Gene 907 (2024) 148279

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Large-scale DNA sequencing identifies rare variants associated with Systemic Lupus Erythematosus susceptibility in known risk genes

Andrea Latini [a,*], Paola Borgiani [a], Giada De Benedittis [a], Cinzia Ciccacci [b], Lucia Novelli [b], Gerardo Pepe [c], Manuela Helmer-Citterich [c], Isabella Baldini [d], Carlo Perricone [e], Fulvia Ceccarelli [f], Fabrizio Conti [f], Generoso Ianniciello [d], Juan Caceres [d], Riccardo Ottalevi [d], Mattia Capulli [g], Giuseppe Novelli [a,h,i]

[a] Department of Biomedicine and Prevention, Genetics Section, University of Rome "Tor Vergata", Rome, Italy
[b] UniCamillus, Saint Camillus International University of Health Sciences, Rome, Italy
[c] Department of Biology, Centro di Bioinformatica Molecolare, University of Rome "Tor Vergata", Rome, Italy
[d] Dante Genomics Inc., New York, USA
[e] Rheumatology, Department of Medicine, University of Perugia, Piazzale Giorgio Menghini, 1, Perugia, Italy
[f] Reumatologia, Dipartimento di Scienze Cliniche Internistiche, Anestesiologiche e Cardiovascolari, Sapienza Università di Roma, Rome, Italy
[g] Department of Biotechnological and Applied Clinical Sciences, University of L'Aquila, L'Aquila, Italy
[h] IRCCS NEUROMED, Pozzilli, IS, Italy
[i] School of Medicine, Department of Pharmacology, Reno University of Nevada, NV, USA

ARTICLE INFO

Edited by: Fahri Ucar

Keywords:
Systemic Lupus Erythematosus
Genome sequencing
Rare variants

ABSTRACT

The identification of rare genetic variants associated to Systemic Lupus Erythematosus (SLE) could also help to understand the pathogenic mechanisms at the basis of the disease. In this study we have analyzed a cohort of 200 Italian SLE patients in order to explore the rare protein-coding variants in five genes (*TNFAIP3, STAT4, IL10, TRAF3IP2,* and *HCP5*) already investigated for commons variants found associated in our previous studies.

Genomic DNA of 200 SLE patients was sequenced by whole exome sequencing. The identified variants were filtered by frequency and evaluated by *in silico* predictions. Allelic association analysis was performed with standard Fisher's exact test.

Introducing a cutoff at MAF < 0.01, a total of 19 rare variants were identified. Seven of these variants were ultra-rare (MAF < 0.001) and six were absent in the GnomAD database. For *TNFAIP3* gene, the variant c.A1939C was observed in 4 SLE patients and it is located in a region enriched in phosphorylation sites and affects the predict affinity of specific kinases. In *TRAF3IP2* gene, we observed 5 different rare variants, including the novel variant c.G410A, located in the region that mediates interaction with TRAF6, and therefore a possible risk factor for SLE development. In *STAT4* gene, we identified 6 different rare variants. Among these, three missense variants decrease the stability of this protein. Moreover, 3 novel rare variants were detected in 3 SLE patients. In particular, c.A767T variant was predicted as damaging by six prediction tools.

Concluding, we have observed that even in genes whose common variability is associated with SLE susceptibility, it is possible to identify rare variants that could have a strong effect in the disease development and could therefore allow a better understanding of the functional domain involved.

1. Introduction

Systemic Lupus Erythematosus (SLE; OMIM#152700) is a chronic and heterogeneous autoimmune disease characterized by abnormal T- and B-cells response and production of autoantibodies (Karrar and Cunninghame Graham, 2018). The etiology of SLE is still partially

Abbreviations: SLE, Systemic Lupus Erythematosus; MAF, Minor Allele Frequency; GWAS, Genome-Wide Association Studies; WES, Whole Exome Sequencing; WGS, Whole Genome Sequencing; SNVs, Single Nucleotide Variants; BAM, Binary Alignment Map; VCF, Variant Call Format; CADD, Combined Annotation Dependent Depletion; VUS, Variants of Uncertain Significance.

* Corresponding author.
E-mail address: a.latini@med.uniroma2.it (A. Latini).

https://doi.org/10.1016/j.gene.2024.148279
Received 11 December 2023; Received in revised form 25 January 2024; Accepted 8 February 2024
Available online 13 February 2024

unknown, but it likely involves complex interactions among genetic and environmental factors. Several studies evidenced that the genetics cover an important role in the SLE development. Indeed, twin studies showed a higher concordance rate for monozygotic (24 %-35 %) and dizygotic (2 %-5%) (Ramos et al., 2010). On the other hand, genome-wide association studies (GWAS) have identified common alleles in several genes which increase the risk to develop SLE (Kwon et al., 2019). Moreover, rare genetic variants that can cause or contribute significantly to SLE susceptibility have been described, some of them with a strong effect, and > 30 causative genes for monogenic SLE have been reported (Demirkaya et al., 2020). For example, a recent study described rare variants in 98 SLE-associated candidate genes by using GWAS data from 4254 SLE patients (Martínez-Bueno and Alarcón-Riquelme, 2019). The identification of rare genetic variants associated to SLE can help to understand the pathogenic mechanisms at the basis of the disease. Most of the identified genes encode proteins involved in key pathogenic pathways, including type I interferon or NF-κB signalling pathways, the maintenance of immunological tolerance, DNA damage repair and inflammatory processes (Pan et al., 2020). This large genetic and phenotypic heterogeneity suggests that many novel genes involved in SLE etiology could be identified in the near future, also considering the rapidly evolving sequencing technology. Indeed, whole exome sequencing (WES) and whole genome sequencing (WGS) studies of the general population have revealed that the average subject carries dozens of potentially deleterious rare variants. Moreover, genetic variant identified by GWASs are often located within non-protein-coding sequences and collectively they are still far from fully explaining the measured heritability for SLE, suggesting additional genetic variants to be discovered yet. In this context, a WES study would be a reasonable complement to GWASs in terms of analyzing coding regions and searching for rare genetic variants.

