Positive Selection Affects the Expression of Systemic Lupus Erythematosus Associated Loci in Human Populations

Victoria Oberreiter

University of Vienna

Tobias Goellner

University of Vienna

David L. Morris

King's College London

Helmut Schaschl (✉ helmut.schaschl@univie.ac.at)

University of Vienna

Research Article

Keywords: human genetic adaption, positive selection, systemic lupus erythematosus, immunity, eQTLs

Posted Date: March 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-255082/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Systemic lupus erythematosus (SLE) shows marked population-specific disparities in disease prevalence, including substantial variation in manifestations and complications according to genetic ancestry. Several recent studies suggest that a substantial proportion of variation of gene expression shows genetic ancestry-associated differences in gene regulation on immune responses. Positive selection may act in a population-specific manner on expression quantitative trait loci (eQTLs) and thereby contributes to the difference in the differences of SLE prevalence and manifestation in human populations. We tested the hypotheses that some of the identified SLE risk polymorphisms display pleiotropic effects or polygenicity driven by positive selection. We performed a genome-wide scan for recent positive selection by using integrated Haplotype Score (iHS) statistics in different human populations. In addition, we estimated the timing of beneficial mutations to understand what possible selective pressures drive positive selection at SLE-associated loci.

Results: We identified several SLE risk loci that are population-specifically under positive selection. Almost all SNPs that are under positive selection function as *cis*-eQTLs in different tissue types. We determined that adaptive eQTLs affect the expression of fewer genes than non-adaptive eQTLs, suggesting a limited range of effect of an eQTL at SLE risk sites that show signatures of positive selection. Furthermore, some positively selected SNPs are located in transcription factor binding sequences. The timing of positive selection for the studied loci suggests that both environmental and recent lifestyle changes during as well as after the Neolithic Transition may have become selectively effective. We propose a novel link between positively selected eQTLs at a certain SLE risk locus in Europeans and a physiological pathway not previously considered in SLE.

Conclusions: We conclude that population-specific adaptive eQTLs contribute to the observed variation in specific manifestations and complications of SLE in different ethnicities. Our results suggest also that human populations adapt more rapidly to environmental and lifestyle stimuli via modification of gene expression without having to alter the genetic code.

Background

Over the last few years, numerous genome-wide association studies (GWAS) have been performed in an attempt to uncover the genetic basis of human autoimmune diseases such as SLE. This inflammatory autoimmune disorder is characterised by an impaired clearance of apoptotic cells, production of auto-antibodies against nuclear antigens, deposition of immune complexes and involvement of multiple organs, leading to considerable variation in clinical manifestations. SLE can lead to increased morbidity and early mortality from end-organ damage; more than half of all SLE patients will develop lupus nephritis, which is one of the severe complications of SLE [1]. Sex-specific hormones and/or certain environmental factors are thought to trigger specific genetic and epigenetic mechanisms that lead to

failure of self-tolerance with increased inflammatory responses. These responses include both the innate and adaptive immune systems, ultimately leading to damage or destruction of multiple organs and tissues such as the skin, joints, kidneys, heart and brain [2]. A striking feature of SLE is that it affects mainly women of reproductive age (c. 90% of cases). A recent study reported that SLE ranks among the top 20 leading causes of death in females between 5 and 64 years of age [3]. This striking female predominance has been explained by the modulating effect estrogen has on the activation of the immune system (or alternatively, in males, the immunosuppressive effects of testosterone leads to reduced susceptibility for SLE in males). Moreover, SLE shows substantial differences in prevalence – it is more common in African-Americans and Asians than in Europeans [4, 5]. People with African genetic ancestry who have migrated to North America or Europe are at a higher risk of developing SLE compared to those of North American or European origin [6, 7]. The exact underlying mechanisms for the population-specific disparities are currently under revision and remain elusive. Evidence from family and linkage analyses indicates that a strong genetic contribution to SLE, with an approximately ten-fold higher risk in monozygotic versus dizygotic twins [8]. The concordance rate in monozygotic twins, however, appears to vary substantially among studied cohorts, ranging from 25–57% [8–10]. Nonetheless, the genetic heritability of SLE is probably substantial, with reports from 44% [11] to 66% [12]. The strongest genetic association signal among the common variants from GWAS was obtained from the human leukocyte antigen (HLA) region and from genes involved in type I interferon (IFN) signalling and production; many other SLE susceptibility loci are located in gene regulatory acting sequences within or near genes with functional relevance to the immune system [13, 14]. Importantly, many risk-variants are common, have small effect sizes and are often in near-perfect linkage disequilibrium (LD) with surrounding SNPs: this makes it challenging to unambiguously identify the causal genetic variants through traditional fine-mapping methods. In addition, the frequencies of risk loci can differ among human populations, resulting in different haplotypes and LD properties. The consensus is that genes close to the top associated risk SNP identified in GWAS are the most likely causal genes. Thus, however, does not apply to all the cases, as shown in recent studies [15]. For example, the SLE risk SNP rs4917014, which is a regulatory variant located close to the transcription factor encoding gene *IKZF1*, acts as a *trans*-eQTL altering *C1QB* and five type 1 interferon response genes which are both hallmarks of SLE [15]. For GWAS these observations have proven to be highly challenging in trying to move from uncovering risk variants through GWAS to their biological interpretation [16, 17].

