0.6173	0.6613	rs9854771	C0235974	Pancreatic carcinoma	1.3591	0.1101	0.0299