In previous studies, associations between genetic polymorphisms in the *TNFAIP3, STAT4, IL10, TRAF3IP2*, and *HCP5* genes and SLE susceptibility have been described (Ciccacci et al., 2019; Ciccacci et al., 2014; Perricone et al., 2013). Single Nucleotide Variants (SNVs) located on these genes have also been extensively described in other studies, on different populations (Fan et al., 2011 Apr; Shancui-Zheng and Guoyuan-Lu, 2022; Yuan et al., 2019). However, each common variant in these genes has shown a modest effect size and altogether common variants are able to explain only in part the disease clinical variability. This prompted us to explore these genes, also for rare variants to evaluate if these could confer a predisposition to SLE susceptibility.

We re-analyzed a cohort of 200 Italian SLE patients by WES, in order to investigate rare protein-coding variants in selected genes previously studied for commons SNVs associated to SLE.

2. Materials and methods

2.1. Patients recruitment

Two hundred patients with SLE, classified in accordance with the revised 1997 American College of Rheumatology criteria were enrolled from Lupus Clinic of the Rheumatology Unit, Sapienza University of Rome. The study protocol included a complete physical examination and blood drawing. The clinical and laboratory data were collected in a standardized, computerized, electronically completed form including demographics, past medical history with date of diagnosis, comorbidities, and previous and concomitant treatments. Patient characteristics are described in Table 1. We included 200 SLE patients (89.2 % women), with a mean age of 43.08 + -11.15 years and an age at onset of 32.17 + -11.72 years. All subjects provided written informed consent in accordance with the Helsinki Declaration. Peripheral blood samples from all patients have been collected and stored at −20 ∘C until usage.

Table 1
Demographic and clinical characteristics of the 200 SLE patients.

Age (years ± SD)	43.08 ± 11.15
Age at onset (years ± SD)	32.17 ± 11.72
Sex (female)	89.2 %
Musculo-skeletal involvement	63.8 %
Photosensitivity	62.0 %
Malar rash	57.6 %
Aphthous/ulcers	16.4 %
Pericarditis	26.6 %
Pleurisy	17.6 %
Nephritis	32.0 %
Neuropsychiatric	13.3 %
Anemia	43.5 %
Leucopoenia	41.9 %
Thrombocytopenia	17.2 %
ANA	94.9 %
Anti-dsDNA	66.9 %
Anti-Sm	16.5 %
Anti-RNP	18.4 %
Anti-Ro/SSA	35.0 %
Anti-La/SSB	16.4 %
Anti-CL IgG or IgM	34.5 %
Anti-β2GPI IgG or IgM	20.1 %
LAC	26.5 %
C3 (below normal values)	55.5 %
C4 (below normal values)	44.5 %

aCL: anticardiolipin antibodies; ANA: antinuclear antibodies; Anti-b2GPI: anti- b2 glycoprotein I; Anti-dsDNA: anti-double-stranded DNA; C3: complement 3; C4: complement 4; Ig: immunoglobulin; LA: lupus anticoagulant; SLE: systemic lupus erythematosus.

2.2. Whole exome sequencing and data pre-processing

Genomic DNA was extracted from peripheral blood using Qiagen Blood DNA Mini Kit (Qiagen, Hilden, Germany). Library for all the samples was prepared using Twist Library Preparation Kit 2. WES was conducted on Novaseq 6000. Once the machine finished the last cycle of sequencing, we performed standard bioinformatic analysis to retrieve FASTQ, BAM and VCF file. Fastq files have been checked to assess the quality of the sequencing according to Q score. The alignment is performed using Dragen Bio-IT platform from Illumina. Variant calling is performed using Dragen variant caller replicating GATK best practices pipeline to ensure good quality variants. Sequencing data were aligned to the hg19 human reference genome. A minimum average depth coverage of 100X was considered suitable for sample analysis, according to the guidelines of the American College of Medical Genetics and Genomics.

The identified variants were filtered by frequency under the polymorphism threshold (MAF ≤ 0.01) and evaluated by genetic tolerance of resulting amino acid changes, and by *in silico* predictions of damage and function of domains bearing mutation. The summary data of the European population from gnomAD v2.1.1 and two hundred and thirteen age- and ethnicity-matched healty subjects were used as population controls. Allelic association analysis was performed with standard Fisher's exact test using default parameters. P value below 0.05 was considered as significant. A burden test has been performed using the TRAPD tool, to compare the variants in the VCF files of SLE patients and healty controls, filtered for effect prediction and for MAF lower than 0.01.

