PROFESSOR MARIA-LUIZA PETZL-ERLER (Orcid ID : 0000-0002-0345-5276)

Artide type : Original Artide

Long non-coding RNA polymorphisms influence susceptibility to endemic pemphigus foliaceus

S.C. Lobo-Alves[1], D.G. Augusto[1,2], W.C.S. Magalhães[3,4], E. Tarazona-Santos[4], M.F. Lima-Costa[5], M.L. Barreto[6], B.L. Horta[7], R.C. de Almeida[8], M.L. Petzl-Erler[1]

[1]Laboratório de Genética Molecular Humana, Departamento de Genética, Universidade Federal do Paraná, Curitiba, Brazil; [2]Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Ilhéus, Brazil; [3]Núcleo de Ensino e Pesquisa - NEP, Instituto Mário Penna, Belo Horizonte, Minas Gerais, Brazil; [4]Instituto de Ciências Biológicas, Departamento de Biologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [5]Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG, Brazil; [6]Instituto de Saúde Coletiva, Universidade Federal da Bahia, Salvador, BA, Brazil; [7]Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil; [8]Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.

Corresponding author: Prof. Dr. Maria Luiza Petzl-Erler
E-mail: perler@ufpr.br

Short title: lncRNA polymorphisms in endemic pemphigus foliaceus

Funding: This research was supported by grants of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação Araucária, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Conflicts of interest: none to declare.

What is already known about this topic?

- The multifactorial autoimmune blistering skin disease pemphigus foliaceus presents a genetic susceptibility component that is not fully understood.
- Although pemphigus foliaceus is rare worldwide, it reaches the prevalence of 1.5% to 3% in some regions of Brazil, the highest ever reported for an autoimmune disease.
- LncRNA polymorphisms have been associated with some complex diseases but were not yet studied in any form of pemphigus.

What does this study add?

- Genetic variation of lncRNA may influence susceptibility to pemphigus foliaceus via its effect on lncRNA structure, on transcription of nearby genes, and on microRNA-lncRNA interactions.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bjd.17640

- We showed, for the first time, that variation in lncRNA genes may influence pemphigus foliaceus pathogenesis.

ABSTRACT

Background: Pemphigus foliaceus (PF) is an epidermal autoimmune disease, characterized by the presence of autoantibodies against the desmosomal protein desmoglein 1. Genetic and environmental factors contribute to PF, a complex disease that is endemic in Brazil and Colombia and neighbouring countries, and in Tunisia. Long non-coding RNAs (lncRNAs) may participate in gene regulation by interacting with DNA, proteins, and other RNAs. Dysregulation of lncRNAs has recently been recognized as an important co-player in the onset or progression of complex diseases. In addition, single-nucleotide polymorphisms (SNP) located in lncRNA genes have been associated with differential risk to cancer, autoimmunity and infection.

Objectives: Here, we aimed to investigate whether SNPs in lncRNA genes are associated with differential susceptibility to endemic PF.

Methods: We integrated data from the lncRNASNP database with genome-wide genotype data obtained for 229 patients and 6,681 controls. We tested the association between endemic PF and 2,080 SNPs located in lncRNAs applying logistic regression.

Results: The most significantly associated SNP was *rs7144332* (OR = 1.63, $P = 2.8 \times 10^{-6}$), located in the lncRNA gene *AL110292.1*. Results for other five SNPs were suggestive of association ($P < 0.001$). *In silico* analysis indicated that five of the six SNPs impact transcription, four may influence lncRNA secondary structure, and three of them may alter microRNA-lncRNA interactions.

Conclusion: We showed, for the first time, that variation in lncRNA genes may influence pemphigus pathogenesis. Our findings highlight the importance of lncRNA variation in autoimmune and possibly other complex diseases and suggest polymorphisms for functional validation.

