Received: 5 July 2021 | Revised: 19 January 2022 | Accepted: 23 January 2022

DOI: 10.1111/hae.14503

Haemophilia WILEY

ORIGINAL ARTICLE

Genetic variants at the chromosomal region 2q21.3 underlying inhibitor development in patients with severe haemophilia A

Silvia Spena[1] | Andrea Cairo[1] | Emanuela Pappalardo[2] | Marcin M. Gorski[2] | Isabella Garagiola[1] | Shermarke Hassan[2,3] | Roberta Gualtierotti[1,2] | Flora Peyvandi[1,2]

[1] Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, and Luigi Villa Foundation, Milan, Italy

[2] Department of Pathophysiology and Transplantation, Università degli Studi di Milano, Milan, Italy

[3] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands

Correspondence
Flora Peyvandi, Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, and Luigi Villa Foundation; Department of Pathophysiology and Transplantation, Università degli Studi di Milano, via Pace 9, 20122 Milan, Italy.
Email: flora.peyvandi@unimi.it

Funding information
Italian Ministry of Health: Ricerca Finalizzata, Grant/Award Number: RF-2016-02363246; Bando Ricerca Corrente 2020

[Correction added on 17 May 2022, after first online publication: CRUI-CARE funding statement has been added.]

Abstract

Introduction: Inhibitor development affects about 30% of patients with severe haemophilia A (HA) and results from different environmental and genetic risk factors. Previously, we identified the missense variant rs3754689 in the *LCT* gene linked with this predisposition. Since rs3754689 variant is benign and is located in a conserved haplotype region, we hypothesized that the association signal captured by this variant is located in coinherited, neighbouring genes.

Aim: To identify novel genetic risk factors associated with inhibitor development in coding regions of *R3HDM1*, *UBXN4*, *CXCR4*, *MCM6*, *DARS* and *miR128-1* genes.

Methods: Targeted sequencing was performed in 246 severe HA patients (72 with and 174 without inhibitor): 181 previously and 65 newly enrolled.

Results: Forty-one common and 152 rare variants passed the quality control. Logistic regression analysis of common variants identified rs3754689 and four additional variants ($.011 < P < .047$; FDR ranging .2-.38). Logistic regression analysis performed only in the 220 Italian patients showed similar results ($.004 < P < .05$; FDR ranging .12-.22). Three of these variants (rs3213892 and rs3816155 in the *LCT* intron 13 and rs961360 in the *R3HDM1* intron10-exon11 junction) may affect the expression of *UBXN4* and *R3HDM1*, respectively. Rare variants did not show association with inhibitor development. Identified variants were not replicated in the multi-ethnic SIPPET cohort of 230 severe HA patients.

Conclusion: Due to the limited sample size that may be responsible of the high FDR values, we could not confirm with certainty the analysed association. Further evaluation of the expression levels of analysed genes will confirm or not their role in inhibitor development.

KEYWORDS
haemophilia A, high-throughput DNA sequencing, neutralizing antibodies, risk factors, single nucleotide polymorphism

1 | INTRODUCTION

Haemophilia A (HA, OMIM #306700) is a bleeding disorder characterized by the deficiency of coagulation factor VIII (FVIII). Patients with FVIII coagulant activity in plasma less than 1% have severe disease and suffer from frequent and spontaneous bleeding episodes (up to 30/year) mainly at joints and muscles.[1,2] Treatment of bleeding episodes consists of on-demand or prophylactic replacement therapy with plasma-derived or recombinant FVIII products. About 30% of patients with severe HA develop, mainly in the first 20 days of therapy, anti-FVIII neutralizing alloantibodies (i.e. inhibitors) that decrease the efficacy of the replacement therapy.[3] In these cases, haemostasis can only be achieved using FVIII-bypassing agents (recombinant activated FVII and activated prothrombin complex concentrate)[4] or a bispecific antibody mimicking the function of activated FVIII (emicizumab).[5] Since both alternative treatments are very costly,[6] understanding the aetiology of inhibitor development is an urgent need. Inhibitor development is a multifactorial complication resulting from environmental and genetic factors.[7,8] Null mutations in the FVIII coding gene (*F8*), leading to the complete lack of FVIII activity, have been described as the strongest genetic risk factors for inhibitor development.[8–10] In addition, polymorphisms in antigen presenting HLA class II molecules[11–14] and in immune regulatory *IL-1*, *IL-2*, *IL-10*, *CTLA4*, TNFa,[13–19] *HMOX1*, *MAPK9* and *CD32* genes[20] have been reported to be associated with this complication although their role is still under debate due to the poor reproducibility of the genotyping results, the small replication cohorts, and the different genetic background of the analysed populations.

Results from our recent study, point towards the presence of novel risk factor(s) for inhibitor development in a highly conserved haplotype region surrounding the *LCT* gene on chromosome 2q21.3[21] were six genes (*R3HDM1*, *UBXN4*, *CXCR4*, *MCM6*, *DARS* and *miR128-1*), previously linked with autoimmune disorders and pathways of the immune response and therefore potential clinically relevant targets, are localized.