2.3. Prediction of the genetic variants impact on protein stability

Comprehending the precise influence of variants on protein stability is a crucial step for predicting their impact on protein conformation and structure. Existing methods have encountered difficulties in accurately assessing the effects of genetic variants on protein stability, often exhibiting a bias towards destabilizing variants and limited predictive capabilities. To overcome these challenges, we have utilized DDMut

(Zhou et al., 2023), an advanced and rapid network specifically designed to predict changes in Gibbs Free Energy resulting from single-point mutations. DDMut achieves this by considering both forward and hypothetical reverse mutations, effectively addressing the issue of model anti-symmetry.

2.4. Prediction of the kinase specific affinity

The affinity for a specific kinase was calculated using the Kinase Prediction tool accessible from Phosphositeplus (https://www.phosphosite.org/kinaseLibraryAction) (Hornbeck et al., 2015). This tool relies on a database of motifs associated with kinases identified through experimental work (Johnson et al., 2023) and collected in The Kinase Library. For each of the peptides centred on the phosphorylatable residue, affinity scores for the kinases were calculated for the wild-type and mutant sequences, respectively.

3. Results

We examined 200 Italian SLE patients for rare variants in five candidate genes previously investigated in our laboratory. Sequencing data from 199 patients passed quality check and were evaluated for subsequent analyses. Introducing a cutoff at MAF < 0.01, a total of 19 rare variants were identified, including 8 synonymous, 10 missense and 1 non-coding variants in *TNFAIP3*, *TRAF3IP2*, *STAT4* and *HCP5* genes (Table 2). Seven of these variants were ultra-rare (MAF < 0.001) and six were absent in the GnomAD database. In addition, several *in silico* scores have been calculated to assess the functional impact of identified non-synonymous variants such as the widely used CADD score (Table S1).

In *TNFAIP3* gene, we identified seven different rare variants, three missense and four synonymous variants. For five of these variants, the allelic frequency reported in GnomAD for the European reference population is very low and we detected a statistically significant difference compared to our SLE cohort. None of these five variants was observed in the analysed control group (see Table S2). In particular, the rare variant c.A1939C (rs142253225) was observed in four unrelated SLE patients and it was already reported in the literature in an individual with periodic fever and autoinflammatory disease (Fathalla et al., 2021), and in a family with autoinflammation due to haploinsufficiency of protein A20 (Mulhern et al., 2019). Instead, the variant c.G336C identified in a heterozygous patient, is not present in the GnomAD database and the *in silico* analysis gave uncertain significant as computational verdict.

In *TRAF3IP2* gene, we observed five different rare variants, including one novel variant. The four known variants were found with frequencies in line with those reported for the reference population in the GnomAD database and with those observed in our healty controls. Regarding the novel variant c.G410A, although it was predicted as likely benign by prediction tools, it is located in the region that mediates interaction with TRAF6, a critical and functional domain.

In *STAT4* gene, we identified six different rare variants, three missense and three synonymous variants. Among these, c.A383T was detected in three unrelated SLE patients, with a higher frequencies compared to that reported in the European non-Finnish population of the GnomAD database and it is not observed in our healty subject cohort. Moreover, three novel rare variants were detected in three different SLE patients. In particular, c.A767T (Asp256Val) variant was predicted as damaging by six prediction tools. Indeed, the Asp residue at position 256 is hydrophilic and has a negatively charged side chain, while the variant residue Val has an aliphatic side chain. The sequence alignment of the STAT4 protein with its orthologous proteins shows that the wild type residue is conserved, implying a structural or functional role for this residue in the STAT4 protein.

In *HCP5* gene, we observed only a rare non-coding variant, absent in our healty controls group and in the GnomAD database. No rare variant was detected in the *IL10* gene.

We examined clinical data from patients with SLE in whom we found rare variants in the selected genes, but these patients did not appear to have a different phenotype from other patients, nor did they show suggestive clinical signs.

We also performed a Burden Test to evaluate a possible difference in the total number of variants with MAF < 0.01 in SLE patients compared to healty control group. The analysis did not reveal an enrichment of rare variants for any of the examined genes in SLE patients. Since these tests involved many variants, regardless of their effect, the power of this analysis is certainly diminished and therefore, we performed a more detailed investigation of the rare variants identified in our study.

For the missense variants, an analysis of the protein stability was performed, to explore the possible effect of amino acid changes on the protein structure, providing a 3D model of the predicted results. The results showed that the three missense variants of *STAT4* gene decrease the stability of the protein, causing loss of secondary interactions (Fig. 1). In contrast, the predicted 3D models of the *TRAF3IP2*-encoded

Table 2
Rare variants (MAF ≤ 0.01) identified in patients with SLE.