From a population genetics viewpoint, the question arises whether any of the identified risk loci are adaptive to the environment, i.e., do changing conditions contribute to the risk of developing autoimmunity. Over the last 100,000 years, behaviourally modern humans have been subjected to a wide range of environmental changes, including the Out-of-Africa migration. The transition of most human populations from a hunter-gatherer form of subsistence to a sedentary, agro-pastoral lifestyle within the last 9,000–13,000 years ('Neolithic Transition') included substantial changes of environments and lifestyles related to the domestication of plants and animals. The transformations during the Neolithic demographic transition resulted, for example, in new dietary habits, increased demographic growth and elevated exposure to new pathogens (zoonosis) due to close contact with domesticated animals. The

subsequent transition to modernity led to additional fundamental lifestyle and ecological changes such as lower fertility, excess of caloric availability, longer life and an ageing population, exposure to novel man-made pollutants and a decline of global biodiversity. According to the mismatch hypothesis, natural selection could not catch up with the speed of rapidly changing environments; this resulted in genomes maintaining alleles that were beneficial to survival in prior environments but may harm fitness in contemporary habitats. Alleles that predispose post-transition populations to autoimmune-related diseases must have had substantial fitness benefits in prior environments. Due to strong selective forces in the past and despite their costs, these risk loci are being maintained in the genome. Ramos et al. [18] support the idea of recent positive selection as one important mechanism causing elevated population-specific frequencies of SLE risk loci. Moreover, Raj et al. [19] demonstrated that loci associated with several inflammatory-diseases, including several loci which show evidence of *cis*-regulatory effects on genes within the associated locus, are enriched for genomic signatures of recent positive selection. These two studies support the hypothesis that recent positive selection is a key mechanism causing an elevated population-specific frequency of certain autoimmune risk loci.

Although most risk loci are consistently present among populations with different ancestries, some loci are exclusive to specific populations. For example, risk alleles that are mapped to the genes *PTPN22* and *SH2B3-ATXN2* were identified as SLE risk variants only in Europeans and are apparently absent in Chinese populations [20]. Ramos et al. [18], who used integrated Haplotype Score (iHS) and F_{ST} analysis, found evidence of positive selection at the *PTPN22* locus in Africans. This suggests that the frequency of alleles at this locus differs among populations due to the effects of natural selection on allele frequency. Furthermore, many of the alleles associated with SLE function as expression quantitative trait loci (eQTLs), some of which have pleiotropic effects. Several recent studies suggest that a substantial proportion of variation of (immune-) gene expression shows genetic ancestry-associated differences in gene regulation on the immune response [21, 22]. For African and European individuals, the study by Nedelec et al. [21] reported that *cis*-eQTLs of immune genes were significantly enriched for high iHS values, indicating the importance of regulatory genetic variation in recent human evolution. Accordingly, positive selection may act in a population-specific manner on eQTLs and thereby contributes to the difference in the population-specific prevalence of different diseases. We therefore hypothesise that some of the identified SLE risk polymorphisms display pleiotropic effects or polygenicity driven by positive selection. To investigate this idea, we conducted tests for potential eQTLs at genetic polymorphisms associated with SLE. We performed a genome-wide scan for recent positive selection by using iHS in populations with African, European, South Asian and East Asian genetic ancestry and investigated which tissues are mainly affected by positively selected eQTLs. Furthermore, we estimated the timing of selection in order to understand when a beneficial mutation arose and spread in human populations.

Results

SLE risk loci under positive selection

Among the analysed 134 SLE risk loci (not included were the loci of the human leucocyte antigen region, HLA, which is known to be under some form of balancing selection), we identified 20 loci under positive selection (Table 1). The F_{ST} analysis revealed high ($F_{ST} > 0.25$) genetic differentiations for most locus-specific F_{ST} -values across populations. The allele frequencies and population pairwise F_{ST} -values for SNPs under positive selection are given in Additional file 1 and 2, respectively. The following risk loci are population-specifically under positive selection: *TNIP1*, *CD44*, *CLEC16A* and the region *SH2B-ATXN2* is under positive selection only in Europeans; in East Asians the loci *IL12A*, *BLK* and *SYNGR1*; in Africans the loci *BANK1*, *JAK2*, *RAD51B* and *UBE2L3*. The loci *IL12RB2*, *STAT4*, *ABHD6-PXK*, *LPP-TPRG1* and *UHRF1BP1* are under positive selection in Europeans and South Asians; the loci *XKR6-C8orf12-FAM167A* and *DRAM1* in Europeans, South Asians and East Asians. The genomic risk region at 7q11.23 shows signatures of positive selection across the *GTF2IRD1-GTF2I-GTF2IRD* and *HIP1* region in all analysed populations.

Table 1

SLE risk loci under positive selection in different human populations (AFR: Africa, EUR: Europe, SAS: South Asia, EAS: East Asia) presented by the average |iHS| values across SNPs under positive selection at the respective locus.

SLE risk region	SLE risk locus	SNP-allele iHS	Population
1p31.3	<i>IL12RB2</i>	rs4655706-G: 2.5	EUR, SAS
2q32.3	<i>STAT4</i>	rs10176621: 2.7	EUR, SAS
3p14.3	<i>ABHD6-PXK</i>	rs4681674-G: 2.8	EUR
		rs4681834-T: 2.8	SAS
3q25.33	<i>IL12A</i>	rs583911-G: 2.1	EAS
3q28	<i>LPP-TPRG1</i>	rs4479588-G: 2.3	EUR
		rs9682946-A: 2.8	SAS
4q24	<i>BANK1</i>	rs9307257-G: 3.5	AFR
5q33.1	<i>TNIP1</i>	rs7708392-G: 2.1	EUR
6p21.31	<i>UHRF1BP1</i>	rs11755393-A: 2.0	EUR, SAS
7q11.23	<i>GTF2IRD1-GTF2I-GTF2IRD2</i>	rs2428793-C: 2.7	AFR
		rs6460087-A: 3.2	SAS, EAS
	<i>HIP1</i>	rs1015877-C: 2.7	EUR, SAS, EAS
8p23.1	<i>XKR6-C8orf12-FAM167A</i>	rs10086521-T: 2.9	EUR, SAS, EAS
	<i>BLK</i>	rs4841563-T: 2.7	EAS
9p24	<i>JAK2</i>	rs75346988-G: 3.3	AFR
11p13	<i>CD44</i>	rs2732552-T: 2.3	EUR
12q23.2	<i>DRAM1</i>	rs6539014-A: 2.9	EUR, SAS, EAS
12q24.12	<i>SH2B-ATXN2</i>	rs10774625-A: 3.2	EUR
14q24.1	<i>RAD51B</i>	rs1314919-T: 2.6	AFR
16p13	<i>CLEC16A-CIITA-SOCS1</i>	rs7197422 -C: 2.3	EUR
22q11.21	<i>UBE2L3</i>	rs140490-G: 3.2	AFR
22q13.1	<i>SYNGR1</i>	rs61616683-T: 2.3	EAS