In this study, the targeted next generation sequencing (NGS) of the coding regions of these genes located in the 586-kb region surrounding the *LCT* gene has been performed in a cohort of 246 patients with severe HA to identify novel variants/genes associated with inhibitor formation.

2 | MATERIALS AND METHODS

2.1 | Patient cohorts

Two hundred and forty-six unrelated patients with severe haemophilia A (FVIII:C < 1%) regularly followed at the Angelo Bianchi Bonomi Hemophilia and Thrombosis Center, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Ca' Granda, Ospedale Maggiore Policlinico (Milan, Italy) were enrolled (181 patients are from the replication cohort reported in Gorski et al[21] and 65 are new patients). All patients received at least 50 infusions of FVIII during on-demand or prophylactic treatment. Of 246 patients, 72 experienced inhibitor development (cases) and the remaining 174 not (controls). The FVIII inhibitor was measured in patients' plasma every 3–5 exposure days in children who started treatment, once a year or after a lack of response to the treatment or before any surgery in previously treated patients by means of the Bethesda assay with the Nijmegen modification.[22] Inhibitor was defined as positive with two measured values > .5 Bethesda units (BU)/mL.

The replication study was conducted in an independent population of 230 patients with severe HA enrolled as part of the international SIPPET (Survey of Inhibitors in Plasma-Product Exposed Toddlers) study,[23] of which 70 were inhibitor-positive and 160 inhibitor-negative.

Written informed consent was obtained from parents of underage patients and adult patients. The study was approved by the Ethics Committee of the Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (Milan, Italy) and was carried out in accordance with the Declaration of Helsinki.

2.2 | NGS and data analysis

Genomic DNA was isolated from peripheral blood samples by the standard salting-out method.[24] Coding regions including 10 bp of intron-exon boundaries and 100 bp of 5′ and 3′ untranslated regions of *R3HDM1*, *UBXN4*, *CXCR4*, *MCM6* and *DARS* genes were amplified using TruSeq Custom Amplicon strategy (Illumina) (Table S1). Intronic single nucleotide polymorphisms (SNP) already reported in linkage disequilibrium (LD) with the missense variant rs3754689[21] and the intronic region of the *R3HDM1* gene encoding the *miR128-1* were also included in the panel design producing a total target size of 19,573 bp (Table S1).

Paired-end sequencing was conducted on a MiSeq sequencing platform (Illumina). Obtained reads were mapped to the reference human genome hg19 using Burrows–Wheeler aligner (BWA-MEM algorithm). Variant calling and filtering was performed according to the guidelines reported by the Broad institute (https://software.broadinstitute.org/gatk/best-practices/). All 246 gVCF files were merged and cumulative VCF file was annotated by ANNOVAR.[25] A second quality control step was performed by KGGSEquation (http://grass.cgs.hku.hk/limx/kggseq/): variants with low quality (Phred score Q < 30), Hardy-Weinberg equilibrium P < 1.0E-04 and genotypes with low depth (< 10) were excluded. VCFtools was used to select common/low-frequency (minor allele frequency, MAF ≥ 1%) and rare (MAF < 1%) variants. Variants with a call rate of < 80% were excluded from the analysis.

2.3 | Statistical analysis

Case control association test for common/low-frequency variants was performed by logistic regression model using Plink 1.07.[26] Results were reported as odds ratios (ORs) with 95% confidence intervals

(CI). Adjustment for multiple testing was performed using the Benjamini and Hochberg's (FDR_BH) method. For each statistical comparison, a standard unadjusted p-value was calculated, as well a multiplicity adjusted p-value. LD of common variants was evaluated by squared correlation of allele frequencies (r2). For rare variants, Burden Test and Sequence Kernel Association Test were performed using PlinkSEquation (https://atgu.MGH.Harvard.edu/plinkseq). Cumulative gene-based association tests were performed on three different subgroups of rare variants: all identified rare variants; only potential damaging variants (missense, nonsense, splicing variants and in-frame deletions/insertions); variants with Combined Annotation Dependent Depletion (CADD) score higher than 20 (see in-silico analyses).

2.4 | Replication

The replication study was performed using TaqMan SNP genotyping assays C_12054686_20 (rs1050115), C_27505260_10 (rs3816155), C_2104738_10 (rs3754689), C_15793430_10 (rs3087343) (Thermo Fisher Scientific) on a StepOnePlus RealTime Polymerase Chain Reaction system (Applied Biosystems) and direct sequencing of rs3213892 with the BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) on an ABI PRISM 3130 Genetic Analyzer (Applied Biosystems).

2.5 | In silico analyses

Prediction of the pathogenicity of variants was estimated by CADD tool (https://cadd.gs.washington.edu/). The impact of variants on exonic splice enhancer (ESE) motifs and on splicing was evaluated with the ESE finder 3.0 software (rulai.cshl.edu/) and the Neural Network Splice Site prediction program (https://www.fruitfly.org/seq_tools/splice.html), respectively. To assess localization of variants in functional elements in the human genome, Encyclopedia of DNA Elements (ENCODE) (https://www.encodeproject.org/) and Enhancer Atlas (http://enhanceratlas.org/) databases were examined.

