Ancient implications for today's precision medicine: How the first Near East farmers shaped the European genetic risk architecture for IBD

Ben Krause-Kyora (b.krause-kyora@ikmb.uni-kiel.de)
Kiel University https://orcid.org/0000-0001-9435-2872

Guillermo Torres
Institute of Clinical Molecular Biology, Kiel University https://orcid.org/0000-0002-4548-7815

Nicolas da Silva
Institute of Clinical Molecular Biology, Kiel University

Daniel Kolbe
Institute of Clinical Molecular Biology, Kiel University

Janina Dose
Institute of Clinical Molecular Biology, Kiel University

Sabine Schade-Lindig
Landesamt für Denkmalpflege Hessen, hessenARCHÄOLOGIE, Schloss Biebrich

Joachim Wahl
Institute for Archaeological Sciences, Palaeoanthropology Working Group, University of Tübingen

Carola Berszin
Anthropologische Dienstleistungen Konstanz

Michael Francken
Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart

Irina Görner
Museumslandschaft Hessen Kassel, Sammlung Vor- und Frühgeschichte

Kerstin Schierhold
LWL-Altertumskommission für Westfalen

Joachim Pechtl
Institut für Archäologie, University of Innsbruck https://orcid.org/0000-0002-3478-9084

Gisela Grupe
Biocenter of the Ludwig-Maximilians-Universität

Amke Caliebe
Institute of Medical Informatics and Statistics, Kiel University

Johannes Müller
Institute of Pre- and Protohistoric Archaeology, Kiel University

Stefan Schreiber

Institute of Clinical Molecular Biology, Kiel University

Almut Nebel

Institute of Clinical Molecular Biology, Kiel University

Article

Keywords:

Posted Date: September 28th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2075746/v1

Abstract

Inflammatory bowel disease (IBD) is often described as a model for modern civilization diseases in which environmental factors trigger disease manifestation in genetically compromised individuals. Little is known about the evolutionary history of variants associated with IBD in modern Europeans. Here, we analysed 610 IBD-variants in 2445 ancient datasets from human remains spanning the last 12,000 years, including genotypes generated from 172 newly collected individuals from the European Neolithic. We found statistically significant differences in the frequencies of 97 IBD variants between Neolithic and modern populations that can be explained by the adoption of an agricultural lifestyle and behaviour and concomitant possible microbiome changes in the earliest farmers. Later admixture events and selection against pathogens largely influenced the genetic risk architecture of IBD in contemporary Europeans. A better understanding of the evolutionary history of disease variants is an important first step in translating genetic findings into preventive health care.

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the two main subphenotypes of inflammatory bowel disease (IBD). IBD is seen as paradigmatic for civilization diseases that arise from perturbed gene-by-environment interactions. More than 200 lead single-nucleotide variants (SNVs) with predisposing or protective effect on IBD have been detected so far. Interestingly, many of the implicated loci have been found to play a role in other inflammatory diseases as well, suggesting in part a shared etiology[1]. While the exact triggers for disease manifestation remain unknown, lifestyle factors such as a western diet, use of antibiotics and hygiene are often blamed for contributing to the development of IBD through altering the composition of the intestinal microbiota[2].

The genetic risk architecture paints a colorful picture of mechanisms that are involved in the pathogenesis of IBD. Most importantly, genetic discovery has highlighted an impaired barrier function, intestinal epithelial cell-microbiome interactions and cellular health as important themes that add to a dysregulation of adaptive immunity. Despite the large number of identified IBD-associated variants, not much is known about their history, i.e., when and where they emerged, or why they are still present in the human gene pool, sometimes as coding variants at relatively high frequencies (e.g., *ATG16L1*[3]). Many IBD loci have been affected by natural selection[4–9], with positive and counterbalancing selection in populations often leading to the enrichment or retention of alleles with proinflammatory and enhanced immune functions (e.g., in *CARD9, FUT2, ATG16L* or *IL23R*)[5,8]. These variations might represent beneficial adaptations to recurrent and severe epidemics that humans experienced throughout history[5,10]. Today the same alleles are risk loci for IBD, most likely due to mismatches with modern environmental and lifestyle conditions[10,11].

With one exception[12], evolutionary investigations of genetic diversity related to IBD have relied on data from modern populations (e.g.,[5,8]), which results in a number of limitations. For example, such studies do

not allow reliable conclusions about the place and date of origin of individual variants, or the extent and pace of their subsequent geographical spread. Ancient genome analyses can provide more accurate estimates as they examine human remains from specific locations and periods, making it possible to observe mutations as they occurred in a historical population and track their frequencies over time[13].

Another advantage of ancient genomic data is that they enable researchers to disentangle the effects of selection from those of past demographic events such as admixture, bottlenecks or expansions that may also have influenced the spread of variants that are the genetic basis of IBD today. Recently, it has been shown that each of the three prehistoric source populations (western hunter-gatherers, Neolithic farmers from the Near East and Steppe populations) that essentially formed the gene pool of contemporary Europeans also contributed certain ancestry-specific signatures to the modern diversity of complex traits and diseases[14–16].

Here, we systematically analysed IBD variants associated in modern Europeans in various ancient population samples from Eurasia. Given the importance of nutrition and the intestinal microbiome in IBD pathogenesis, we first focused on the European Neolithic (7000 - 2200 BCE) when the introduction of agriculture led to drastic dietary changes[17]. For this purpose, we generated 172 new datasets from this period. In a second step, we included genotype information from additional ancient populations, which allowed us to establish allele frequency trajectories and identify major shifts over the last 12,000 years.

Results

In this study, we analysed the history of alleles that are known IBD risk or protective factors in the European population today. For this purpose, 610 SNVs associated with IBD, CD and UC were enriched and genotyped in ancient DNA (aDNA) libraries generated from human skeletal remains that were excavated from 7 Neolithic sites. The investigated Neolithic individuals represent two populations, i.e., early farmers (EF, n = 47, 5500-4500 BCE) and late farmers (LF, n = 125, 3500-2900 BCE) that differ considerably in their genomic compositions as shown previously[18,19]. After quality control filtering, 595 SNVs were available for analysis. Since many of these alleles were correlated with each other in the Neolithic populations, we applied pairwise variant pruning that resulted in 353 independent SNVs (Supplementary Table 1). Of these, 97 showed a statistically significant minor allele frequency (MAF) difference between at least one of the two Neolithic groups and a modern reference dataset of Non-Finnish Europeans (NFE; n ≈ 7670) (Supplementary Table 1). A permutation test revealed that the 97 significant markers represented a significantly higher number of variants than expected by chance ($p < 2.2 \times 10^{-16}$). Of the 97 alleles, only a few have been functionally linked to IBD, such as the 7 missense SNVs shown to be involved in microbiome sensing and/or T_H17 signaling (Table 1).