GTEx eQTLs and functional analysis

Almost all of the SLE risk SNPs function as *cis*-eQTLs in different human tissue types for several genes (Table 2). The relative magnitude of genes influenced by the adaptive and non-adaptive SLE eQTLs in various tissue types is presented in Fig. 1. The eQTLs (both adaptive and non-adaptive eQTLs) influence gene expression (eQTL–eGene interactions) mainly in tissues such as the thyroid, different brain tissues, skin, whole blood, adipose, oesophagus and muscle, suggesting that SLE-associated genetic variants in regulatory regions have tissue-specific effects. In comparison to the non-adaptive eQTLs, adaptive eQTLs influence significantly fewer ($p < 0.03$, $t = 1.8459$, t-test is one-tailed) genes in the different tissues. The funMotifs database shows (Table 2) that several positively selected loci are located in binding motifs for different transcription factors (TFs). The TFs that potentially bind to these sequence motifs are involved in a large spectrum of important biological functions including cell differentiation and proliferation, hematopoiesis, apoptosis, tissue development (NR4A2, MEF2C, Mitf, Mecom, THAP1, TFAP2A, TFAP2C), type I and type III interferon (IFN) signaling (STAT1::STAT2), insulin signaling and regulation of metabolic homeostasis in response to oxidative stress, regulation of T cell homeostasis (Foxo1), and regulation of melanocyte development (Mitf).

Table 2

Functions of SNPs under positive selection at SLE risk loci. The positively selected SNPs function as eQTLs for different genes in various GTEx tissues. The predictions for the TF motifs of the adaptive SNPs were obtained from the funMotifs database for blood tissue. Given is also timing of positive selection in thousand years ago (kya).

Beneficial allele/ alternate allele	Gene location	Timing of selection	dbSNP/GTEx functions	TF motifs (blood)	GTEx eQTL–eGene
rs4655706-G/A	<i>IL12RB2</i>	9–11	intron/eQTL	-	<i>IL12RB2, SERBP1</i>
rs10176621-T/C	<i>STAT4</i>	19–21	intergenic/eQTL	-	<i>STAT4</i>
rs4681674-G/A; rs4681834-T/G	<i>ABHD6</i>	7–10 9–13	intron/eQTL; intron/eQTL	- -	<i>ABHD6, DNASE1L3, KCTD6, PXX, PDHB, RPP14, ACOX2, ABHD6, PDHB, PXX, RPP14</i>
rs583911-G/A	<i>IL12A</i>	13–19	intron/eQTL	-	<i>IL12A</i>
rs4479588-G/T; rs9682946-A/G	<i>TPRG1</i>	10–16 14–19	intron; intron	NR4A2 Mitf	- -
rs9307257-G/A	<i>BANK1</i>	5–9	intron/eQTL	STAT1::STAT2	<i>FLJ20021</i>
rs7708392-G/C	<i>TNIP1</i>	15–20	intron/eQTL	-	<i>TNIP1, ZNF300P1</i>
rs11755393-A/G	<i>UHRF1BP1</i>	12–15	missense/eQTL	-	<i>UHRF1BP1, SNRPC, C6orf106, FANCE, RPS10, TAF11</i>
rs2428793-C/G; rs6460087-A/G	<i>GTF2I</i>	15–21 14–19	intron; 3'-UTR/eQTL	- PAX5	- <i>GTF2IRD2, GTF2I, NCF1, PMS2P5, STAG3L2</i>
rs1015877-C/G	<i>HIP1</i>	20–25	intron/eQTL	MEF2C	<i>HIP1</i>
rs10086521-T/C	<i>XKR6</i>	24–31	intron/eQTL	TFAP2A, TFAP2C	<i>BLK, CTSB, FAM167A, FDFT1, MSRA, MTMR9, NEIL2, RP1L1, SLC35G5, SUB1P1, TDH</i>

Beneficial allele/ alternate allele	Gene location	Timing of selection	dbSNP/GTEx functions	TF motifs (blood)	GTEx eQTL–eGene
rs4841563-T/G	<i>BLK</i>	28–35	intergenic/eQTL	-	<i>BLK, DEFB134, FAM167A, FAM66A, FDFT1, MTMR9, NEIL2, TDH</i>
rs75346988-G/A	<i>JAK2</i>	8–9	intron	Mecom	-
rs2732552-T/C	<i>CD44</i>	10–14	intergenic/eQTL	-	<i>CD44, APIP</i>
rs6539014-A/G	<i>DRAM1</i>	7–11	intron/eQTL	-	<i>DRAM1, CHPT1, GNPTAB, RNY1P16</i>
rs10774625-A/G	<i>ATXN2</i>	2–3	intron/eQTL	Foxo1	<i>ALDH2, ADAM1B, LINC01405, SH2B3, TMEM116</i>
rs1274641-G/A	<i>RAD51B</i>	30–35	intron/eQTL	-	<i>RDH12</i>
rs7197422-C/G	<i>CLEC16A</i>	35–45	intron/eQTL	THAP1	<i>CLEC16A, HNRNPCP4</i>
rs140490-G/T	<i>UBE2L3</i>	23–28	intron/eQTL	Mitf	<i>UBE2L3, CCDC116, TOP3BP1, YDJC</i>
rs61616683-T/C	<i>SYNGR1</i>	23–33	intron/eQTL	THAP1	<i>SYNGR1, MGAT3</i>

Estimate of timing of positive selection

The timing of the selection of the loci *IL12A*, *STAT4*, *TPRG1*, *TNIP1*, *HIP1*, *XKR6*, *BLK*, *RAD51B*, *UBE2L3*, *SYNGR1* falls in the period before the Neolithic Transition, between 45 kya and 13 kya. The timing of selection for the loci *IL12RB2*, *ABHD6*, *BANK1*, *JAK2*, *CD44*, *DRAM1* and *SH2B3-ATXN2*, were estimated to a time frame of about 13 kya to 2 kya, i.e, during or after the Neolithic Transition (Table 2). Because most of these genes are involved in immunity, cell signaling, and metabolism, it is difficult to determine which selective mechanisms drive positive selection (Additional file 3). However, timing and eQTLs indicate that both environmental (i.e., pathogens) and lifestyle changes during and after the Neolithic Transition become possibly selective effective. The most recent timing for positive selection was calculated for the rs10774625-A allele at the *SH2B3-ATXN2* locus which is only in Europeans under positive selection. This SNP functions as eQTLs in different tissues for the downstream gene aldehyde dehydrogenase 2 (*ALDH2*). The *ALDH2* enzyme plays a key role in the major oxidative pathway of alcohol metabolism. The beneficial allele, which reaches a frequency of almost 50% in Europeans, is associated with increased expression. Because of the timing of positive selection, we think that the primary target of selection is the

ALDH2 gene rather than *ATXN2* or *SH2B3*. We therefore discuss below a possible novel link between *ALDH2* and SLE.