3 | RESULTS

3.1 | Identification of variants associated to inhibitor development

Target sequencing has been performed in the discovery cohort of 246 patients with severe HA (72 inhibitor-positive and 174 inhibitor-negative). Data analysis revealed a total of 365 variants across all samples. After filtering, 193 variants (41 common; MAF ≥ 1% and 152 rare; MAF < 1%) passed the quality controls and included one in-frame deletion, two splicing, four nonsense, 42 synonymous, 54 missense and 90 intronic (Tables S2 and S3). One hundred and four variants are listed in dbSNP(150) and 89 are not reported. Case-control association analysis for common variants was carried out by logistic regression.

Association with inhibitor development was found for five common variants (Figure 1, Table 1): the missense variant rs3754689 (p.Val219Ile) in the *LCT* gene ($P = .047$; OR = .65; 95% CI = .43-.99), previously identified in our laboratory,[21] and four additional common variants: rs1050115 ($P = .011$; OR = .56; 95% CI = .35-.87) in the *UBXN4* gene, rs3213892 ($P = .015$; OR = .58; 95% CI = .37-.90) and rs3816155 ($P = .023$; OR = .60; 95% CI = .39-.93) in the *LCT* gene, and rs3087343 ($P = .014$; OR = .56; 95% CI = .36-.89) in the *MCM6* gene (Table 1). To overcome the issue of genetic factors that may lead to confounding, logistic regression analysis was further performed in the sub-cohort of 220 patients with Italian ethnicity (i.e. the discovery cohort without the 26 non-Italian patients) (Supplementary Table 1). Similar results were obtained for the five common variants identified in *UBXN4*, *LCT* and *MCM6* (Table 2) and additional three common variants with a nominal $P < .05$: rs961360 ($P = .030$; OR = .59; 95% CI = .36-.95) and rs12466487 ($P = .030$; OR = .57; 95% CI = .34-.95) in *R3HDM1* and rs371309040 ($P = .028$; OR = .59; 95% CI = .37-.94) in *DARS1* (Table 2).

Additional variants with no statistical significance were found in *R3HDM1* (41 variants), *DARS* (23), *CXCR4* (13), *UBXN4* (18), *LCT* (60) and *MCM6* (33) and no variants in the *miR128-1* were identified (see Tables S2 and S3 for a comprehensive list of common and rare variants, respectively). The top variants are potentially protective against inhibitor development (.48 < OR < .64) despite the adjusted p-values were high (FDR ranging .12-.22) (Table 2). LD analysis evidenced that these variants were strongly associated with each other.

To test for independency of signal, a conditional logistic regression analysis was carried out for the most significant variant (rs1050115). All other variants lose the statistical significance ($P > .05$) (Tables 1 and 2).

Concerning rare variants, cumulative tests were carried out to evaluate the association between our target genes and the inhibitor development; none of them resulted significantly associated (Table S4).

3.2 | Replication of variants associated to inhibitor development

The replication study of five identified variants performed in the multiethnic SIPPET cohort of 230 patients with severe HA (70 inhibitor-positive and 160 inhibitor-negative) showed no protection against inhibitor development (1.02 < OR < 1.19) (Table 1).

3.3 | In silico analysis of identified variants

To evaluate the potential pathogenic role of the identified variants, in silico analyses were performed. Due to the exonic localization of rs961360 (*R3HDM1*), rs3754689 (*LCT*) and rs1050115 (*UBXN4*) variants, the potential impact of nucleotide substitutions on ESE motifs was assessed. No ESE motifs were predicted in the region encompassing rs961360 (*R3HDM1*) and rs3754689 (*LCT*). Despite prediction of a binding site for the ESE-binding SRSF1 protein in the region encom-

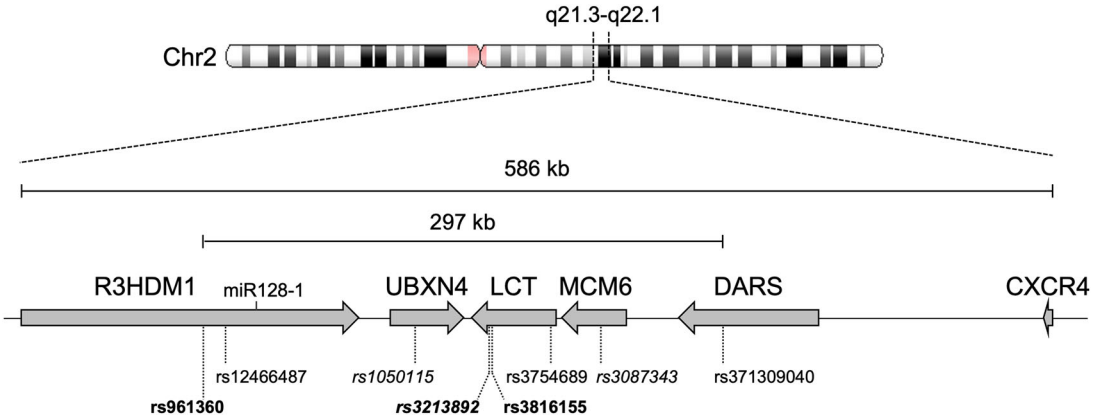

FIGURE 1 Localization of analysed target genes. The chromosome bands q21.3-q22.1 are shown in the ideogram of chromosome 2 (upper panel). A schematic representation of analysed target genes is drawn to scale (lower panel). The arrows indicate the transcriptional orientation of each gene. The size of the analysed cluster and of identified variants are indicated. The localization of *miR128-1* and of the identified variants is shown: variants previously found in LD[21] with the rs3754689 in *LCT* are in italic, those with putative role in *R3HDM1* and *UBXN4* gene expression are in bold

passing rs1050115 (*UBXN4*), no disruption of this putative ESE motif was predicted.