Table 1. Overview of the 7 missense mutations

Gene	rsID	Trajectory	Functional allele	IBD OR	CD OR	UC OR	Protein change	Effect
PTPN22	rs2476601	2	A (protective)	-	0.79	-	R [CGG] > W [TGG]	Protective effect in CD by mediating the intestinal microbiome[20–22]
FCGR2A	rs1801274	1	A	1.12	1.8	1.19	H [CAT] > R [CGT]	Adaptive immunity, microbe sensing[23,24]
IL23R	rs11209026	2	A (protective)	0.50	0.38	0.57	R [CGA] > Q [CAA]	Downregulating the T_H17 immune response and alleviating inflammation[25]
PLAU	rs2227564	1	T (protective)	-	0.91	-	L [CTG] > [CCG]	Increased T_H17 response[26]
CD6	rs11230563	3	T (protective)	0.92	0.92	0.93	R [CGG] > W [TGG]	Increased CD6 levels, increased T_H17 response[27,28]
SH2B3	rs3184504	1	T	-	1.06	1.05	W [TGG] > R [CGG]	Adaptive immunity, microbe sensing, increase protection against bacterial infection [29]
LACC1	rs3764147	1	G	1.11	1.16	-	I [ATC] > V [GTC]	Microbicidal mechanism, microbe sensing that is mediated via NOD2[30]

For the 97 SNVs, we investigated the dynamics of their reported IBD-associated allele in different ancient populations that differ regarding subsistence, time periods and location (hunter-gatherers, farmers, Steppe herders, Bronze Age, Iron Age and medieval populations). In addition to genotypes from our 172 Neolithic samples, we also considered publicly available information from 2273 other individuals (Supplementary Table 2). Using this large dataset, we established individual spatial-temporal trends for all 97 alleles (Supplementary Figure 1). Interestingly, the 7 missense variants could be divided into 3 groups based on similar frequency trends (1: *FCGR2A* rs1801274, *PLAU* rs2227564, *SH2B3* rs3184504, *LACC1* rs3764147; 2: *PTPN22* rs2476601, *IL23R* rs11209026 and 3: *CD6* rs11230563; Figure 1).

To assess whether all alleles followed common trend patterns, we explored their similarities through a hierarchical clustering analysis that grouped them into six major trajectories (Figure 2, Supplementary Figure 2). Trajectories 1, 2 and 3 were found to be the most frequent ones, as they comprised 78 of the 97 SNVs, including the 7 missense mutations (Figure 2, Table 1). Alleles in trajectory 1 had a low frequency in hunter-gatherers and Neolithic farmers compared to the other populations. Trajectory 2 was characterized by a sharp rise in allele frequency from hunter-gatherers to Neolithic individuals and then a decline to very low frequencies today. Trajectory 3 showed the mirror image (mirrored on the x-axis) of trajectory 2, with a very high frequency in hunter-gatherers and a steep decrease in the Neolithic (Figure 2). Trajectories 4, 5 and 6 (Supplementary Figure 2) are not further discussed here due to their small number of SNVs (total N = 19).

The SNVs in trajectories 1, 2 and 3 were selected for functional enrichment analysis. Based on the genomic positions of the SNVs, a gene set was assembled for each trajectory (Figure 3). Each trajectory contained 26 genes for the enrichment process. For trajectory 1, we observed 7 pathways that represented 3 functional groups, i.e., pertussis, cytokine-cytokine receptor-cytokine interaction and Fc gamma R-mediated phagocytosis (Figure 3A). For trajectory 2, 20 pathways in 9 functional groups were identified. The most representative pathways were mainly related to mechanisms of bacterial or other antigen recognition (e.g., shigellosis, allogeneic rejection, antigen processing and presentation) and T_H17 differentiation (Figure 3B). Trajectory 3 yielded 48 signalling pathways including, for instance, TGF-beta, Wnt, Toll-like receptor, C-type lectin receptor or B-cell receptor pathways (Figure 3C).

To investigate whether the allele frequency changes of the 7 missense variants that occurred over time were the result of selection or demographic events, we performed an admixture-informed analysis. We assessed whether the observed frequency in a population was in line with the admixture-informed expected frequency (based on the mixture proportions of the parental populations and their allele frequencies). Since aDNA data of the ancestors of the hunter-gatherers and early farmers in the Near East are not available, the admixture-informed analyses were limited to populations that derived from the two groups.

For the variants in trajectory 2 (*PTPN22* rs2476601, *IL23R* rs11209026) and 3 (*CD6* rs11230563), we found that only one observed frequency significantly differentiated from the expectation (*CD6* rs11230563 in LF, binomial test $p < 0.05$) but the others did not (binomial test $p > 0.05$; Supplementary Figure 3), indicating that these frequencies were likely the result of mixture between the parental populations. However, this was not the case for the variants in trajectory 1 (*FCGR2A* rs1801274, *PLAU* rs2227564, *SH2B3* rs3184504, *LACC1* rs3764147). Here we noted clear deviations between observed and expected frequencies for several populations (binomial test $p < 0.05$; Supplementary Figure 4), suggesting that admixture alone could not explain the allele frequency differences between populations. It is thus likely that selection, rather than admixture, played a larger role in shaping trajectory 1.

Discussion

IBD represents a modern disease. The first medical case reports date back to the late 18th century AD[31]. The genetic variation underlying IBD is assumed to be older, although it is still unclear how much. Some IBD-SNVs were shown to be already present in Neanderthals[32], a few others were detected in the genomes of prehistoric individuals[33], suggesting that the respective alleles have been conserved in the human gene pool for a long period of time. To extend our knowledge beyond these isolated observations, we have performed a systematic analysis of IBD-associated SNVs in ancient populations from Eurasia.

Given the important role that diet (and, through diet, the intestinal microbiome) plays in the manifestation of IBD, our starting point was the Neolithic period that was characterized by the transition from a hunter-gathering subsistence to farming (domestication of plants and animals), a sedentary lifestyle and a carbohydrate- and dairy-rich diet, among other things[34–36]. Co-housing with animals may have resulted in

exposure to animal microbiome and zoonotic diseases. The Neolithic began in the Near East about 12,000 years ago and was brought to Europe by migration about 4500 years later (~5500 BCE)[37]. This period marks the beginning of a long-term development that ultimately led to the genotype-environment mismatch that is responsible for IBD nowadays and that first manifested during industrialization. We used samples from the European Neolithic (5500 - 2900 BCE) to compare their genetic IBD variation with that of modern Europeans. Almost all of the independent 353 IBD-associated alleles analysed (99.4%) already existed in the Neolithic, and about 72.8% of them had a similar frequency as today, indicating that a large part of the genetic IBD component is thousands of years old.