Gene	Genomic position	Rs id	annotation	hgvs_c	hgvs_p	SLE patients MAF	GnomAD as control	
							MAF	P
TNFAIP3	6:138199789	rs767086093	synonymous	c.T1207C	p.L403L	0.002 (1/398)	0.000009 (1/113764)	**0.007**
	6:138202378	rs5029956	synonymous	c.C2295T	p.P765P	0.002 (1/398)	0.0002 (23/124472)	0.07
	6:138199959	rs369402265	synonymous	c.G1377A	p.S459S	0.002 (1/398)	0.00003 (3/113120)	**0.01**
	6:138200335	rs190199205	missense	c.G1753A	p.D585N	0.002 (1/398)	0.00003 (3/113572)	**0.01**
	6:138201240	rs142253225	missense	c.A1939C	p.T647P	0.01 (4/398)	0.002 (299/129188)	**0.01**
	6:138196151	rs140424499	synonymous	c.G465A	p.T155T	0.007 (3/398)	0.0006 (79/128878)	**0.002**
	6:138196022	n.a.	missense	c.G336C	p.M112I	0.002 (1/398)	n.a.	/
TRAF3IP2	6:111912641	rs139282334	missense	c.C649A	p.P217T	0.005 (2/398)	0.009 (1035/118574)	0.59
	6:111896863	rs139767840	missense	c.A1184G	p.N395S	0.007 (3/398)	0.003 (450/128614)	0.16
	6:111901465	rs146226365	missense	c.C957A	p.S319R	0.002 (1/398)	0.001 (151/128638)	0.37
	6:111887758	rs200213270	synonymous	c.C1362T	p.T454T	0.002 (1/398)	0.0002 (21/129162)	0.06
	6:111912880	n.a.	missense	c.G410A	p.R137H	0.002 (1/398)	n.a.	/
STAT4	2:191929596	rs61756200	missense	c.G719A	p.R240Q	0.002 (1/398)	0.003 (323/128766)	1
	2:191937906	rs140675301	missense	c.A383T	p.E128V	0.007 (3/398)	0.002 (203/129036)	**0.026**
	2:191927613	rs140221173	synonymous	c.G816A	p.L272L	0.002 (1/398)	0.0006 (74/128902)	0.20
	2:191903973	n.a	synonymous	c.T1386C	p.N462N	0.002 (1/398)	n.a.	/
	2:191929548	n.a.	missense	c.A767T	p.D256V	0.002 (1/398)	n.a.	/
	2:191927640	n.a.	synonymous	c.A789T	p.T263T	0.002 (1/398)	n.a.	/
HCP5	6: 31,431,149	n.a.	non-coding	n.T102C	n.a.	0.002 (1/398)	n.a.	/

MAF: minor allele frequency; hgvs: Human Genome Variation Society; n.a.: not available; GnomAD: Genome Aggregation Database P: p-value.
RefSeq accession numbers: *TNFAIP3,*NM_001270507*; TRAF3IP2* NM_001164381*; STAT4,* NM_001243853; *HCP5,* NR_040662.
Genomic position was based on GRCh37. P-values were obtained using Fisher's exact test.

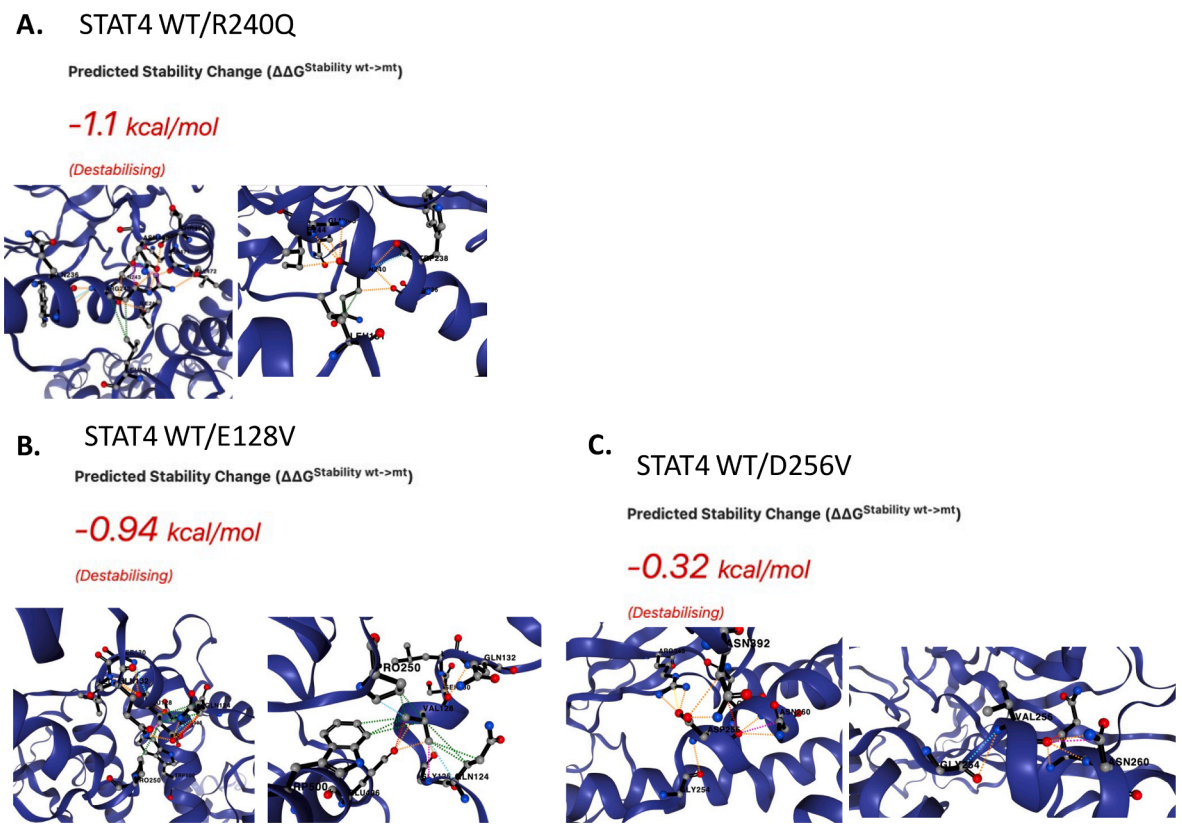


Fig. 1. Predicted 3D model of STAT4 protein. Predicted structure o f STAT4 showing the change in secondary interactions and protein stability resulting from p. R240Q (A), p.E128V (B) and p.D256V (C). The predicted models are based on DDMut tool.

protein with missense variants do not show structure stability alterations, compared to the wild-type structure. The prediction of the effect on *TNFAIP3* variants could not be performed because the structure of this protein is not yet fully characterized.