Cryptic splice site activation was also assessed for all exonic and intronic variants. No activation was predicted in all cases with the exception of the rs961360 variant localized at the first nucleotide of the *R3HDM1* exon 11 for which no alteration of the acceptor splice site (3′ss) (score from .98 to .99; score range 0–1) but activation of a new strong (score .89) donor splice site were predicted (Figure 2A).

The localization of all identified variants in regulatory regions was further evaluated. This additional analysis evidenced the localization of both rs3213892 and rs3816155 variants (*LCT*) in predicted cis-acting transcriptional regulatory elements: a 400-bp promoter and a 200-bp enhancer spanning the *LCT* exon14-intron13 junction and including the first 15 nucleotides of the exon 14 (Figure 2B). Moreover, a deep analysis with Enhancer Atlas evidenced the enhancer as the potential regulatory element of the target *UBXN4* gene (Figure 2B).

4 | DISCUSSION

Inhibitor formation is a multifactorial complication with a heterogeneous genetic component. Although the type of mutation in *F8* gene is ascertained as the strongest risk factor in patients with severe HA, the mutation alone does not completely explain all the genetic components and increasing evidence supports a combined role of several other genetic factors. In this scenario, the present work continues our recent whole exome sequencing study that identified the missense variant p.Val219Ile (rs3754689) in the *LCT* (lactase-phlorizin hydrolase) gene associated with inhibitor development.[21] Since this variant, a polymorphism with a predicted benign effect,[27] is located in a conserved haplotype block surrounding the *LCT* gene at 2q21.3, we hypothesized that the association signal, captured by the rs3754689 variant, was located in coinherited genes. Five neighbouring genes (*R3HDM1*, *UBXN4*, *MCM6*, *DARS*, *CXCR4*) that we selected for target sequencing are directly or indirectly involved in autoimmune diseases and/or in pathways of immune response. A reduced expression of *R3HDM1* gene, coding a poorly characterized RNA binding protein, was recently found in patients with celiac disease.[28] Moreover, the expression of *miR128-1*, located in the *R3HDM1* intron 18 (Figure 1), has been shown to be upregulated in naïve CD4+ T cells from patients with multiple sclerosis[29] and dysregulated in CD4+ lymphocytes of patients with systemic sclerosis,[30] both autoimmune diseases. The UBXN4 protein is a member of the UBXN family; while UBXN1 is involved in innate immunity,[31] UBXN4 has been demonstrated to function in endoplasmic reticulum–associated protein degradation (ERAD) and in degradation of the ERAD substrate CD3δ.[32] The MCM6 protein is essential for the initiation of eukaryotic genome replication and is highly expressed in patients with systemic lupus erythematosus.[33] The *DARS* gene, encodes the cytosolic aspartyl tRNA synthetase, with multiple non-canonical functions including mediation of inflammatory and immune responses.[33,34] Finally, overexpression of chemokine receptor CXCR4 in patients with systemic lupus erythematosus has been reported to promote inflammatory cell infiltration into renal tissue.[35] Given the alloimmune nature of inhibitor formation, the aforementioned genes and microRNA are interesting targets with putative role in FVIII inhibitor development.

To identify variants associated with inhibitor development, we performed target sequencing in the 586-kb chromosomal region 2q21.3-q22.1 encompassing the *LCT* gene in a large cohort of patients with a putative common genetic background (90% are of Italian ethnicity); 72.6% of patients are from the replication cohort reported in Gorski et al[21] and 26.4% are new patients accounting for 1/3 of the here analyzed cases (patients with inhibitor). The protective effect of the previously identified *LCT* p.Val219Ile missense variant was confirmed by the odd ratio almost unvaried than previous one (OR = .65 vs .58 and .64 in total and Italian discovery cohorts, respectively). Moreover, we identified seven additional common variants, four in both cohorts and three only in the Italian cohort, with a protective effect