Discussion

In this study, we identified several SLE-associated loci that are population-specifically under positive selection. The timing of positive selection of some loci were estimated at time frames during or after the Neolithic Transition (< 13 kya); this indicates that both environmental factors and recent human cultural-driven lifestyle possibly played roles in driving selection at these loci. The variants at the SLE risk region 7q11.23 appear to be under positive selection in all analysed human populations. Nonetheless, the different haplotype structures and the occurrence of gene copy number variation at this genomic region pose challenges in identifying the true causal SLE variants and complicate the allocation of the primary target of positive selection. The F_{ST} analysis revealed high genetic ($F_{ST} > 0.25$) differentiation for most of the positively selected loci. Thus, positive selection is an important factor influencing the frequency of SLE risk alleles in different human populations. This is in line with two studies reporting positive selection as an important mechanism that can cause elevated population-specific frequencies of autoimmune risk loci [18, 19]. Similar to a previous study, we found that the number of positively selected SLE GWAS loci is higher in Europeans than in Africans [23]. Additionally, we found that no signatures of positive selection on SLE risk SNPs were shared between Africans and Europeans (exception: SLE risk region 7q11.23). Most SLE GWAS, however, analysed mainly individuals with European genetic ancestry, which could result in a bias of selected loci when studying positive selection at the SLE-associated loci.

Almost all SNPs that are under positive selection function as eQTLs in different tissue types (Table 2). This is in line with a recent study which showed – for African and European individuals – that *cis*-eQTLs were significantly enriched for high iHS values, indicating the importance of regulatory genetic variation in recent human evolution [21, 22]. We also detected that several TF binding motifs are located in adaptive sequences. Thus, it is not fully clear whether the potential SLE risk genes are *per se* under positive selection, or whether this pertains to eQTLs and TF binding motifs that regulate the expression of other genes. The adaptive eQTLs influence gene expressions mainly in tissues such as the thyroid, brain, skin, whole blood, adipose, oesophagus and muscle, suggesting that these tissues were the primary targets of recent genetic adaptation. Moreover, few positive selected SNPs that function as eQTLs also influence genes that are involved in immunity such as defensin beta 134 (*DEFB134*), APAF1 interacting protein (*APIP*) and neutrophil cytosolic factor 1 (*NCF1*). The latter gene, which is located in very close vicinity to *GTF2I* locus, has been reported to be associated with SLE in Japanese populations [24]. Furthermore, while some adaptive eQTLs affect the expression of multiple genes, the majority affect – as opposed to non-adaptive eQTLs – significantly fewer genes and tissues. This finding is in line with [25], who reported that adaptive eQTLs tend to affect fewer tissues than non-adaptive eQTLs. These results suggest that genetic adaptation reduces the effect of pleiotropy.

Purifying selection is expected to purge deleterious genetic polymorphisms from a population's gene pool. Importantly, our findings show that risk variants are retained in the human genome. This somewhat

contradictory result is attributable to the fact that phenotypic adaptations (i.e., epigenetics and eQTLs) can occur faster than genotypic adaptations. These changes entail the potential of a population to adapt more rapidly to environmental stimuli via modification of gene expression without having to alter the genetic code. Moreover, for a genetic polymorphism to be subjected to negative selection, it must exert a drastic harmful effect in terms of reducing individual fitness. SLE mainly affects women of reproductive age and this disease therefore clearly reduces the fitness of affected individuals. From an evolutionary perspective, the question arises as to why the underlying alleles have not been driven to low frequencies if SLE reduces survival and reproduction chances? Pathogen-mediated positive selection is thought to drive genetic adaptation at loci involved in immune responses. However, recent lifestyle changes during as well as after the Neolithic Transition may have become selectively effective (e.g., as shown for dairy farming and the associated genetic adaptation of the lactase gene).

Our study found a very recent timing of positive selection (2.2 kya – 3.2 kya) for the beneficial allele rs10774625-A in Europeans at the SLE risk locus *SH2B3-ATXN2*. The protein encoded by *SH2B3* influences a variety of signalling pathways mediated by Janus kinase (JAK) and receptor tyrosine kinases (RTKs), acts as a negative regulator in cytokine signalling, and is involved in growth factor signalling activities [26]. The gene *ATXN2* lies very close to this gene. It encodes a polyglutamine protein involved in RNA metabolism and metabolic homeostasis [27]. The *SH2B3-ATXN2* region was found to be associated with various other diseases such as rheumatoid arthritis [28], type 1 diabetes [29] and coronary artery disease [30]. Our study shows that the *SH2B3-ATXN2* region is exclusively under strong positive selection in Europeans, as reported in two previous studies [31, 32]. *SH2B3* has been suggested to play a role in protecting against bacterial infection because a risk allele for celiac diseases was found to be associated with activation of the NOD2 recognition pathway. This could potentially explain the selective sweep at this locus [33]. Importantly, we recorded signatures of positive selection across the genes *SH2B3-ATXN2* (Additional file 4) which are also characterised by strong LD (Additional file 5); their close proximity on the chromosome hinders determining which of these two genes is actually under positive selection. Either one of the genes or potentially both are under positive selection; another possibility is that the identified eQTLs within this region are under positive selection because they control the expression of the downstream *ALDH2* gene (Additional file 5 shows that *SH2B3-ATXN2* and *ALDH2* are located in the same extended LD block in Europeans). This suggests that the *ALDH2* gene is the primary target of positive selection in Europeans. The encoded protein of this gene, which functions in the mitochondrial matrix, belongs to the aldehyde dehydrogenase family of proteins and represents the second enzyme of the major oxidative pathway of alcohol metabolism; it plays a key role in oxidising acetaldehyde into nontoxic acetate. The alcohol flush reaction is predominantly shown by people of Asian ancestry carrying an *ALDH2*2* allele, i.e., carrying the inactive isozyme with limited activity to convert acetaldehyde into acetate. On contrast, people of Caucasian ancestry usually have only the active isozyme (*ALDH2*1* allele). The positively selected allele rs10774625-A as well as other positively selected SNPs in this genomic region in Europeans function as eQTLs and are associated with increased *ALDH2* expression. We therefore conclude that the culturally driven European lifestyle involving high alcohol consumption is the driving force for positive selection at this locus. The result is an increasing frequency