Lastly, since protein phosphorylation is one of the most common post-transcriptional modifications and it plays an important role in intracellular signal transduction, in which *TNFAIP3, TRAF3IP2* and *STAT4* are involved, we investigated whether the identified missense variants could affect this mechanism. Among the variants located in *TNFAIP3* gene, c.A1939C causes the substitution of Thr 647 with a proline, precisely at a phosphorylation site. Moreover, this variant is in a region enriched in phosphorylation sites, and the amino acid change also affects the affinity of specific kinases for surrounding phosphorylation sites, such as Ser 645 or Ser 649 (Fig. 2). For *TRAF3IP2* gene, we observed that two of the identified variants (p.P217T and p.N395S) cause an amino acid change near to a phosphorylation site. In particular, the prediction analysis showed that the substitution of proline with threonine at position 217 could alter the affinity of specific kinases both with Thr 219 and with Ser 220 (Fig. 3). On the contrary, p.N395S although being close to the phosphorylation site in position 390, it does not seem to alter its recognition capacity by the kinases. For *STAT4* gene, only p.E128V variant is located near a known phosphorylation site. The prediction analysis showed that this amino acid substitution changes the specificity of individual kinases in recognizing and phosphorylating the serine at position 130 (Fig. 3).

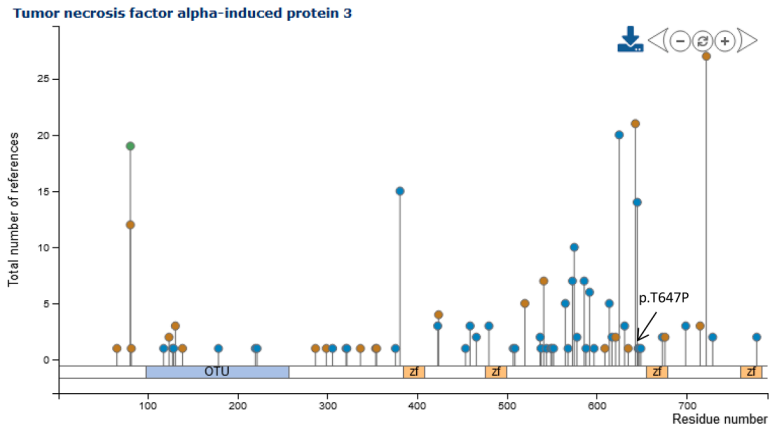
4. Discussion

It is nowadays recognised that common genetic variants can explain only a fraction of the heritability of complex traits (Manolio et al., 2009) and that the missing heritability can be searched, among other factors, including rare genome variations. Indeed, recent data have shown that rare variants can explain a greater part of the variability of some complex phenotypes, such as weight and height (Wainschtein et al., 2022) or even the variability of complex disorders susceptibility and clinical phenotypes (Sazonovs et al., 2022). In autoimmune disorders, few studies have started to investigate the possible role of rare variants in commonly known risk loci. For example, one study identified rare variants in *IL2RA* and *IL2RB* as associated with Rheumatoid Arthritis (Diogo et al., 2013). In addition, it is worth to highlight that rare genetic variants are likely to be more deleterious than common variants and their identification in association with complex disorders could allow to identify domains whose integrity is essential for the physiological function of proteins, and therefore useful also to understand the molecular mechanisms underlying the diseases.

Following the previous considerations, in the current study we explored the rare protein-coding variants in five candidate genes already investigated in our previous studies for common risk polymorphisms. In an Italian SLE cohort, we identified six rare and seven ultra-rare (MAF < 0.001) known variants and six novel variants, absent in our healty controls group and in the GnomAD database.

TNFAIP3 (Tumour Necrosis Factor Alpha Induced Protein 3) and *TRAF3IP2* (Tumour Necrosis Factor Receptor-Associated 3 Interacting Protein 2) encode for two proteins of NF-κB signalling pathway. NF-κB signaling regulates multiple aspects of innate and adaptive immune functions, inducing the expression of various pro-inflammatory genes and participating in inflammasome regulation and immune cell activation and differentiation (Liu et al., 2017). Numerous association studies have highlighted genetic variants located in the genes involved in this pathway which seem to increase the risk of developing SLE and other autoimmune diseases, including *TRAF3IP2* and *TNFAIP3* (Barnabei et al., 2021). TNFAIP3, also known as A20, is a ubiquitin-editing enzyme and a negative regulator of the NF-κB signaling pathway. Genetic variants of this gene, in particular nonsense, frameshift, and splice site

A.