However, 97 SNVs had significantly different frequencies compared to contemporary Europeans. A permutation test revealed that this large number was unlikely to be due to chance. We investigated these SNVs in more detail by establishing allele frequency trends over time including datasets from additional populations that differed in subsistence (hunting and gathering *versus* farming), period and geographic location. The allele frequency trends followed 6 main trajectories (Figure 2, Supplementary Figure 2). It appears that the variants within one trajectory, though not genetically linked, had a common history and changed their allele frequencies simultaneously, whether caused by demographic processes or selection.

Strikingly, the two mirror-inverted trajectories 2 and 3 alone accounted for 52 of the 97 SNVs, suggesting a functional commonality on which selection might have acted. The functions assigned to trajectory 2 (T_H17 pathway) and trajectory 3 (signalling, Figure 3) are relevant to inflammatory responses; these functions are also reflected in the 3 missense mutations. In trajectory 2, two missense mutations in the genes *IL23R* and *PTPN22* were detected (Table 1). Interestingly, the variant rs11209026 in *IL23R* is the variant with the greatest known protective effect on IBD and UC[38]. The IL23R protein, a receptor of the cytokine IL23, is a key regulator in the pathogenesis of chronic inflammatory disorders. Among the factors that trigger the endogenous IL23 production are microbial dysbiosis (caused by commensal bacteria or pathogens) in the intestine and dietary antigens[39]. Recently, monoclonal antibodies that block IL23R have been developed for several inflammatory diseases, where they have an enormous anti-inflammatory efficacy probably through downregulating the T_H17 immune response[40]. A similar effect has been described for *IL-23R* rs11209026[25]. The variant rs2476601 in *PTPN22* has been widely associated with a host of immune mediated diseases (e.g., rheumatoid arthritis, type-1 diabetes); however, in CD, this variant has a protective effect[5] that is partly mediated via an interaction between NLRP3 signalling and the intestinal microbiome[41,42]. In trajectory 3, the missense mutations rs11230563 in *CD6* and rs2227564 in *PLAU* were reported as risk factors for high CD6 levels (leading to an increased T_H17 response[43] and Alzheimer's disease[26]).

The missense variants (rs11209026 in *IL23R*, rs2476601 in *PTPN22*) in trajectory 2 are rare in Europeans today but were present at a high frequency of 21% and 25%, respectively, in the earliest group of Neolithic farmers in the Near East (Figure 1). The health status of this population was presumably poor due to harsh living conditions and a heavy workload[44,45]. The farmers subsisted on a diet dominated by

carbohydrates, dairy and brewery[44,46–49]. The increased intake of fermented products such as alcohol at the beginning of the Neolithic could have altered the gut microbiome[46,50] as well as the faecal-oral transmission of bacteria from domesticated animals[51]. Such a scenario is supported by studies of contemporary hunter-gatherer populations who often suffer from intestinal microbial dysbiosis and inflammation after adopting agricultural or western lifestyles (reviewed in[51]). Therefore, it is plausible that genetic variants that compensated for these deleterious effects by conveying an anti-inflammatory influence, such as rs11209026 in *IL23R* or rs2476601 in *PTPN22* (trajectory 2), increased in frequency from hunter-gatherers to earliest farmers, while pro-inflammatory variants, such as rs11230563 in *CD6,* decreased (trajectory 3). In small populations, directional selection can lead to large allele frequency shifts. Indeed, both the archaeological record and ancient genomic studies showed that farmers in the Near East were isolated and characterized by small population sizes[33,52,53]. With so many variants affected by the same frequency changes (58 SNVs in trajectory 2 and 3), genetic drift is unlikely to have been a major driver. Thus, it seems conceivable that selection and adaptation to the agricultural lifestyle were mainly responsible for the observed allele frequency differences between hunter-gatherers and farmers.

The later phases of the Neolithic were marked by unprecedented population growth which allowed the Near East farmers to migrate to Europe and across the continent (~5500 BCE)[54]. Interestingly, the three variants rs11209026, rs2476601 and rs11230563 persisted at high (the first two) or low frequencies (the last) for several thousand years during the Neolithic expansion and the early European Neolithic (Figure 1), suggesting balancing selection that may have been influenced by increased exposure to pathogens. For the post-Neolithic periods, our results showed no evidence of directional selection. Rather, the later frequency changes can be explained by the two major admixture events that significantly influenced the European gene-pool: the influx of hunter-gatherers into the farmers (~3800-3500 BCE) and the arrival of the Steppe populations (~2600 BCE).

We therefore hypothesize that the genetically dampened immune reactions against intestinal dysbiosis and protection against excessive inflammation observed in the farmers of the Near East might represent the ancient phenotype at the core of the three missense mutations in trajectory 2 and 3. Compared to these functionally well-studied variants, no similarly detailed information was available for the other 49 SNVs in the two trajectories. However, when we performed a manual literature search on their functions, we found that 17 of the annotated genes (or flanking genes in the case of SNVs in intergenic regions) in trajectory 2 (65%) and 7 in trajectory 3 (26%) were also associated with (dys-)regulation of the gut microbiome and/or cytokine signaling (Supplementary Table 3), supporting our hypothesis. Evidence for selection towards an anti-inflammatory phenotype is also provided by modern genetic studies which have identified many protective IBD alleles as targets of positive selection and, conversely, risk alleles as targets of negative selection[4,8]. However, a strong anti-inflammatory profile comes at the cost of increased susceptibility to infection. Noteworthy, trajectory 1 was associated with bacteria/antigen recognition, cytokine-mediated regulation and phagocytosis functions in our analyses (Figure 3A). It mainly differed from trajectories 2 and 3 by a steep rise in allele frequencies between the Iron Age and the

Middle Ages (Figure 2). The resulting high frequencies have been maintained until today. Four missense variants (rs3764147 in *LACC1*, rs1801274 in *FCGR2A*, rs3184504 in *SH2B3,* rs2227564 in *PLAU*) were detected in trajectory 1. All four genes are involved in microbe recognition mediated via NOD2, an important receptor for bacterial pathogens[29,55]. Interestingly, the risk allele rs3184504 in *SH2B3* was shown to increase protection against bacterial infection[29]. An analysis of data from modern-day Europeans suggested a selective sweep on *SH2B3* estimated to have occurred between 1200 and 1700 years ago[29]. This time frame fits remarkably well with the increase in allele frequency observed for rs3184504 (and the other alleles in trajectory 1). Additionally, historically reported epidemics date to the same period (e.g., the Justinianic plague, 541-549 AD). The frequency of the rs3184504 risk allele has remained unchanged until today, possibly due to recurrent outbreaks of bacterial infections caused by *Yersinia pestis*, *Salmonella enterica*, *Mycobacterium leprae* or *tuberculosis*, amongst others[56,57]. A literature search on the other 17 (65%) variants in trajectory 1 supported our hypothesis that the emergence of the ancient phenotype of trajectory 1 was related to selection pressure from bacterial infections (Supplementary Table 3).