of the SLE risk allele in Europeans. The frequency (in the 1000 Genomes populations) of the beneficial allele rs10774625-A in Europeans is about 0.48, whereas it is virtually absent in East-Asians. In Africans the allele frequency at this locus is 0.02; in South-Asians about 0.07 (Additional file 1). Finally, Neanderthal and Denisovan sequences show the ancestral allele (G) at this SNP.

How is this beneficial allele possibly linked to SLE susceptibility in Europeans? *ALDH2* is not only a major detoxification enzyme for ethanol-derived acetaldehyde but also mediates the activation of nitric-oxide synthases (NOS), a family of enzymes catalysing the production of nitric oxide (NO) [34]. NO is involved in various vital physiological processes including host defence. However, NO can also form peroxynitrite (ONOO^-), reactive NO species (RNOS) exerting various proinflammatory actions [34, 35]. Acute or chronic exposure to ethanol is apparently associated with increased RNOS through induction of NOS expression, protein nitration and lipid oxidation. Studies indicate that increased inducible NOS activity is associated with the progression of SLE [36]. Other studies, however, indicate that moderate alcohol consumption is associated with decreased SLE risk [37, 38]. We propose that the beneficial allele associated with increased *ALDH2* expression has positive effects on ethanol detoxification (after increased alcohol consumption) and NOS activation. Nonetheless, in SLE progression, increased inducible NOS activity (potentially due to oxidative stress) limits *ALDH2* activity, possibly leading to (sex-specific [39]) increased formation of RNOS and reactive aldehydes (such as 4-hydroxynonenal and malondialdehyde). This alters the immune response in SLE. We thus suggest that this risk SNP acts as a 'promoter' rather than 'trigger' of SLE in Europeans. Furthermore, this SNP is located within enhancer histone marks, and the funMotifs database shows that it is also located within the binding sequence for the transcription factor forkhead box protein O1 (Foxo1). This TF is involved in regulating glucose metabolism, insulin signalling and regulates metabolic homeostasis in response to oxidative stress. Foxo1 also interacts with NAD-dependent deacetylase sirtuin1 (*SIRT1*), which has been shown to inhibit T cell activation [40]. Thus, the regulatory elements at this locus, including the effects of positive selection, could contribute to substantial phenotypic/trait differences among individuals with different genetic ancestries.

Conclusion

Positive selection acts mainly on regulatory elements such as *cis*-eQTLs and transcription factor binding motifs, thereby affecting the expression of the various SLE-associated genes and genes located near the risk loci. Most loci analysed are population-specifically under positive selection, suggesting local adaptation to specific environmental/lifestyle conditions. Thus, positive selection influences the frequency of haplotypes encompassing the genetic risk polymorphisms in populations differently. Moreover, we also determined that adaptive eQTLs affect the expression of fewer genes than non-adaptive eQTLs, suggesting a limited range of effect of an eQTL at SLE risk sites that show signatures of positive selection. We conclude that population-specific adaptive eQTLs contribute to the observed variation in specific manifestations and complications of SLE in different ethnicities. We also highlight the proposed novel link between positively selected eQTLs at the *SH2B3-ATXN2* SLE risk locus in Europeans and the physiological pathway via *ALDH2* in SLE.

Methods

SLE GWAS data

From the NHGRI-EBI GWAS catalog (<https://www.ebi.ac.uk/gwas/>) we obtained (accessed November 2019 - January 2020) 31 publications and the reported risk loci for SLE (Additional file 6). We included risk loci with a reported statistical threshold of $< 5 \times 10^{-8}$ in the studies, which yielded 145 genetic loci associated with SLE. However, we excluded the human leucocyte antigen (HLA) region because several studies already showed that HLA loci bear signatures of both positive and balancing selection.

Human 1000 genomes project phase 3 data

We used the phased genomic datasets from the 1000 genomes projects (phase 3; <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>) [41]. We obtained single nucleotide polymorphism (SNP) data from 12 human populations with the following genetic ancestries: African ancestry (AFR), European ancestry (EUR), South-Asian ancestry (SAS) and East-Asian ancestry (EAS) (Additional file 7). We used the software program PLINK 1.9 [42] (<https://www.cog-genomics.org/plink/1.9/> and VCFtools v0.1.14 [43] (<https://vcftools.github.io/index.html>) to process variant call format (VCF) files. We excluded structural variants and all SNPs that deviated from Hardy-Weinberg equilibrium (with p -value $< 1e^{-6}$). SNP positions are in accordance with the human genome version GRCh37/hg19 (<https://genome-euro.ucsc.edu/>).