B.

kinase	kinase group	log_2(score)	site percentile	percentile rank
PRP4	CMGC	3.596	97.073 %	1
GSK3B	CMGC	1.317	91.908 %	2
NLK	CMGC	2.710	90.870 %	3
PKG1	AGC	-0.259	88.692 %	4
PKG2	AGC	0.698	88.632 %	5

kinase	kinase group	log_2(score)	site percentile	percentile rank
PRP4	CMGC	3.191	96.419 %	1
KIS	Other	3.820	93.369 %	2
CDK7	CMGC	3.994	92.645 %	3
GSK3B	CMGC	1.160	89.098 %	4
NLK	CMGC	2.362	87.049 %	5

C.

kinase	kinase group	log_2(score)	site percentile	percentile rank
CHAK1	Alpha	1.873	97.816 %	1
TSSK2	CAMK	2.423	96.820 %	2
SNRK	CAMK	1.049	96.490 %	3
IRE1	Other	0.931	94.816 %	4
STK33	CAMK	0.437	94.752 %	5

kinase	kinase group	log_2(score)	site percentile	percentile rank
CK1E	CK1	2.079	98.933 %	1
CHAK1	Alpha	2.331	98.819 %	2
MPSK1	Other	2.318	97.433 %	3
IRE1	Other	1.508	97.096 %	4
CHAK2	Alpha	1.487	96.407 %	5

Fig. 2. Bioinformatic prediction of phosphorylation sites in TNFAIP3 protein (A); bioinformatic prediction of main kinases that phosphorylate Ser 645 (B) and Ser 649 (C) of TNFAIP3, in protein wild-type and p.T647P variant.

mutations, could also cause monogenic forms of early onset auto-inflammatory diseases, due to A20 haploinsufficiency (Zhang et al., 2021). The effect of missense common variants in this gene are instead more difficult to evaluate and have rarely been classified as pathogenic. In this study, we identified three different missense variants of the *TNFAIP3* gene, none of which was observed in the analysed control healthy group. In particular, the rare variant c.A1939C was observed in four unrelated SLE patients. This variant resides in the region with E3 ubiquitin ligase activity, involved in recruiting adaptor proteins, and an *in vitro* functional study demonstrated that significantly reduces the inhibitory effect of A20 on NF-κB activity. Moreover, this variant is located in an enriched region in phosphorylation sites and predictive tools suggest that could affects the affinity of specific kinases. A study demonstrated that the genetic variant C243Y eliminates almost entirely A20 phosphorylation and causes a severe inflammatory disease in both mice and humans, comparable to the phenotype of a subject with a truncated A20 protein (Zammit et al., 2019). At the light of the role of A20 in the suppression of inflammation, immune cell proliferation and the immune response, the reduction of its activity and the following excess activation of NF-κB signalling could represent a strong risk factor for the onset of autoimmune diseases such as SLE (Kadowaki et al., 2021).

TRAF3IP2 encodes Act1, a positive signalling adapter in IL-17-mediated cellular immune responses and a negative regulator of adaptive immunity. The association between polymorphisms of the *TRAF3IP2* gene and susceptibility to SLE has been described for the first time in a study on Italian population (Perricone et al., 2013). Here, our results did not show differences between frequencies observed in SLE patients and those reported for the reference population in the GnomAD database with regard to the four known genetic variants. Instead, we identified a novel variant (c.G410A) located in the region that mediates interaction with TRAF6. This region represents a critical and functional domain, because TRAF6 is a key signalling molecule in B-cell activation through CD40/TLR and the lack of its regulation could lead to aberrant activation of B cells. Thus, although *in silico* analyses classify this *TRAF3IP2* variant as benign, it may have a minor effect on TRAF6 interaction and represent a risk factor for SLE development. p.P217T.

STAT4 encodes a cytoplasmic transcription factor, belonging to the STAT protein family, essential for mediating responses to IL12 in lymphocytes and regulating T-helper cells differentiation. The gene is a confirmed genetic risk factor for many autoimmune diseases, such as Sjogren's syndrome, rheumatoid arthritis and SLE. In *STAT4* gene, we identified one rare missense variant c.A383T, in three unrelated SLE patients, with a significantly higher frequency compared to that reported in GnomAD database. This genetic variant, already described as associated with rheumatoid arthritis (Saevarsdottir et al., 2022), leads to an amino acid change from negatively charged glutamic acid to non-polar hydrophobic valine: This change falls in a loop connecting the N-terminal domain, important for tetramer formation at DNA binding sites and the coiled coil domain, which provides a carbonised hydrophilic surface that binds to regulatory factors. Also, the novel variant c.A767T identified in one of ours SLE patients is located in the coiled coil domain and it was predicted as damaging by six prediction tools. The amino acid sequence and secondary structure of these domains are highly conserved between species, and they are probably necessary for full transcriptional activation of many of its target genes. Furthermore, all three rare missense variants identified in SLE patients cause an amino acid residue change that decreases the thermodynamic stability of the protein, altering hydrophobic interactions and ionic bonds, and thus favours misfolding.