Keywords: Long non-coding RNA, pemphigus, fogo selvagem, genetic polymorphism, case-control association study, SNP, genetic susceptibility, autoimmunity

INTRODUCTION

Pemphigus foliaceus (PF) is a rare autoimmune disease characterized by autoantibodies against desmoglein 1 (DSG1), a desmosomal cell-adhesion protein that is expressed in the epidermis[1]. The main clinical feature of PF is the formation of intra-epidermal blisters due to keratinocytes' loss of adhesion, a process called acantholysis[2]. In Central-Western Brazil, PF is endemic (EPF) and popularly known as *fogo selvagem* (that means "wild fire" in Portuguese). PF aetiology is multifactorial, with the participation of environmental and genetic factors. Several environmental factors have been speculated to be involved in PF pathogenesis, which for EPF includes precarious living conditions and hematophagous insect bites[2-4]. Diverse studies have demonstrated the importance of genetic factors. The first and strongest genetic associations ever reported were with the HLA class II genes encoding the beta chain of the HLA-DR and HLA-DQ molecules[5,6]. Polymorphisms in other protein-coding genes have been found associated with EPF[7-14]. A recent study showed that a 3' untranslated region polymorphism is associated with susceptibility to EPF by affecting a microRNA (miRNA) binding site[15]. Despite that, polymorphisms in non-coding regions have not been systematically analysed in any form of pemphigus.

For a long time, genetic studies of complex diseases focused mainly on protein-coding genes. Along the last two decades, however, the importance of non-coding regions and non-coding RNAs in biological processes became evident, such as gene expression regulation, dosage compensation, and cellular differentiation[16,17]. Interestingly, more than 90% of complex disease- and trait-associated variants that emerged from genome-wide association studies lie within non-coding sequences[18]. Recently discovered, long non-coding RNAs (lncRNAs) comprise a heterogeneous class of transcripts with over 200 nucleotides in length that may exhibit important molecular functions distinct from those transcripts that encode proteins[19].

LncRNAs genes are structurally and transcriptionally similar to protein-coding genes, with exons, introns and post-transcriptional processing; however, they show lower and more tissue-specific expression than protein-coding genes[20, 21]. Mature lncRNAs may be located in the nucleus and the cytoplasm and can interact with DNA, other RNAs, and proteins; yet the specific functions are still unknown for most of them. They participate in many cellular processes including gene expression regulation, and act, for example, in transport and maturation of RNA, epigenetic regulation, chromatin remodelling and as competitor endogenous RNAs (ceRNAs) [22, 23].

Aberrant lncRNA expression has been associated with susceptibility to diseases, likely playing a role in their onset or development[24-27]. Furthermore, single nucleotide polymorphisms (SNPs) in lncRNAs genes may influence genetic predisposition to complex diseases, such as infectious and autoimmune diseases and cancer[28-31]. However, lncRNA polymorphisms have been analysed only in a few diseases and never in any form of pemphigus. Therefore, in order to investigate if lncRNA SNPs influence susceptibility to endemic pemphigus foliaceus, we used genome-wide genotype dataset of nearly 7,000 individuals and integrated them with an updated lncRNA polymorphisms database. We found six lncRNA SNPs that may influence susceptibility to EPF. Additional *in silico* analysis

demonstrated that four of these SNPs may alter lncRNAs conformation, while three of them may influence potential miRNA-lncRNA interactions. Together, our results suggest that genetic variation of lncRNA affects the susceptibility to EPF and probably also other forms of pemphigus.

MATERIAL AND METHODS

Population sample

The sample comprised 229 EPF patients and 6,681 controls without the disease. Patients were diagnosed based on the clinical characterization combined with immunological test, histopathology and/or immunohistochemistry of skin biopsies. The patients were contacted in different Brazilian hospitals: Hospital Adventista do Pênfigo, in Campo Grande, Mato Grosso do Sul State; Hospital de Clínicas, UFPR, Hospital de Dermatologia Sanitária São Roque, Hospital Santa Casa de Misericórdia, in Curitiba, Paraná State; Hospital das Clínicas, USP, in Ribeirão Preto, São Paulo State; and Lar da Caridade Hospital do Fogo Selvagem, in Uberaba, Minas Gerais State. Part of the controls (194 individuals) was recruited in the endemic regions. The additional control sample of 6,487 individuals was collected in Minas Gerais state, Rio Grande do Sul state, and Bahia state, by the EPIGEN-Brasil Project[32]. The national ethics committee for research with human subjects (CONEP - Comissão Nacional de Ética em Pesquisa), approved the present project (protocol number 02727412.4.0000.0096. All individuals voluntarily participated and signed an informed consent form. Peripheral blood was collected, from which DNA was extracted by the phenol-chloroform-isoamyl alcohol protocol[33].