Trajectories 2 and 3 are probably the result of the Neolithization process in the Near East 12,000 years ago and admixture in the subsequent millennia. In contrast, trajectory 1 was more influenced by events that took place only in the last 2,000 years. It appears that the current frequencies of the 97 IBD variants, and thus a large part of the complex genetic risk architecture of IBD in the contemporary European population, are the consequence of past adaptations and demographic processes. A better understanding of the evolutionary history of disease variants is an important first step in translating genetic findings into life-style interventions as an element of health maintenance. Palaeogenetics may support a precision medicine approach by facilitating trajectory-informed genetic stratification of patients and opening new perspectives for personalized disease prevention.

Materials And Methods

Material

Sample collection

We generated and analysed data from 172 individuals collected from seven archeological sites that all represent farming communities (Table 2). The sites are located in present-day Germany and cover the transition from the Early to the Late Neolithic.

Table 2. Origin and dating of the 172 individuals newly examined in this study

Site	Sample size	Dating (BCE)
Niedertiefenbach[19,58]	89	3300 – 3200
Altendorf[59]	15	3250 – 3100
Warburg[60, 58]	18	3400 –2900
Rimbeck[61]	3	3300 – 2900
Niederpöring[62]	7	5200 – 4900
Fellbach-Öffingen [63,64]	17	5700 – 4900
Trebur[65]	23	5000 – 4500
TOTAL	**172**	

Publicly available data

For comparative purposes, allele frequency information of individuals from the modern Non-Finish European (NFE) population was retrieved from the Genome Aggregation Database (gnomAD; https://gnomad.broadinstitute.org (release version V2.1.1). Additionally, sequencing data (FASTQ) or mapping data (BAM) of individuals listed in the Allen Ancient DNA Resource (AADR; https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data (release version V50.0) was downloaded. The selected 2273 individuals were from Western Eurasia (n=2141, 35.05 < lat < 69.65, -24.07 < long < 27.86), the Near East (n=48, 31.79 < lat < 41.9, 21.35 < long < 36.44) and the Russian Steppe (n=84, 42.18 < lat < 55.99, 39.91 < long < 92.85). Sequencing data was mapped and processed as described below.

Methods

Ancient DNA extraction, capture and sequencing

Bone or tooth samples were cleaned with bleach and drilled or powdered to obtain material for DNA extraction. Extraction was performed following established protocols published elsewhere[66] using partial Uracil-DNA Glycosylase (UDG) libraries. Genomic loci were captured using the myBaits hybridization kit from Arbor Biosciences. Our targets were SNVs associated with IBD, UC or CD in Europeans based on the GWAS catalog[67] (accessed on 04.01.2019). For the candidate loci, a total of 807 unique probes were designed with an average length of ~60 bp. Probes were to specifically capture the target regions while avoiding homologous or repetitive elements. However, only 610 probes passed the quality control thresholds specified by the manufacturer. The enriched libraries were sequenced using the Illumina NovaSeq 6000 platform (2x100). In addition to the genomic capture, whole-genome shotgun sequencing was performed to assess the ancestry of the samples.

Processing and mapping of sequencing reads

Sequencing data was preprocessed as described in Immel et al. 2021[19] and subsequently mapped to the human genome build hg19 (International Human Genome Sequencing Consortium, 2001) using BWA[51] with the reduced mapping stringency parameter "-n 0.01". Authenticity of aDNA was verified as described in[19]. To reduce genotyping errors arising from the ends of the reads, we estimated deamination patterns with DamageProfiler v1.1[52] and trimmed the ends of the reads with bamUtils v1.0.15[68] to achieve a terminal damage rate below 5%. In addition, the mapped data was filtered by quality, keeping reads with mapping quality and base alignment quality greater than 20.

Genotyping and quality control

Genotypes were estimated from BAM files using Bcftools v1.12[69]. Reads from whole-genome shotgun sequencing that aligned to the target loci were included in the dataset to be genotyped. Since the loci examined were enriched and had low similarity with other regions of the genome, we decided to call diploid genotypes rather than pseudo-haploid ones in order to obtain more accurate estimates of the allele frequencies. Genotype calls that did not meet the following criteria were removed: QUAL $\geq$ 20; GQ > 20; DP > 3; “not indel”. To reduce duplicate signals, we applied pairwise variant pruning by running Plink 1.9 using -indep-pairwise. We considered a window of 50 variants, a shift window step and a pairwise genotypic correlation threshold of 5 and 0.2, respectively. To ensure that the SNVs to be analysed were not the product of genotyping errors, purifying selection, inbreeding or population substructure in early or late farmers, we excluded SNVs that deviated from Hardy-Weinberg equilibrium (HWE; $p < 0.0001$) in any of the populations.

Allele frequency analysis

For allele frequency calculations, we considered 10 groups according to their respective archaeological context and dating: hunter-gatherers (published), early farmers from the Near East (published), early farmers from Europe (published and our own EF), late farmers (published and our own LF), Steppe herders (published), Bronze Age/Iron Age (published), medieval populations (published) and NFE. Based on the minor alleles of the target SNVs (reported for the modern European population), we calculated allele frequencies in each of the abovementioned populations. Group frequencies were used to identify SNVs with significant minor allele frequency (MAF) changes between early or late farmers and the modern NFE. A two-sided Fisher's exact test with mid p-value adjustment using the R package exact2x2[70] was applied to test for frequency differences. The p-value of the binomial tests was adjusted for multiple testing using the false discovery rate method. A permutation test was used to assess whether the number

of significant SNVs was different from that expected by chance. For this, a two-sided Fisher's exact test (as described above) was applied using 685,807 SNVs that are commonly applied to assess population ancestry and that were in the MAF range of the 385 SNVs studied.