Our study has some limitations: at present, rare variants identified have limited power, indeed larger sizes cohorts and advanced variant interpretation algorithms are needed to confirm their effect on SLE; the interpretation of variants of uncertain significance (VUS) remains a

A.

kinase	kinase group	log_2(score)	site percentile	percentile rank
ULK1	Other	1.961	99.625 %	1
NEK7	Other	2.363	99.301 %	2
TGFBR2	TKL	2.710	99.196 %	3
DSTYK	TKL	2.854	99.079 %	4
NEK5	Other	1.359	98.958 %	5

kinase	kinase group	log_2(score)	site percentile	percentile rank
ULK1	Other	1.578	99.272 %	1
RAF1 [!]	TKL	1.670	97.824 %	2
CHAK2	Alpha	1.724	97.370 %	3
ULK2	Other	1.094	97.201 %	4
BRAF [!]	TKL	0.370	96.281 %	5

B.

kinase	kinase group	log_2(score)	site percentile	percentile rank
VRK1	CK1	1.712	99.791 %	1
MEK2 [!]	STE	-0.029	98.802 %	2
VRK2	CK1	0.577	97.982 %	3
NEK1	Other	0.318	96.735 %	4
MPSK1	Other	1.878	96.113 %	5

kinase	kinase group	log_2(score)	site percentile	percentile rank
MEK2 [!]	STE	0.803	99.701 %	1
VRK1	CK1	1.166	99.474 %	2
NEK1	Other	0.888	98.382 %	3
VRK2	CK1	0.416	97.039 %	4
NEK4	Other	0.406	96.706 %	5
MPSK1	Other	-0.615	76.733 %	67

C.

kinase	kinase group	log_2(score)	site percentile	percentile rank
TTBK1	CK1	1.858	99.232 %	1
CAMKK2	Other	1.137	98.923 %	2
CAMKK1	Other	0.654	97.882 %	3
PLK4	Other	0.915	97.150 %	4
ZAK	TKL	0.789	97.141 %	5

kinase	kinase group	log_2(score)	site percentile	percentile rank
TTBK1	CK1	2.391	99.734 %	1
IRAK1	TKL	1.589	99.027 %	2
RIPK2	TKL	0.581	98.031 %	3
CAMKK1	Other	0.641	97.839 %	4
CAMKK2	Other	0.545	97.614 %	5

Fig. 3. Bioinformatic prediction of main kinases that phosphorylate Ser 130 of STAT4, in protein wild-type and p.E128V variant (A); Bioinformatic prediction of main kinases that phosphorylate Thr 219 (B) and Ser 220 (C) of TRAF3IP2, in protein wild-type and p.P217T variant.

challenge; we evaluated only coding or splice-altering variants, but expanding the analysis to include noncoding variants would likely increase the burden of rare variants associated with the phenotype (Ribeiro and Delaneau, 2023); the absence of specific consensus and biological material does not allow us to confirm by western blotting whether the presence of these rare SNVs modifies the phosphorylation state on the proteins.

Concluding, we have observed that even in genes whose common variability is associated with SLE susceptibility, it is possible to identify rare variants that could have a strong effect in the disease development and could therefore allow a better understanding of the functional domain involved. In last decades, GWAS and association studies for candidate genes have identified thousands of polymorphisms underlying complex diseases, including SLE. However, common variants have often modest effects on the disease risk and limited clinical utility. Rare variants located in coding regions may allow a direct link between altered genes function and specific phenotypes. Understanding the role of rare variants in common diseases could be useful to explain disease etiogenesis and to develop targeted drug.

Funding

This study was supported by MUR-PNRR M4-C2-I1.3 PE6 project PE00000019 Heal Italia [CUP: E83C22004670001] to G.N. and European Union - NextGenerationEU: National Center for Gene Therapy and Drugs based on RNA Technology, CN3 - Spoke 7 (code: CN00000041; PNRR - Mission 4, Component 2; Investment 1.4) to M.H.C and G.N.

CRediT authorship contribution statement

Andrea Latini: Writing – original draft, Methodology, Investigation, Data curation. **Paola Borgiani:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Giada De Benedittis:** Writing – original draft, Investigation, Data curation. **Cinzia Ciccacci:** Writing – original draft, Data curation. **Lucia Novelli:** Resources. **Gerardo Pepe:** Investigation. **Manuela Helmer-Citterich:** Writing – review & editing, Investigation, Funding acquisition. **Isabella Baldini:** Formal analysis. **Carlo Perricone:** Resources. **Fulvia Ceccarelli:** Resources. **Fabrizio Conti:** Resources. **Generoso Ianniciello:** Formal analysis. **Juan Caceres:** Formal analysis. **Riccardo Ottalevi:** Formal analysis, Data curation. **Mattia Capulli:** Formal analysis. **Giuseppe Novelli:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2024.148279.

References

Barnabei, L., Laplantine, E., Mbongo, W., et al., 2021. NF-κB: At the borders of autoimmunity and inflammation. Front. Immunol. 12, 716469.

Ciccacci, C., Perricone, C., Ceccarelli, F., et al., 2014. A multilocus genetic study in a cohort of Italian SLE patients confirms the association with STAT4 gene and describes a new association with HCP5 gene. PLoS One 9, e111991.

Ciccacci, C., Latini, A., Perricone, C., et al., 2019. TNFAIP3 gene polymorphisms in three common autoimmune diseases: Systemic Lupus Erythematosus, rheumatoid arthritis, and primary Sjogren syndrome-association with disease susceptibility and clinical phenotypes in Italian patients. J. Immunol. Res. 2019, 6728694.

Demirkaya, E., Sahin, S., Romano, M., et al., 2020. New horizons in the genetic etiology of Systemic Lupus Erythematosus and lupus-like disease: Monogenic lupus and beyond. J. Clin. Med. 9, 712.