Positive selection and F_{ST} analysis

We used the integrated Haplotype Score (iHS) statistics [44] which are implemented in the software program selscan version 1.2.0a (<https://github.com/szpiech/selscan>) [45] to detect positive selection in genomic population data. These statistics are based on the model of selective sweeps, i.e., a beneficial mutation rises quickly in frequency until it reaches fixation within a population and thereby simultaneously reduces genetic diversity at sites in very close genetic vicinity. The iHS statistic compares integrated EHH values between alleles at a given SNP and is based on the decay of the haplotype homozygosity model as a function of recombination distance. The underlying rationale is that selected alleles will have an unusually long-range linkage disequilibrium (LD) given their frequency in the population. Significant (based on empirical p -values < 0.05) negative iHS values (absolute iHS score < -2.0) indicate unusually long haplotypes carrying the derived allele, and significant positive values (absolute iHS score > 2.0) are associated with long haplotypes carrying the ancestral allele [44]. All scans were run on phased whole chromosome data with the default selscan model parameters. The unstandardized iHS scores were normalized in frequency bins across the entire genome using the script *norm*, provided with the selscan program. We considered candidate regions of positive selection as genomic regions containing clustering of SNPs with high iHS statistics. This was quantified as ≥ 20 SNPs with $|iHS| > 2.0$ in 100 kb nonoverlapping windows. We used the LDlink, a web-based application (<https://analysistools.cancer.gov/LDlink/?tab=home>) [46], to explore population-specific linking of correlated alleles between the SLE risk SNPs and SNPs under positive selection. Pairwise F_{ST} were

calculated for each SNP under positive selection using the Weir & Cockerham F_{ST} calculation [47] implemented in the PLINK program. Negative F_{ST} values were set to zero.

Genotype-Tissue Expression (GTEx) eQTL and functional analysis

We used the GTEx Portal V8 Release (<https://www.gtexportal.org/home/>) [48] to obtain expression quantitative trait loci (eQTLs) (accessed January - August 2020 (dbGaP Accession phs000424.v8.p2)). In recent years, hundreds of thousands of human eQTLs have been catalogued in the GTEx database. GTEx reports global RNA expression within individual tissues: treating the expression levels of genes as quantitative traits enables identifying variations in gene expression that are highly correlated with genetic variation as eQTLs [49]. We used the R program [50] and the R packages ggplot2 [51] to calculate and plot the relative magnitude of eQTLs–eGene interaction in different human tissues (eGene is defined as a gene in which the expression level is significantly associated with an eQTL). We used funMotifs DB (<http://bioinf.icm.uu.se:3838/funmotifs/>) (accessed August 2020) to analyse whether any of the SNPs under selection contain transcription factor (TF) motifs [52]. In this analysis we used as input type ‘genomic region’ (and the tissue blood) containing the core SNP under selection including 10 SNPs up- and downstream of that SNP. The UniProt database (<https://www.uniprot.org/>) was used to extract information on gene/protein functions. The Reactome Database (<https://reactome.org/>) was used to obtain locations of the gene in the Pathway Browser. Genomic feature plots were generated by using R package Gviz [53].

Estimation for the timing of selection on beneficial alleles

We estimated the timing of selection on a beneficial allele by applying a recently published method Startmrca [54]. The method applies a Markov chain Monte Carlo simulation (MCMC) that samples over the unknown ancestral haplotype to generate a sample of the posterior distribution for the time of the common ancestor (TMRCA). The model takes advantage of both the length of the ancestral haplotype on each chromosome and the accumulation of derived mutations on the ancestral haplotype to generate a sample of the posterior distribution for the time to the TMRCA. The model requires a sample (panel) containing the selected allele and a reference panel of haplotypes without the selected allele. We chose four representative populations for the analyses: LWK (Africa), GBR (Europe), BEB (South Asia) and CHB (East Asia). We used recombination rates from the sex-averaged recombination map from the deCODE to model recombination rate variation across the human genome. We analysed 1 Mb regions surrounding each genetic variant under selection with an assumed mutation rate of 1.6×10^{-8} . We ran three independent MCMC chains, each with 25,000 iterations. Subsequently, we discarded the first 9000 iterations (burn-in), retaining the remaining successful iterations. Timing of selection (TMRCA estimates) are given in thousand years ago (kya). We assumed 25 years as generation time.

Abbreviations

IL12RB2: Interleukin-12 receptor subunit beta-2

ABHD6: Abhydrolase domain-containing protein 6

PXK: PX Domain Containing Serine/Threonine Kinase Like

IL12A: Interleukin-12 subunit alpha

LPP: LIM Domain Containing Preferred Translocation Partner In Lipoma

TPRG1: Tumor protein p63-regulated gene 1 protein

BANK1: B-cell scaffold protein with ankyrin repeats

TNIP1: TNFAIP3-interacting protein 1

UHRF1BP1: UHRF1-binding protein 1-like

GTF2IRD1: General transcription factor II-I repeat domain-containing protein 1

GTF2I: General transcription factor II-I

GTF2IRD2: General transcription factor II-I repeat domain-containing protein 2A

HIP1: Huntingtin-interacting protein 1

XKR6: XK-related protein 6

C8orf12: FAM167A-AS1

FAM167A: Family With Sequence Similarity 167 Member A

BLK: Tyrosine-protein kinase Blk

JAK2: Tyrosine-protein kinase JAK2

CD44: CD44 antigen

DRAM1: DNA damage-regulated autophagy modulator protein 1

SH2B3: SH2B adapter protein 3

Ataxin-2: Ataxin-2

RAD51B: DNA repair protein RAD51 homolog 2

CLEC16A: Protein CLEC16A

CIITA: Class II Major Histocompatibility Complex Transactivator

SOCS1: Suppressor Of Cytokine Signaling 1

UBE2L3: Ubiquitin-conjugating enzyme E2 L3

SYNGR1: Synaptogyrin-1

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The SLE-GWAS data can be obtained from <https://www.ebi.ac.uk/gwas/>. The gnomc data can be obtained from 1000 Genomes database (see Additional file 6). The generated iHS dataset is available from the corresponding author on reasonable request.

Competing interests

No competing interests.

Funding

Not applicable.

Authors' contributions

HS conceived the study; OV, HS, TG and DM analysed the data; VO and HS wrote the paper. All authors have read and approved the manuscript.

Acknowledgements

We thank Michael Stachowitsch from Department of Evolutionary Anthropology for valuable comments on the manuscript.