Including the frequencies of all the 10 populations studied allowed us to outline the dynamics of the allele frequency of each SNV over time and across populations. We refer to these dynamics as trends. Thus, having an "m x n" matrix, with number of rows "m" equal to the number of SNVs (97), "n" columns equal to the number of populations (10) and where each element s_{ij} denotes the value of the observed frequency for the minor allele of SNV_i in $population_j$; the trends represent each of the rows of this matrix. We then cluster the trends according to their similarity in trajectories using a custom R script. The script first transforms the trends (scale and centers on the mean) and then computes a pairwise Euclidean distance between them. Euclidean distances are used to perform agglomerative hierarchical clustering using average linkage. The optimal number of clusters was determined using the gap statistics of the R package factoextra[71].

Functional context analysis

We assigned genes to SNVs based on their genomic location (gene closest to SNV); we then investigated KEGG pathways for enrichment of those genes by functional enrichment analysis employing the R package pathfindR[72].

Ancestry-informed admixture analysis

To determine whether the allele frequencies in two parental populations A and B could explain allele frequencies in admixed offspring populations, a two-sided binomial test was performed. The admixture proportions (k_A, k_B; $k_A+k_B=1$) and allele frequencies (f_A, f_B) of the parental populations were used to calculate the expected allele frequency of a target population (f_{exp}).

$$f_{exp} = (k_A \times f_A) + (k_B \times f_B)$$

Then a binomial test was applied to test for a significant deviation of the observed frequency in the offspring population from f_{exp}.

Declarations

Ethics approval and consent to participate

This study was carried out following the principles for ethical DNA research on human remains as described in[73].

Availability of data and materials

BAM files of aligned reads can be found at the European Nucleotide Archive (project accession no. PRJEB55520). All other data needed to evaluate the findings in the paper are present in the main text or the supplementary materials.

Competing interests

The authors declare no competing interests.

Funding

This study was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2167 390884018 and EXC 2150 390870439 – and CRC 1266 project ID 290391021.

Author contributions

B.K.-K., S.Sch. and A.N. developed the idea for this study. S. Sch.-L., J.W., C.B., D.K., M.F., I.G., K.Sch., M.R., J.P. and G.G. assembled archaeological material. B.K.-K. and N.d.S. generated ancient DNA data. G.G.T., N.d.S., D.K., A.C., and B.K.-K. analysed the ancient DNA data. B.K.-K., G.G.T., N.d.S., D.K., J.D., S. Sch.-L., J.W., C.B., M.F., I.G., K.Sch., J.P., G.G., A.C., J.M. S.Sch., A.N. interpreted the findings. A.N., B.K.-K., S.Sch. wrote the manuscript with contributions from G.G.T., N.d.S., D.K., J.D. as well as input from all other authors.

References

1. Ellinghaus, D. et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* **48**, 510-518 (2016).

2. Caruso, R., Lo, B. C. & Núñez, G. Host-microbiota interactions in inflammatory bowel disease. *Nat Rev Immunol* **20**, 411-426 (2020).

3. Hampe, J. et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* **39**, 207-211 (2007).

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

5. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* **491**, 119-124 (2012).

6. Nakagome, S. et al. Crohn's disease risk alleles on the NOD2 locus have been maintained by natural selection on standing variation. *Mol Biol Evol* **29**, 1569-1585 (2012).

7. Cagliani, R. et al. Crohn's disease loci are common targets of protozoa-driven selection. *Mol Biol Evol* **30**, 1077-1087 (2013).

8. Raj, T. et al. Common risk alleles for inflammatory diseases are targets of recent positive selection. *Am J Hum Genet* **92**, 517-529 (2013).

9. Mesbah-Uddin, M., Elango, R., Banaganapalli, B., Shaik, N. A. & Al-Abbasi, F. A. In-silico analysis of inflammatory bowel disease (IBD) GWAS loci to novel connections. *PLoS One* **10**, e0119420 (2015).

10. Karlsson, E. K., Kwiatkowski, D. P. & Sabeti, P. C. Natural selection and infectious disease in human populations. *Nat Rev Genet* **15**, 379-393 (2014).

11. Quintana-Murci, L. Understanding rare and common diseases in the context of human evolution. *Genome Biol* **17**, 225 (2016).

12. Domínguez-Andrés, J. et al. Evolution of cytokine production capacity in ancient and modern European populations. *Elife* **10**, e64971 (2021).

13. Marciniak, S. & Perry, G. H. Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet* **18**, 659-674 (2017).

14. Cox, S. L., Ruff, C. B., Maier, R. M. & Mathieson, I. Genetic contributions to variation in human stature in prehistoric Europe. *Proc Natl Acad Sci U S A* **116**, 21484-21492 (2019).

15. Marnetto, D. et al. Ancestral genomic contributions to complex traits in contemporary Europeans. *Curr Biol* **32**, 1412-1419.e3 (2022).

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

17. Richards, M. P., Schulting, R. J. & Hedges, R. E. Archaeology: sharp shift in diet at onset of Neolithic. *Nature* **425**, 366 (2003).

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

19. Immel, A. et al. Genome-wide study of a Neolithic Wartberg grave community reveals distinct HLA variation and hunter-gatherer ancestry. *Commun Biol* **4**, 113 (2021).

20. Hedjoudje, A. et al. rs2476601 polymorphism in PTPN22 is associated with Crohn's disease but not with ulcerative colitis: a meta-analysis of 16,838 cases and 13,356 controls. *Ann Gastroenterol* **30**, 197-208 (2017).

21. Rabaneda-Bueno, R. et al. PTPN22 gene functional polymorphism (rs2476601) in older adults with frailty syndrome. *Mol Biol Rep* **48**, 1193-1204 (2021).

22. Armitage, L. H., Wallet, M. A. & Mathews, C. E. Influence of PTPN22 Allotypes on Innate and Adaptive Immune Function in Health and Disease. *Front Immunol* **12**, 636618 (2021).

23. Zhang, C., Wang, W., Zhang, H., Wei, L. & Guo, S. Association of FCGR2A rs1801274 polymorphism with susceptibility to autoimmune diseases: A meta-analysis. *Oncotarget* **7**, 39436-39443 (2016).

24. López-Martínez, R. et al. The FCGR2A rs1801274 polymorphism was associated with the risk of death among COVID-19 patients. *Clin Immunol* **236**, 108954 (2022).

25. Abdollahi, E., Tavasolian, F., Momtazi-Borojeni, A. A., Samadi, M. & Rafatpanah, H. Protective role of R381Q (rs11209026) polymorphism in IL-23R gene in immune-mediated diseases: A comprehensive review. *J Immunotoxicol* **13**, 286-300 (2016).

26. Riemenschneider, M. et al. A functional polymorphism within plasminogen activator urokinase (PLAU) is associated with Alzheimer's disease. *Hum Mol Genet* **15**, 2446-2456 (2006).