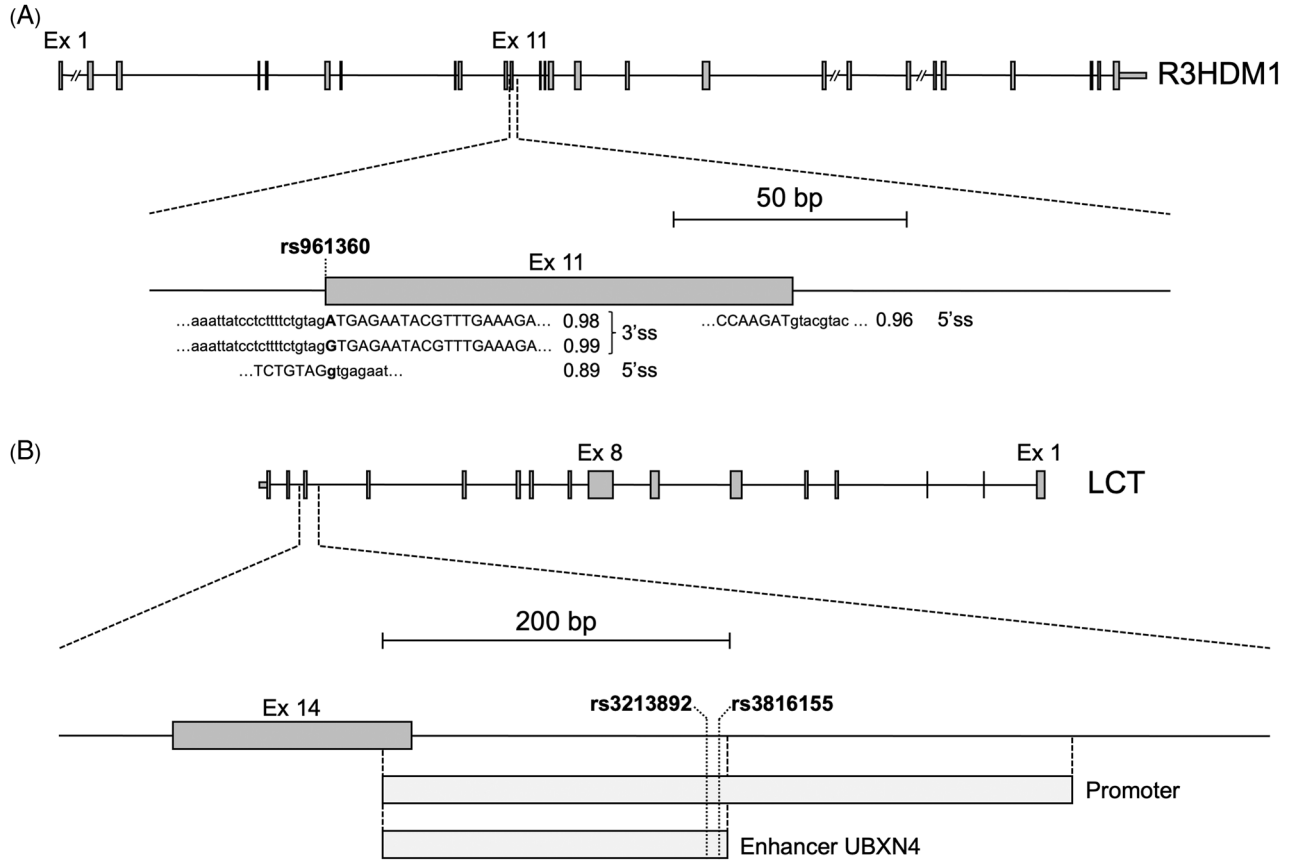

FIGURE 2 In silico analysis of rs961360 in *R3HDM1* and rs3213892 and rs3816155 in *LCT*. A schematic representation of *R3HDM1* (A) and *LCT* (B) genes is drown to scale; exons are represented by gray boxes and introns by straight lines. (A) Physiologic and predicted acceptor (3′) and donor (5′) splice sites of *R3HDM1* exon 11 are showed with the corresponding score (score range 0–1; threshold .6). (B) The LCT region with predicted regulatory elements is represented. The localization of rs3213892 and rs3816155 variants in intron 13 is shown

(OR < 1). Three of these variants (rs1050115 in *UBXN4*, rs3087343 in *MCM6* and rs3213892 in *LCT*) were previously found in LD with the missense variant p.Val219Ile,[21] confirming the highly conserved haplotype region encompassing the *LCT*, the lactase enzyme responsible for the digestion of lactose. Persistence of LCT in adulthood is partly promoted by a single nucleotide polymorphism (rs4988235) at 14 kb upstream from *LCT* in the cis-acting enhancer element localized in the intron 13 of the *MCM6* gene. Interestingly, two of the identified variants (rs3213892 and rs3816155) are located close together (10 nt distance) in the *LCT* intron 13, a region predicted as a potential regulatory element of the *UBXN4* gene and localized 53 kb downstream from the start of transcription of the adjacent UBXN4 gene. Hence, the molecular mechanisms underlying the association between the identified variants and inhibitor development could rely on the different expression of the *UBXN4* gene that, as ERAD protein, could be involved in FVIII antigen presentation of antigen-presenting cells. Since one of the variants identified in the Italian cohort (rs961360) is located in the R3HDM1 exon 11 acceptor splice site and activation of a cryptic donor splice site around the mutated site was predicted in silico, an additional/alternative mechanism could rely on an alternative processing of the R3HDM1 transcript and different synthesis/function of the corresponding protein. Since limitations of this study relies on (i) the low statistical power (51.5%) due to the disease rarity and the unavailability of the right sample size that may be responsible of the high FDR values (.12-.22) and (ii) the sequencing restricted to the coding regions of target genes that did not allow the identification of possible variants in regulatory regions not covered by sequencing, assessment of the expression levels of the analysed immunological target genes (*R3HDM1*, *UBXN4*, *MCM6*, *DARS*, *CXCR4* and *miR128-1*) in patients with and without the identified variants should be undertaken to confirm or not our findings.