References

1. Mohan C, Putterman C: **Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis.** *Nature Reviews Nephrology* 2015, **11**(6):329-341.
2. Katsuyama T, Tsokos GC, Moulton VR: **Aberrant T Cell Signaling and Subsets in Systemic Lupus Erythematosus.** *Frontiers in Immunology* 2018, **9**.
3. Yen EY, Singh RR: **Lupus-An Unrecognized Leading Cause of Death in Young Females: A Population-Based Study Using Nationwide Death Certificates, 2000-2015.** *Arthritis & Rheumatology* 2018, **70**(8):1251-1255.
4. Pons-Estel GJ, Alarcon GS, Scofield L, Reinlib L, Cooper GS: **Understanding the Epidemiology and Progression of Systemic Lupus Erythematosus.** *Seminars in Arthritis and Rheumatism* 2010, **39**(4):257-268.
5. Ghodke-Puranik Y, Niewold TB: **Immunogenetics of systemic lupus erythematosus: A comprehensive review.** *Journal of Autoimmunity* 2015, **64**:125-136.
6. Lim SS, Bayakly AR, Helmick CG, Gordon C, Easley KA, Drenkard C: **The Incidence and Prevalence of Systemic Lupus Erythematosus, 2002-2004.** *Arthritis & Rheumatology* 2014, **66**(2):357-368.
7. Somers EC, Marder W, Cagnoli P, Lewis EE, DeGuire P, Gordon C, Helmick CG, Wang L, Wing JJ, Dhar JP *et al*: **Population-Based Incidence and Prevalence of Systemic Lupus Erythematosus The Michigan Lupus Epidemiology and Surveillance Program.** *Arthritis & Rheumatology* 2014, **66**(2):369-378.
8. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Royburman P, Walker A, Mack TM: **A revised estimate of twin concordance in systemic lupus erythematosus.** *Arthritis and Rheumatism* 1992, **35**(3):311-318.
9. Reichlin M, Harley JB, Lockshin MD: **Serologic studies of monozygotic twins with systemic lupus erythematosus.** *Arthritis and Rheumatism* 1992, **35**(4):457-464.
10. Ulf-Moller CJ, Svendsen AJ, Viemose LN, Jacobsen S: **Concordance of autoimmune disease in a nationwide Danish systemic lupus erythematosus twin cohort.** *Seminars in Arthritis and Rheumatism* 2018, **47**(4):538-544.
11. Kuo CF, Grainge MJ, Valdes AM, See LC, Luo SF, Yu KH, Zhang WY, Doherty M: **Familial Aggregation of Systemic Lupus Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families.** *Jama Internal Medicine* 2015, **175**(9):1518-1526.
12. Lawrence JS, Martins CL, Drake GL: **A family survey of lupus erythematosus. 1. Heritability.** *Journal of Rheumatology* 1987, **14**(5):913-921.
13. Bentham J, Morris DL, Graham DSC, Pinder CL, Tombleson P, Behrens TW, Martin J, Fairfax BP, Knight JC, Chen L *et al*: **Genetic association analyses implicate aberrant regulation of innate and**

- adaptive immunity genes in the pathogenesis of systemic lupus erythematosus.** *Nature Genetics* 2015, **47**(12):1457-+.
14. Fujio K, Takeshima Y, Nakano M, Iwasaki Y: **Review: transcriptome and trans-omics analysis of systemic lupus erythematosus.** *Inflammation and Regeneration* 2020, **40**(1).
 15. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE *et al*: **Systematic identification of trans eQTLs as putative drivers of known disease associations.** *Nature Genetics* 2013, **45**(10):1238-U1195.
 16. Westra HJ, Franke L: **From genome to function by studying eQTLs.** *Biochimica Et Biophysica Acta-Molecular Basis of Disease* 2014, **1842**(10):1896-1902.
 17. Pierce SE, Booms A, Prah J, van der Schans EJC, Tyson T, Coetzee GA: **Post-GWAS knowledge gap: the how, where, and when.** *Npj Parkinsons Disease* 2020, **6**(1).
 18. Ramos PS, Shaftman SR, Ward RC, Langefeld CD: **Genes associated with SLE are targets of recent positive selection.** *Autoimmune diseases* 2014, **2014**:203435-203435.
 19. Raj T, Kuchroo M, Replogle JM, Raychaudhuri S, Stranger BE, De Jager PL: **Common Risk Alleles for Inflammatory Diseases Are Targets of Recent Positive Selection.** *American Journal of Human Genetics* 2013, **92**(4):517-529.
 20. Wang YF, Lau YL, Yang WL: **Genetic studies on systemic lupus erythematosus in East Asia point to population differences in disease susceptibility.** *American Journal of Medical Genetics Part C-Seminars in Medical Genetics* 2019, **181**(2):262-268.
 21. Nedelec Y, Sanz J, Baharian G, Szpiech ZA, Pacis A, Dumaine A, Grenier JC, Freiman A, Sams AJ, Hebert S *et al*: **Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens.** *Cell* 2016, **167**(3):657-+.
 22. Quach H, Rotival M, Pothlichet J, Loh YHE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M *et al*: **Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations.** *Cell* 2016, **167**(3):643-+.
 23. Brinkworth JF, Barreiro LB: **The contribution of natural selection to present-day susceptibility to chronic inflammatory and autoimmune disease.** *Current Opinion in Immunology* 2014, **31**:66-78.
 24. Yokoyama N, Kawasaki A, Matsushita T, Furukawa H, Kondo Y, Hirano F, Sada K, Matsumoto I, Kusaoi M, Amano H *et al*: **Association of NCF1 polymorphism with systemic lupus erythematosus and systemic sclerosis but not with ANCA-associated vasculitis in a Japanese population.** *Scientific Reports* 2019, **9**.
 25. Quiver MH, Lachance J: **Adaptive eQTLs reveal the evolutionary impacts of pleiotropy and tissue-specificity, while contributing to health and disease in human populations.** *bioRxiv* 2018:444737.
 26. Devalliere J, Charreau B: **The adaptor Lnk (SH2B3): An emerging regulator in vascular cells and a link between immune and inflammatory signaling.** *Biochemical Pharmacology* 2011, **82**(10):1391-1402.
 27. Carmo-Silva S, Nobrega C, de Almeida LP, Cavadas C: **Unraveling the Role of Ataxin-2 in Metabolism.** *Trends in Endocrinology and Metabolism* 2017, **28**(4):309-318.