27. Swaminathan, B. et al. Fine mapping and functional analysis of the multiple sclerosis risk gene CD6. *PLoS One* **8**, e62376 (2013).

28. Ma, C. et al. Critical Role of CD6highCD4+ T Cells in Driving Th1/Th17 Cell Immune Responses and Mucosal Inflammation in IBD. *J Crohns Colitis* **13**, 510-524 (2019).

29. Zhernakova, A. et al. Evolutionary and functional analysis of celiac risk loci reveals SH2B3 as a protective factor against bacterial infection. *Am J Hum Genet* **86**, 970-977 (2010).

30. Lahiri, A., Hedl, M., Yan, J. & Abraham, C. Human LACC1 increases innate receptor-induced responses and a LACC1 disease-risk variant modulates these outcomes. *Nat Commun* **8**, 15614 (2017).

31. Kaplan, G. G. & Windsor, J. W. The four epidemiological stages in the global evolution of inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* **18**, 56-66 (2021).

32. Sankararaman, S. et al. The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* **507**, 354-357 (2014).

33. Marchi, N. et al. The genomic origins of the world's first farmers. *Cell* **185**, 1842-1859.e18 (2022).

34. Münster, A. et al. 4000 years of human dietary evolution in central Germany, from the first farmers to the first elites. *PLoS One* **13**, e0194862 (2018).

35. Schulting, R. Dietary shifts at the Mesolithic-Neolithic transition in Europe: An overview of the stable isotope data in *The oxford handbook of the archaeology of diet* (eds Lee-Thorp, J. A. & Katzenberg, M. A., Oxford Academic, 2018).

36. Simmons, A. H. *The Neolithic revolution in the Near East: transforming the human landscape* (University of Arizona Press, 2011).

37. Bellwood, P. *First farmers: The origins of agricultural societies* (Wiley-Blackwell, 2004).

38. Ferguson, L. R. et al. IL23R and IL12B SNPs and haplotypes strongly associate with Crohn's Disease risk in a New Zealand population. *Gastroenterol Res Pract* **2010**, 539461 (2010).

39. Gerosa, F. et al. Differential regulation of interleukin 12 and interleukin 23 production in human dendritic cells. *J Exp Med* **205**, 1447-1461 (2008).

40. Reay, W. R. & Cairns, M. J. Advancing the use of genome-wide association studies for drug repurposing. *Nat Rev Genet* **22**, 658-671 (2021).

41. Spalinger, M. R. et al. PTPN22 regulates NLRP3-mediated IL1B secretion in an autophagy-dependent manner. *Autophagy* **13**, 1590-1601 (2017).

42. Larabi, A., Barnich, N. & Nguyen, H. T. T. New insights into the interplay between autophagy, gut microbiota and inflammatory responses in IBD. *Autophagy* **16**, 38-51 (2020).

43. Consuegra-Fernández, M. et al. Genetic and experimental evidence for the involvement of the CD6 lymphocyte receptor in psoriasis. *Cell Mol Immunol* **15**, 898-906 (2018).

44. Larsen, C. S. et al. Bioarchaeology of Neolithic Çatalhöyük reveals fundamental transitions in health, mobility, and lifestyle in early farmers. *Proc Natl Acad Sci U S A* **116**, 12615-12623 (2019).

45. Latham, K. J. Human health and the Neolithic revolution: An overview of impacts of the agricultural transition on oral health, epidemiology, and the human body. *Nebraska Anthropologist* **28**, 95-102 (2013).

46. Dietrich, L. et al. Investigating the function of Pre-Pottery Neolithic stone troughs from Göbekli Tepe–An integrated approach. *J Archaeol Sci Rep* **34**, 102618 (2020).

47. Evershed, R. P. et al. Dairying, diseases and the evolution of lactase persistence in Europe. *Nature* **608**, 336-345 (2022).

48. Rosen, A. M. & Rivera-Collazo, I. Climate change, adaptive cycles, and the persistence of foraging economies during the late Pleistocene/Holocene transition in the Levant. *Proc Natl Acad Sci U S A* **109**, 3640-3645 (2012).

49. Sebald, S. V., Papathanasiou, A. & Grupe, G. Changing subsistence economies in the course of the Neolithic transition: Isotopic sourcing of collagen isotopic ratios in human skeletons from early Neolithic Anatolia and Greece. *J Archaeol Sci Rep* **43**, 103450 (2022).

50. Martino, C. et al. Acetate reprograms gut microbiota during alcohol consumption. *Nat Commun* **13**, 4630 (2022).

51. Crittenden, A. N. & Schnorr, S. L. Current views on hunter-gatherer nutrition and the evolution of the human diet. *Am J Phys Anthropol* **162 Suppl 63**, 84-109 (2017).

52. Kılınç, G. M. et al. The demographic development of the first farmers in Anatolia. *Curr Biol* **26**, 2659-2666 (2016).

53. Kılınç, G. M. et al. Archaeogenomic analysis of the first steps of Neolithization in Anatolia and the Aegean. *Proc Biol Sci* **284**, 20172064 (2017).

54. Furholt, M. Mobility and social change: Understanding the european neolithic period after the archaeogenetic revolution. *J Archaeol Res* **29**, 481-535 (2021).

55. Graham, D. B. & Xavier, R. J. Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* **578**, 527-539 (2020).

56. Susat, J. et al. Yersinia pestis strains from Latvia show depletion of the pla virulence gene at the end of the second plague pandemic. *Sci Rep* **10**, 14628 (2020).

57. Haller, M. et al. Mass burial genomics reveals outbreak of enteric paratyphoid fever in the Late Medieval trade city Lübeck. *iScience* **24**, 102419 (2021).

58. Meadows, J. et al. High-precision Bayesian chronological modeling on a calibration plateau: the Niedertiefenbach gallery grave. *Radiocarbon* **62**, 1261-1284 (2020).

59. Rinne, C., Drummer, C. & Hamann, C. Collective and individual burial practices. Changing patterns at the beginning of the third millennium BC: The megalithic grave of Altendorf. *Journal of Neolithic Archaeology* **21**, 75-88 (2019).

60. Raetsel-Fabian, D. Der nordwestliche Nachbar: Neue Aspekte zur Wartbergkultur. *Kolloquien Des Instituts Für Ur Und Frühgeschichte Erlangen* **1**, 26-28 (1999).

61. Raetzel-Fabian, D. Absolute Chronologie. *Die Kollektivgräber-Nekropole Warburg IV. Bodenaltertümer Westfalens* **34**, 165-178 (1997).