The lack of association of the identified variants with inhibitor development in our available replication cohort can rely on the different genetic background (i.e. ethnicity) of the SIPPET cohort consisting of several populations (mainly Indian, Egyptian and Iranian). Recently, the association of the *LCT* p.Val219Ile missense variant (rs3754689) to inhibitor development in patients with severe HA has been found in Afghans[36] but not confirmed in Brazilian and Iranian populations.[36,37] Since the allele frequencies of all the identified variants reported in the 1000 Genome project are similar in the European and the South Asian populations but are markedly different in the Italian population (Table 1), a replication study performed in an independent cohort of Italian patients could be suitable to overcome the problem of poor reproducibility of genotyping results.

TABLE 1 Statistical analysis of variants identified by NGS in the discovery and replication cohorts

										Discovery (N = 246)							Replication (N = 230)			
dbSNP ID	Position[a]	Variant	Gene	Ex/Int	cDNA	aa variation	Minor allele	MAF[b] EU	MAF[c] Italy	MAF cases	MAF ctr	*P* logistic	OR	95% CI	*P* adjusted	*P* cond[d]	*P* logistic	OR	95% CI	p adjusted
rs1050115	136511817	A > G	UBXN4	Ex 4	c.303A > G		G	.21	.41	.25	.38	.011	.56	.35-.87	.20	/	.58	1.13	.73-1.74	.93
rs3213892	136552517	G > A	LCT	Int 13	c.4977-172C > T	/	A	.21	.41	.26	.38	.015	.58	.37-.90	.20	.93	.93	1.02	.67-1.55	.93
rs3816155	136552526	A > C	LCT	Int 13	c.4977-181T > G	/	A	.21	.41	.27	.38	.023	.60	.39-.93	.24	.59	.41	1.19	.79-1.80	.93
rs3754689	136590746	C > T	LCT	Ex 2	c.655G > A	p.Val219Ile	T	.22	.43	.29	.38	.047	.65	.43-.99	.38	.92	.69	1.09	.72-1.65	.93
rs3087343	136622543	T > G	MCM6	Int 7	c.1078+40A > C	/	G	.22	.43	.25	.38	.014	.56	.36-.89	.20	.89	.79	1.06	.70-1.62	.93

Abbreviations: aa, aminoacid; CHI_SQ, chi square; CI, confidence interval; ctr, controls; EU, European; Ex, exon; Int, intron; MAF, minor allele frequency; N, number of patients; OR, odds ratio; p-adjusted, multiplicity adjusted p-value, calculated using the Benjamini and Hochberg's (FDR_BH) method.

[a]Genomic coordinates of chromosome 2 (hg19).

[b]Frequency from 1000 Genomes Project Phase 3_European population.

[c]Frequency from 1000 Genomes Project Phase 3_Italian Tuscany population.

[d]p (logistic) conditioned rs1050115.

TABLE 2 Statistical analysis of variants identified by NGS in the discovery cohort of Italian patients

										Italian cohort (N = 220)						
dbSNP ID	Position[a]	Variant	Gene	Ex/Int	cDNA	aa variation	Minor allele	MAF[b] EU	MAF[c] Italy	MAF cases	MAF ctr	*P* logistic	OR	95% CI	*P* adjusted	*P* cond[d]
rs961360	136393658	A > G	R3HDM1	Ex 11	c.808A > G	p.Met270Val	G	.22	.41	.25	.37	.030	.59	.36-.95	.18	.98
rs12466487	136407078	T > C	R3HDM1	Int 16	c.1624-2222T > C	/	C	.19	.38	.22	.33	.030	.57	.34-.95	.18	.65
rs1050115	136511817	A > G	UBXN4	Ex 4	c.303A > G		G	.21	.41	.23	.38	.004	.48	.29-.79	.12	/
rs3213892	136552517	G > A	LCT	Int 13	c.4977-172C > T	/	A	.21	.41	.25	.38	.012	.54	.33-.87	.12	.56
rs3816155	136552526	A > C	LCT	Int 13	c.4977-181T > G	/	A	.21	.41	.25	.38	.012	.54	.33-.87	.12	.56
rs3754689	136590746	C > T	LCT	Ex 2	c.655G > A	p.Val219Ile	T	.22	.43	.28	.38	.050	.64	.40-1.0	.22	.44
rs3087343	136622543	T > G	MCM6	Int 7	c.1078+40A > C	/	G	.22	.43	.24	.38	.011	.52	.32-.86	.12	.79
rs371309040	136690425	T > C	DARS1	Int 6	c.505-13T > C	/	C	.21	.44	.28	.39	.028	.59	.37-.94	.18	.71

Abbreviations: aa, aminoacid; CHI_SQ, chi square; CI, confidence interval; ctr, controls; EU, European; Ex, exon; Int, intron; MAF, minor allele frequency; N, number of patients; OR, odds ratio; p-adjusted, multiplicity adjusted p-value, calculated using the Benjamini and Hochberg's (FDR_BH) method.