28. Coenen MJH, Trynka G, Heskamp S, Franke B, van Diemen CC, Smolonska J, van Leeuwen M, Brouwer E, Boezen MH, Postma DS *et al*: **Common and different genetic background for rheumatoid arthritis and coeliac disease.** *Human Molecular Genetics* 2009, **18**(21):4195-4203.
29. Auburger G, Gispert S, Lahut S, Omur O, Damrath E, Heck M, Basak N: **12q24 locus association with type 1 diabetes: SH2B3 or ATXN2?** *World Journal of Diabetes* 2014, **5**(3):316-327.
30. Wild PS, Felix JF, Schillert A, Teumer A, Chen MH, Leening MJG, Volker U, Grossmann V, Brody JA, Irvin MR *et al*: **Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function.** *Journal of Clinical Investigation* 2017, **127**(5):1798-1812.
31. Barreiro LB, Quintana-Murci L: **From evolutionary genetics to human immunology: how selection shapes host defence genes.** *Nature Reviews Genetics* 2010, **11**(1):17-30.
32. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M *et al*: **Genome-wide patterns of selection in 230 ancient Eurasians.** *Nature* 2015, **528**(7583):499-+.
33. Zhernakova A, Elbers CC, Ferwerda B, Romanos J, Trynka G, Dubois PC, de Kovel CGF, Franke L, Oosting M, Barisani D *et al*: **Evolutionary and Functional Analysis of Celiac Risk Loci Reveals SH2B3 as a Protective Factor against Bacterial Infection.** *American Journal of Human Genetics* 2010, **86**(6):970-977.
34. Deng XS, Deitrich RA: **Ethanol Metabolism and Effects: Nitric Oxide and its Interaction.** *Current Clinical Pharmacology* 2007, **2**(2):145-153.
35. Wang G, Pierangeli SS, Willis R, Gonzalez EB, Petri M, Khan MF: **Significance of Lipid-Derived Reactive Aldehyde-Specific Immune Complexes in Systemic Lupus Erythematosus.** *Plos One* 2016, **11**(10).
36. Pan L, Yang SR, Wang JH, Xu M, Wang SF, Yi HF: **Inducible nitric oxide synthase and systemic lupus erythematosus: a systematic review and meta-analysis.** *Bmc Immunology* 2020, **21**(1).
37. Wang J, Pan HF, Ye DQ, Su H, Li XP: **Moderate alcohol drinking might be protective for systemic lupus erythematosus: a systematic review and meta-analysis.** *Clinical Rheumatology* 2008, **27**(12):1557-1563.
38. Barbhaiya M, Lu B, Sparks JA, Malspeis S, Chang SC, Karlson EW, Costenbader KH: **Influence of Alcohol Consumption on the Risk of Systemic Lupus Erythematosus Among Women in the Nurses' Health Study Cohorts.** *Arthritis Care & Research* 2017, **69**(3):384-392.
39. Karpuzoglu E, Ahmed SA: **Estrogen regulation of nitric oxide and inducible nitric oxide synthase (iNOS) in immune cells: Implications for immunity, autoimmune diseases, and apoptosis.** *Nitric Oxide-Biology and Chemistry* 2006, **15**(3):177-186.
40. Zhang JN, Lee SM, Shannon S, Gao BX, Chen WM, Chen A, Divekar R, McBurney MW, Braley-Mullen H, Zaghoulani H *et al*: **The type III histone deacetylase Sirt1 is essential for maintenance of T cell tolerance in mice.** *Journal of Clinical Investigation* 2009, **119**(10):3048-3058.
41. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P *et al*: **A global reference for human genetic variation.** *Nature* 2015, **526**(7571):68-

74.

42. Chang CC, Chow CC, Tellier L, Vattikuti S, Purcell SM, Lee JJ: **Second-generation PLINK: rising to the challenge of larger and richer datasets**. *Gigascience* 2015, **4**.
43. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: **The variant call format and VCFtools**. *Bioinformatics* 2011, **27**(15):2156-2158.
44. Voight BF, Kudaravalli S, Wen XQ, Pritchard JK: **A map of recent positive selection in the human genome (vol 4, pg 154, 2006)**. *Plos Biology* 2006, **4**(4):659-659.
45. Szpiech ZA, Hernandez RD: **selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection**. *Molecular Biology and Evolution* 2014, **31**(10):2824-2827.
46. Machiela MJ, Chanock SJ: **LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants**. *Bioinformatics* 2015, **31**(21):3555-3557.
47. Weir BS, Cockerham CC: **ESTIMATING F-STATISTICS FOR THE ANALYSIS OF POPULATION-STRUCTURE**. *Evolution* 1984, **38**(6):1358-1370.
48. Ardlie KG, DeLuca DS, Segre AV, Sullivan TJ, Young TR, Gelfand ET, Trowbridge CA, Maller JB, Tukiainen T, Lek M *et al*: **The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans**. *Science* 2015, **348**(6235):648-660.
49. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park Y, Parsana P, Segre AV *et al*: **Genetic effects on gene expression across human tissues**. *Nature* 2017, **550**(7675):204+.
50. Team RC: **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. In. Vienna, Austria; 2020.
51. Wickham H: **ggplot2: Elegant Graphics for Data Analysis**. *Ggplot2: Elegant Graphics for Data Analysis* 2009:1-212.
52. Umer HM, Smolinska-Garbulowska K, Marzouka N-a-d, Khaliq Z, Wadelius C, Komorowski J: **funMotifs: Tissue-specific transcription factor motifs**. *bioRxiv* 2019:683722.
53. Hahne F, Ivanek R: **Visualizing Genomic Data Using Gviz and Bioconductor**. *Statistical Genomics: Methods and Protocols* 2016, **1418**:335-351.
54. Smith J, Coop G, Stephens M, Novembre J: **Estimating Time to the Common Ancestor for a Beneficial Allele**. *Molecular Biology and Evolution* 2018, **35**(4):1003-1017.

Figures

Figure 1

Relative magnitude of SLE–associated eQTLs–eGene interactions (circles) across different human tissues (y-axis); x-axis: proportion of adaptive eQTLs that influence gene expressions in tissues.

Supplementary Files

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

- [Additionalfile1.docx](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.docx](#)
- [Additionalfile4.docx](#)
- [Additionalfile5.docx](#)
- [Additionalfile6.docx](#)

- [Additionalfile7.docx](#)
- [SupplementaryInformation.docx](#)