62. Pechtl, J., Hanöffner, S., Staskiewicz, A. & Obermaier, H. Die linienbandkeramische Gräbergruppe von Niederpöring-"Leitensiedlung", Gde. Oberpöring, Lkr. Deggendorf in *Vorträge des 36. Niederbayerischen Archäologentages* Leidorf, Marie, 2018).

63. Biel, J. Ein bandkeramischer Friedhof in Fellbach-Oeffingen, Rems-Murr-Kreis in *Excavaciones arqueológicas en Baden-Württemberg* (ed Theiss, K., Theiss, Konrad, 1987).

64. Trautmann, I. & Wahl, J. Leichenbrände aus linearbandkeramischen Gräberfeldern Südwestdeutschlands-Zum Bestattungsbrauch in Schwetzingen und Fellbach-Oeffingen. *Fundberichte aus Baden-Württemberg* **28**, 7-18 (2005).

65. Müller, J. Zur Belegungsabfolge des Gräberfeldes von Trebur: Argumente der typologieunabhängigen Datierungen. *Prachistorische Zeitschrifte* **77**, 148-158 (2002).

66. Krause-Kyora, B. et al. Ancient DNA study reveals HLA susceptibility locus for leprosy in medieval Europeans. *Nat Commun* **9**, 1569 (2018).

67. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**, D1005-D1012 (2019).

68. Jun, G., Wing, M. K., Abecasis, G. R. & Kang, H. M. An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data. *Genome Res* **25**, 918-925 (2015).

69. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993 (2011).

70. Fay, M. P. Confidence intervals that match Fisher's exact or Blaker's exact tests. *Biostatistics* **11**, 373-374 (2010).

71. Kassambara, A. & Mundt, F. Factoextra: extract and visualize the results of multivariate data analyses. *R package version* (2020).

72. Ulgen, E., Ozisik, O. & Sezerman, O. U. pathfindR: An R Package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front Genet* **10**, 858 (2019).

73. Alpaslan-Roodenberg, S. et al. Ethics of DNA research on human remains: five globally applicable guidelines. *Nature* **599**, 41-46 (2021).

Figures

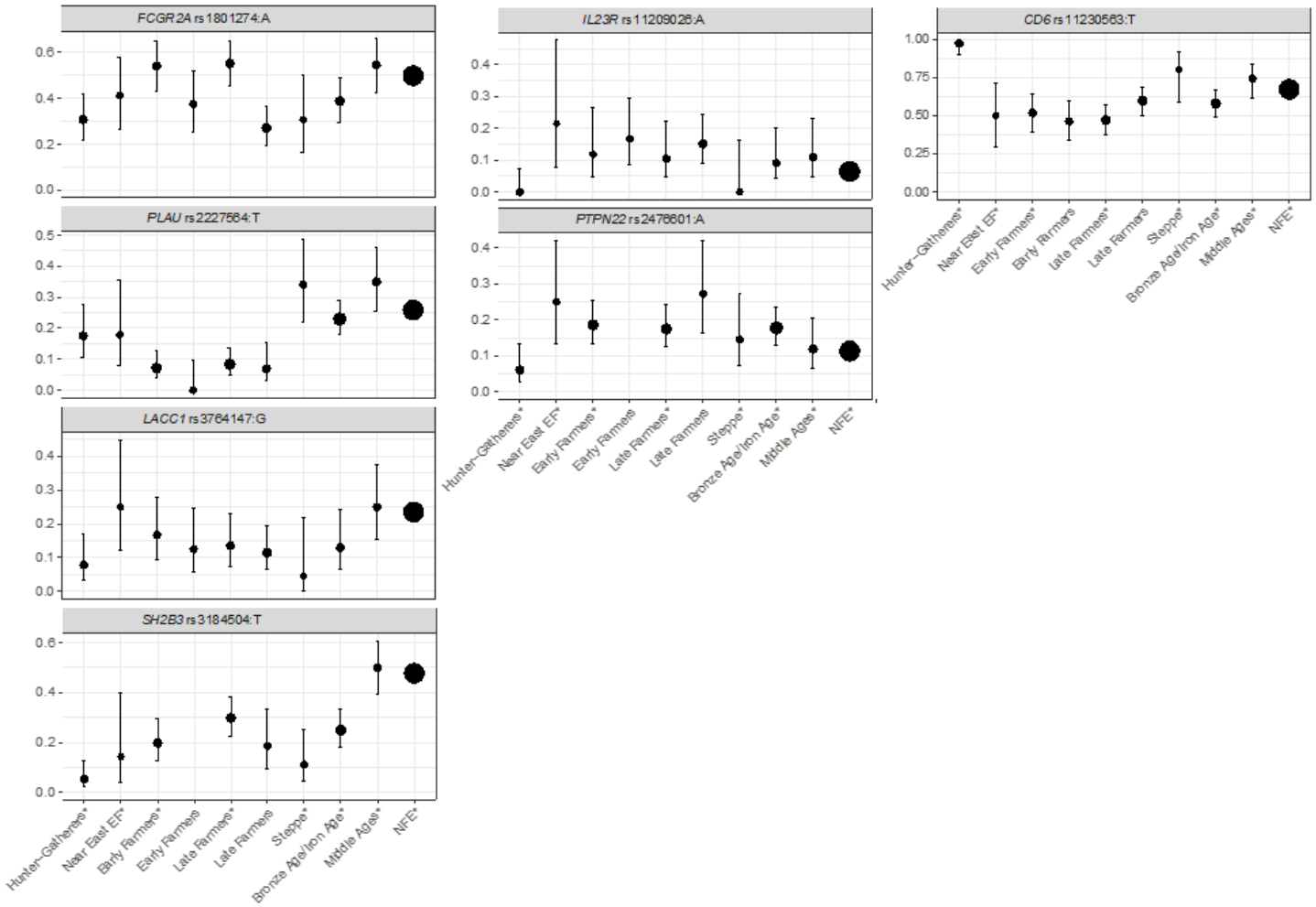

Figure 1

Frequency trends of the 7 missense variants.

The blue dashed line runs through the mean value of the allele frequency for each of the populations evaluated, thus forming the trend. Populations marked with an asterisk (*) contain samples from published sources. The size of the circles represents the relative sample size. The size of the NFE circles was divided by 100 for visualization.