[a]Genomic coordinates of chromosome 2 (hg19).

[b]Frequency from 1000 Genomes Project Phase 3_European population.

[c]Frequency from 1000 Genomes Project Phase 3_Italian Tuscany population.

[d]p (logistic) conditioned rs1050115.

In conclusion, with the forementioned limitations, there is some evidence that LCT locus is a susceptibility locus for inhibitor development in Italian patients with severe HA. Further deeper investigations are needed to understand the effective contribution of *R3HDM1* and *UBXN4* on FVIII inhibitor development. Apart from their possible immunological role, the identified association between the genetic variants in the LCT locus and the risk of inhibitor development may be useful to better predict the risk of inhibitor development in Italian patients with severe HA.

ACKNOWLEDGEMENTS

The authors thank Luigi Flaminio Ghilardini (Università degli Studi di Milano, Milan) for his help with figures and Sonia Marino for her help with data collection. This work was supported by the Italian Ministry of Health: Ricerca Finalizzata RF-2016-02363246 and Bando Ricerca Corrente 2020.

Open Access Funding provided by Universita degli Studi di Milano within the CRUI-CARE Agreement.

CONFLICTS OF INTEREST

RG reports participation in advisory boards for Biomarin, Pfizer, Bayer and Takeda and participation at educational seminars sponsored by Pfizer, Sobi and Roche. FP reports participation in advisory boards for Roche, Sanofi, Sobi, Takeda and Biomarin and participation at educational meeting sponsored by Grifols and Roche. The other authors have no disclosures.

AUTHOR CONTRIBUTIONS

SS performed in-silico analyses, interpreted the results and drafted the manuscript. AC performed data analysis, interpreted the results and participated in manuscript preparation. EP performed NGS experiments. MMG designed the research and critically reviewed the manuscript. IG collected data, helped with the interpretation of the results and critically reviewed the manuscript. SH helped with the interpretation of the results and critically reviewed the manuscript. RG collected patients' data. FP supervised the research, interpreted the results and critically reviewed the manuscript. All authors read and approved the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

REFERENCES

1. Mannucci PM, Tuddenham EG. The hemophilias–from royal genes to gene therapy. *N Engl J Med*. 2001;344:1773-1779.
2. Peyvandi F, Garagiola I, Young G. The past and future of haemophilia: diagnosis, treatments, and its complications. *Lancet*. 2016;388:187-197.
3. Witmer C, Young G. Factor VIII inhibitors in hemophilia A: rationale and latest evidence. *Ther Adv Hematol*. 2013;4:59-72.
4. Turecek PL, Váradi K, Keil B, et al. Factor VIII inhibitor-bypassing agents act by inducing thrombin generation and can be monitored by a thrombin generation assay. *Pathophysiol Haemost Thromb*. 2003;33:16-22.
5. Lenting PJ, Denis CV, Christophe OD. Emicizumab, a bispecific antibody recognizing coagulation factors IX and X: how does it actually compare to factor VIII?. *Blood*. 2017;130:2463-2468.
6. D'Angiolella LS, Cortesi PA, Rocino A, et al. The socioeconomic burden of patients affected by hemophilia with inhibitors. *Eur J Haematol*. 2018;101:435-456.
7. Garagiola I, Palla R, Peyvandi F. Risk factors for inhibitor development in severe hemophilia A. *Thromb Res*. 2018;168:20-27.
8. Spena S, Garagiola I, Cannavò A, et al. Prediction of factor VIII inhibitor development in the SIPPET cohort by mutational analysis and factor VIII antigen measurement. *J Thromb Haemost*. 2018;16:778-790.
9. Oldenburg J, Pavlova A. Genetic risk factors for inhibitors to factors VIII and IX. *Haemophilia*. 2006;12(6):15-22.
10. Gouw SC, van den Berg HM, Oldenburg J, et al. F8 gene mutation type and inhibitor development in patients with severe hemophilia A: systematic review and meta-analysis. *Blood*. 2012;119:2922-2934.
11. Hay CR, Ollier W, Pepper L, et al. HLA class II profile: a weak determinant of factor VIII inhibitor development in severe haemophilia A. *Thromb Haemost*. 1997;77:234-237.
12. Oldenburg J, Picard JK, Schwaab R, et al. HLA genotype of patients with severe haemophilia A due to intron 22 inversion with and without inhibitors of factor VIII. *Thromb Haemost*. 1997;77:238-242.
13. Pavlova A, Delev D, Lacroix-Desmazes S, et al. Impact of polymorphisms of the major histocompatibility complex class II, interleukin-10, tumor necrosis factor-alpha and cytotoxic T-lymphocyte antigen-4 genes on inhibitor development in severe hemophilia A. *J Thromb Haemost*. 2009;7:2006-2015.
14. Pergantou H, Varela I, Moraloglou O, et al. Impact of HLA alleles and cytokine polymorphisms on inhibitors development in children with severe haemophilia A. *Haemophilia*. 2013;19:706-710.
15. Astermark J, Oldenburg J, Carlson J, et al. Polymorphisms in the TNFA gene and the risk of inhibitor development in patients with hemophilia A. *Blood*. 2006;108:3739-3745.
16. Astermark J, Oldenburg J, Pavlova A, et al. Polymorphisms in the IL10 but not in the IL1beta and IL4 genes are associated with inhibitor development in patients with hemophilia A. *Blood*. 2006;107:3167-3172.
17. Astermark J, Wang X, Oldenburg J, et al. Polymorphisms in the CTLA-4 gene and inhibitor development in patients with severe hemophilia A. *J Thromb Haemost*. 2007;5:263-265.
18. Pavlova A, Diaz-Lacava A, Zeitler H, et al. Increased frequency of the CTLA-4 49 A/G polymorphism in patients with acquired haemophilia A compared to healthy controls. *Haemophilia*. 2008;14:355-360.
19. Lozier JN, Rosenberg PS, Goedert JJ, et al. A case-control study reveals immunoregulatory gene haplotypes that influence inhibitor risk in severe haemophilia A. *Haemophilia*. 2011;17:641-649.
20. Bachelet D, Albert T, Mbogning C, et al. Risk stratification integrating genetic data for factor VIII inhibitor development in patients with severe hemophilia A. *PLoS One*. 2019;14:e0218258.
21. Gorski MM, Blighe K, Lotta LA, et al. Whole-exome sequencing to identify genetic risk variants underlying inhibitor development in severe hemophilia A patients. *Blood*. 2016;127:2924-2933.
22. Verbruggen B, van Heerde W, Novákovà I, et al. A 4% solution of bovine serum albumin can be used in place of factor VIII:c deficient plasma in the control sample in the Nijmegen Modification of the Bethesda factor VIII:c inhibitor assay. *Thromb Haemost*. 2002;88:362-364.
23. Peyvandi F, Mannucci PM, Garagiola I, et al. A randomized trial of factor VIII and neutralizing antibodies in hemophilia A. *N Engl J Med*. 2016;374:2054-2064.
24. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16:1215.