Data integration: lncRNAs SNPs and genotype data

Genotyping was performed with two different microarrays: patients and controls for the endemic area were genotyped with Human Core Exome-24 Illumina; the remaining controls were part of the EPIGEN project[32] and were genotyped with HumanOmmni5 and HumanOmni2.5, Illumina. We integrated the genome-wide genotyping data from patients and controls and applied principal component analysis (PCA) to correct for possible population substructure and batch effects. Quality control (QC) of the genotype data was performed as described by Kehdy *et al.*,[32] and Anderson *et al.*,[34]. In summary, we excluded SNPs with (1) a low call rate (< 95%), (2) deviation from Hardy-Weinberg equilibrium with $P < 0.00001$ in the control sample and (3) MAF < 0.01. After QC, a total of 204,967 markers remained for further analysis.

A list of 495,730 SNPs previously mapped to lncRNA genes was obtained from the lncRNASNP database v 1.0, available at http://bioinfo.life.hust.edu.cn/lncRNASNP[35]. This list of SNPs was intersected with the list of 204,967 SNPs that remained after our QC stage. A total of 2,080 intersected SNPs (Table S1) were used for association analysis.

Association analysis

A case-control analysis was performed by logistic regression for the additive model. To avoid bias due to population stratification, a principal component analysis was carried out using the EGENISTRAT algorithm[36]. For this analysis, the 204,967 markers that remained after QC were used to certify that samples from the EPIGEN-Brasil project belong to the same population stratum as the controls from endemic areas. Population structure was adjusted using the two principal components (PC1 and PC2) as covariant in the Plink 1.9 software[37]. Differences between the patient and control samples with *P* value ≤ 0.001 were considered as

suggestive associations, while the Bonferroni corrected *P* value of 2.4 x 10^{-5} was considered unambiguously statistically significant.

In silico analysis

We applied in silico analysis to explore the genetic associations identified in this study. Firstly, we hypothesized that the SNPs associated or suggestively associated with EPF could affect the secondary structure of lncRNAs and, consequently, their interactions with other RNAs or proteins, as well as their function. Next, we explored the functional annotation data for each SNP using publicly available tools and databases. Moreover, we explored the SNPs regarding their potential effect on lncRNA-microRNA interaction because microRNAs (miRNAs) can participate in post-transcriptional regulation of lncRNAs as well as lncRNAs can act as endogenous competitors of miRNAs. For this, were considered only variants present in the mature sequence of lncRNAs, as described below.

Secondary structure prediction of lncRNAs: The online version of RNAfold Web Server based on Vienna RNA package[38] was used to predict the secondary structure of lncRNAs which have variants located in exons of their genes associated or suggestively associated with EPF. To evaluate the impact of those SNPs on the lncRNA secondary structures, we first extracted lncRNA transcript sequences from the human reference genome (Ensembl version 93) and, for each SNP, we changed the corresponding variant site to obtain the structure of the lncRNA with both SNP alleles. Further, the RNAfold software was used to predicted lncRNA secondary structure based on minimum free energy (MFE) calculations.

Functional Annotation: To verify the predicted structural and regulatory impact of the genetic variants and if they are eQTL (expression Quantitative Trait Loci) we used the HaploRegtool[39], RegulomeDB[40], and GTEx Portal (www.gtexportal.org) databases, which assemble public datasets and published data, to provide information about the functional impact of SNPs.