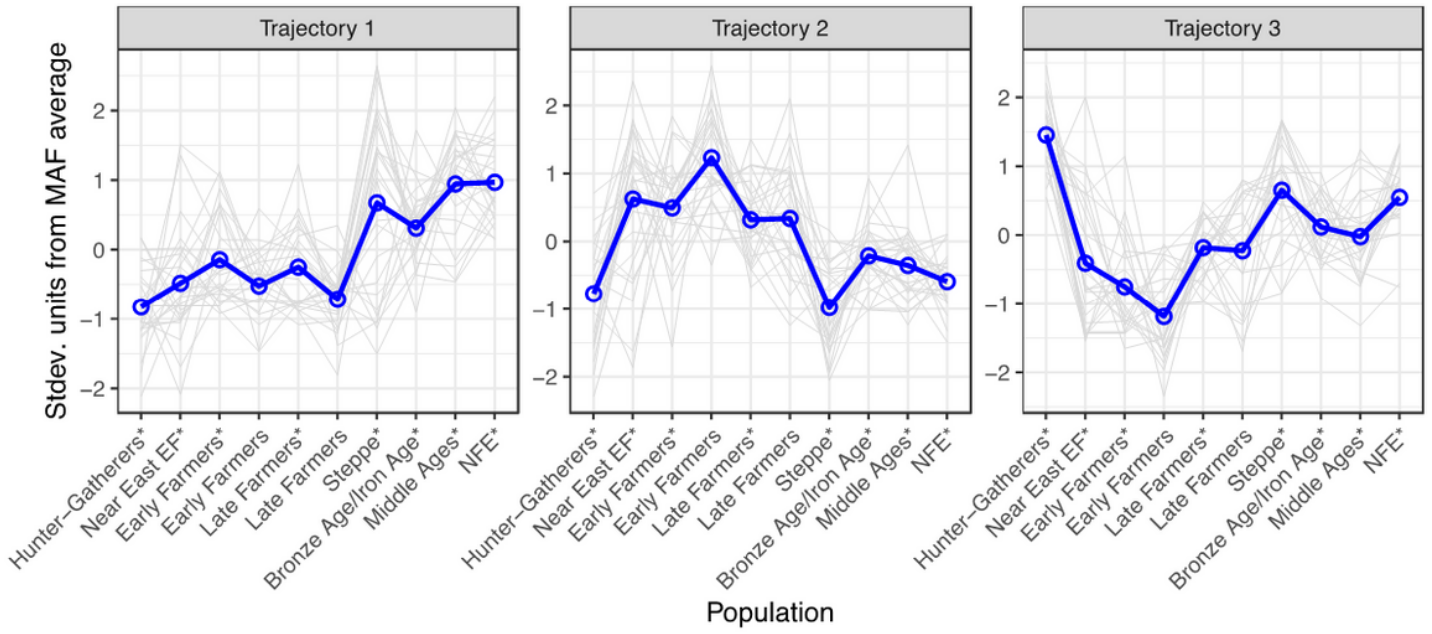

Figure 2

Three frequency trajectories which explain 78 out of 97 variants

Each trend is represented as a light gray line and is scaled and centered on the mean of the trend set in its trajectory. The blue line runs through the mean of the trend sets for each of the populations shaping the overall trajectory. On the y-axis, each trend is scaled and centered on its mean, thus the axis denotes the standard deviation. Asterisk (*) in the x-axis labels refers to published data including the modern Non-Finnish European (NFE) population. The number of markers in each of the 3 trajectories is 26.

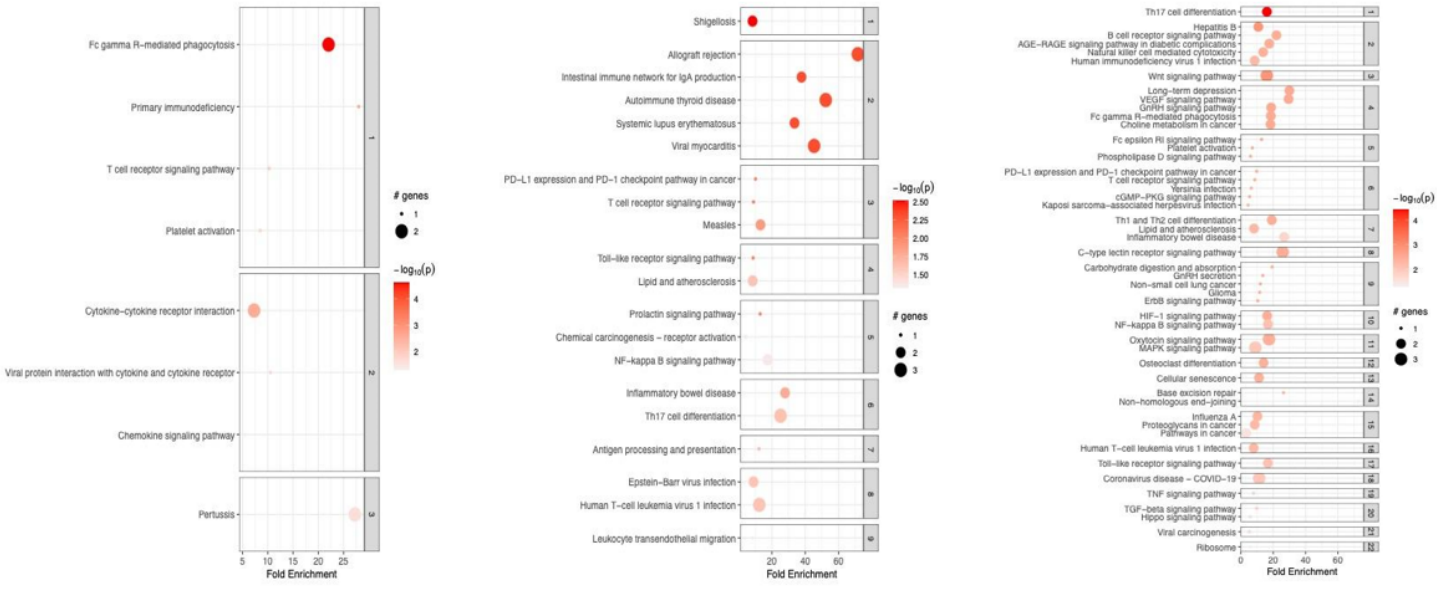

Figure 3

Bubble chart of the functional enrichment for trajectories 1 (A), 2 (B) and 3 (C) Enrichment results were grouped by clusters (labeled on the right-hand side of each panel). The x axis corresponds to fold enrichment values, while the y axis indicates the enriched pathways. The size of the bubble indicates the number of genes in the given pathway. Color indicates the −log10 (lowest-p) value; the more it shifts to red, the more significantly enriched the pathway is.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigurelegends170922.docx
- SupplementaryTable1.xlsx
- SupplementaryTable2.xlsx
- SupplementaryTable3.xlsx
- SupplementaryFigure1.pdf
- SupplementaryFigure2.pdf
- SupplementrayFigure3.pdf
- SupplementaryFigure4.pdf