Diogo, D., Kurreeman, F., Stahl, E.A., et al., 2013. Rare, lowfrequency, and common variants in the protein-coding sequence of biological candidate genes from GWASs contribute to risk of rheumatoid arthritis. Am. J. Hum. Genet. 92, 15–27.

Fan, Y., Tao, J.H., Zhang, L.P., Li, L.H., Ye, D.Q., 2011 Apr. The association between BANK1 and TNFAIP3 gene polymorphisms and systemic lupus erythematosus: a meta-analysis. Int. J. Immunogenet. 38 (2), 151–159. https://doi.org/10.1111/j.1744-313X.2010.00990.x. Epub 2011 Jan 5 PMID: 21208380.

Fathalla, B.M., Alsarhan, A., Afzal, S., et al., 2021. The genomic landscape of pediatric rheumatology disorders in the Middle East. Hum. Mutat. 42, e1–e14.

Hornbeck, P.V., Zhang, B., Murray, B., et al., 2015. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res. 43, D512–D520.

Johnson, J.L., Yaron, T.M., Huntsman, E.M., et al., 2023. An atlas of substrate specificities for the human serine/threonine kinome. Nature 613, 759–766.

Kadowaki, S., Hashimoto, K., Nishimura, T., et al., 2021. Functional analysis of novel A20 variants in patients with atypical inflammatory diseases. Arthritis Res. Ther. 23, 52.

Karrar, S., Cunninghame Graham, D.S., 2018. Abnormal B cell development in systemic lupus erythematosus: What the genetics tell us. Arthritis Rheumatol. 70, 496–507.

Kwon, Y.C., Chun, S., Kim, K., et al., 2019. Update on the genetics of systemic lupus erythematosus: Genome-wide association studies and beyond. Cells 8, 1180.

Liu, T., Zhang, L., Joo, D., et al., 2017. NF-κB signaling in inflammation. Signal. Transduct. Target Ther. 2, 17023.

Manolio, T.A., Collins, F.S., Cox, N.J., et al., 2009. Finding the missing heritability of complex diseases. Nature 461, 747–753.

Martínez-Bueno, M., Alarcón-Riquelme, M.E., 2019. Exploring impact of rare variation in systemic lupus erythematosus by a genome wide imputation approach. Front. Immunol. 10, 258.

Mulhern, C.M., Hong, Y., Omoyinmi, E., et al., 2019. Janus kinase 1/2 inhibition for the treatment of autoinflammation associated with heterozygous TNFAIP3 mutation. J. Allergy Clin. Immunol. 144, 863–866.

Pan, L., Lu, M.P., Wang, J.H., et al., 2020. Immunological pathogenesis and treatment of systemic lupus erythematosus. World J. Pediatr. 16, 19–30.

Perricone, C., Ciccacci, C., Ceccarelli, F., et al., 2013. TRAF3IP2 gene and systemic lupus erythematosus: association with disease susceptibility and pericarditis development. Immunogenetics 65, 703–709.

Ramos, P.S., Brown, E.E., Kimberly, R.P., et al., 2010. Genetic factors predisposing to systemic lupus erythematosus and lupus nephritis. Semin. Nephrol. 30, 164–176.

Ribeiro, D., Delaneau, O., 2023. Non-coding rare variant associations with blood traits on 166,740 UK Biobank genomes. bioRxiv. https://doi.org/10.1101/2023.12.01.569422.

Saevarsdottir, S., Stefansdottir, L., Sulem, P., et al., 2022. Multiomics analysis of rheumatoid arthritis yields sequence variants that have large effects on risk of the seropositive subset. Ann. Rheum. Dis. 81, 1085–1095.

Sazonovs, A., Stevens, C.R., Venkataraman, G.R., et al., 2022. Large-scale sequencing identifies multiple genes and rare variants associated with Crohn's disease susceptibility. Nat. Genet. 54, 1275–1283.

Shancui-Zheng, Jinping-Zhang, Guoyuan-Lu, Liu L, Zhiyong-Deng. Polymorphism in STAT4 Increase the Risk of Systemic Lupus Erythematosus: An Updated Meta-Analysis. Int J Rheumatol. 2022 Apr 22;2022:5565057. doi: 10.1155/2022/5565057. PMID: 35493285; PMCID: PMC9054488.

Wainschtein, P., Jain, D., Zheng, Z., et al., 2022. Visscher Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. Nat. Genet. 54, 263–273.

Yuan Y, Wang X, Ren L, Kong Y, Bai J, Yan Y. Associations between interleukin-10 gene polymorphisms and systemic lupus erythematosus risk: a meta-analysis with trial sequential analysis. Clin Exp Rheumatol. 2019 Mar-Apr;37(2):242-253. Epub 2018 Aug 29. PMID: 30183604.

Zammit, N.W., Walters, S.N., Seeberger, K.L., et al., 2019. A20 as an immune tolerance factor can determine islet transplant outcomes. JCI Insight 4, e131028.

Zhang, D., Su, G., Zhou, Z., et al., 2021. Clinical characteristics and genetic analysis of A20 haploinsufficiency. Pediatr. Rheumatol. Online J. 19, 75.

Zhou, Y., Pan, Q., Pires, D.E.V., et al., 2023. DDMut: predicting effects of mutations on protein stability using deep learning. Nucleic Acids Res. 51, W122–W128.