25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Res*. 2010;38:e164.
26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
27. Boll W, Wagner P, Mantei N. Structure of the chromosomal gene and cDNAs coding for lactase-phlorizin hydrolase in humans with adulttype hypolactasia or persistence of lactase. *Am J Hum Genet*. 1991;48:889-902.
28. Garcia-Etxebarria K, Jauregi-Miguel A, Romero-Garmendia I, et al. Ancestry-based stratified analysis of Immunochip data identifies novel associations with celiac disease. *Eur J Hum Genet*. 2016;24:1831-1834.
29. Guerau-de-Arellano M, Smith KM, Godlewski J, et al. Micro-RNA dysregulation in multiple sclerosis favours pro-inflammatory T-cell-mediated autoimmunity. *Brain*. 2011;134:3578-3589.
30. Galeazzi M, Balistreri E, Giannitti C, et al. MicroRNAs in autoimmune rheumatic diseases. *Reumatismo*. 2012;64:7-17.
31. Hu Y, O'Boyle K, Auer J, et al. Multiple UBXN family members inhibit retrovirus and lentivirus production and canonical NFκB signaling by stabilizing IκBα. *PLoS Pathog*. 2017;13:e1006187.
32. Liang J, Yin C, Doong H, et al. Characterization of erasin (UBXD2): a new ER protein that promotes ER-associated protein degradation. *J Cell Sci*. 2006;119:4011-4024.
33. Xie H, Hua C, Sun L, et al. 17β-estradiol induces CD40 expression in dendritic cells via MAPK signaling pathways in a minichromosome maintenance protein 6-dependent manner. *Arthritis Rheum*. 2011;63:2425-2435.
34. Guo M, Schimmel P. Essential nontranslational functions of tRNA synthetases. *Nat Chem Biol*. 2013;9:145-153.
35. Zhao LD, Liang D, Wu XN, et al. Contribution and underlying mechanisms of CXCR4 overexpression in patients with systemic lupus erythematosus. *Cell Mol Immunol*. 2017;14:842-849.
36. Bahrami Zadegan S, Mousavi SH, Damavandi N, et al. Investigating the influence of LCT rs3754689 polymorphism on inhibitor development in Iranian and Afghan patients with severe hemophilia A. *Blood Coagul Fibrinolysis*. 2020;31:11-15.
37. Zuccherato LW, Elói-Santos SM, Jardim LL, et al. Variation of rs3754689 at lactase gene and inhibitors in admixed Brazilian patients with hemophilia A. *Haematologica*. 2019;104:e527-e529.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Spena S, Cairo A, Pappalardo E, et al. Genetic variants at the chromosomal region 2q21.3 underlying inhibitor development in patients with severe hemophilia A. *Haemophilia*. 2022;28:270–277.
https://doi.org/10.1111/hae.14503