Impact of SNPs on miRNA-lncRNA interaction: Predictions of the lncRNASNP2 database[41] were used to verify if the lncRNA SNPs could be creating or disrupting miRNA binding sites on lncRNA mature sequences. The lncRNASNP2 database obtains miRNA sequences from miRBase (release 21)[42] while predictions of miRNA target sites on the sequences around the SNP (± 25 bp) were achieved by the combination of results from miRanda, TargetScan, and Pita. Besides, the lncRNASNP2 database includes all SNPs from dbSNP from NCBI and also GWAS tag SNPs from NHGRI GWAS Catalog. Finally, by intersecting predictions from the lncRNASNP2 database for the pemphigus-associated SNPs, we were able to obtain a list of miRNAs whose binding motif includes the SNP site.

RESULTS

Association analysis

In order to identify SNPs located in lncRNAs associated with EPF, we tested 2,080 lncRNA-SNPs by performing a logistic regression on nearly 7,000 individuals. We identified six SNPs that possibly have an impact on susceptibility to EPF ($P < 0.001$; Table 1, Figure 1).

The variant *rs7144332 T* located at the lncRNA *AL110292.1* on chromosome 14 showed the most significant association (OR = 1.63, $P = 2.82 \times 10^{-6}$). Furthermore, we observed suggestive associations ($P < 0.001$) with the variants *rs6095016 A, rs7195536 G, rs1542604 T, rs6942557 C* and *rs17774133 T*. These variants are either in intronic or exonic regions of lncRNA genes that partially overlap protein-coding genes or other non-coding genes (Table 1), while the variant *rs6095016 A* (OR = 0.597, $P = 5.45 \times 10^{-5}$) is located in a lncRNA gene distant more than 13,000 bp from any other gene**.** The lncRNAs *AL110292.1, AC133785.1, linc01176, and linc01119* exhibit alternative splicing isoforms, thus, the variants *rs7144332 T, rs1542604 T, rs6942557 C*, and *rs17774133 T* can either be present in, or absent of, the mature lncRNA, depending on the particular mature lncRNAs isoform (Figure S1). Therefore, these SNPs may have different consequences for the distinct transcripts of the same lncRNA gene.

In silico analysis

To identify the possible effects of variants associated or suggestively associated with EPF, we performed three types of in silico analysis.

Initially, we focused on the five SNPs located in exons, which therefore are present in the lncRNA mature sequence, to verify their possible impact on lncRNA secondary structure. Different secondary structures were predicted for each variant of the *rs7195536 in AC009121.1, rs6942557* in *linc01176* (transcript ENST00000628121.1) and *rs1542604 in AC133785.1* (transcript ENST00000446213.2). No differences were seen in the predicted secondary structure for SNPs *rs6095016* and *rs17774133* located in lncRNAs *lnc-PREX1-7:1* and *linc01119* (transcript ENST00000468141.2), respectively (Figure 2).

Then, we explored the available functional annotations for each SNP. The RegulomeDB database[40] and the Haploreg v. 4.1 tool[39] provided evidence that the SNPs *rs7144332*, *rs6095016*, *rs7195536*, *rs1542604*, and *rs6942557* alter motifs for transcription factors, while SNPs *rs7195536*, *rs1542604*, *rs6942557*, and *rs17774133* overlap with enhancers that are important in several tissues, including skin (Table S2. Moreover, according to the GTEx portal (www.gtexportal.org) the SNP *rs7195536* located in lncRNA *AC009121.1* is an eQTL for three nearby genes in chromosome 16: *RMI2* (RecQ mediated genome instability), a protein-coding gene; *RP11485G7.5* (*AC009121.2*), a lncRNA gene; and *CTD-3088G3.4* (*AC099489.2*), a processed pseudogene.

Furthermore, we intersected our results with the predicted lncRNA-miRNA interactions from the lncRNASNP2 database[41] to identify a possible effect of each associated SNP on the gain or loss of miRNA binding sites. Three suggestively associated variants (*rs1542604* T, *rs6942557* C, and *rs17774133* T*)* were predicted to create a new miRNA binding site in the lncRNAs (Table 2). Also, two of these variants (*rs6942557 C* and *rs17774133 T*) may disrupt miRNA binding sites. The variant *rs17774133 T* was the most likely to affect the miRNA:lncRNA interaction since it was predicted to likely affects the binding sites of six miRNAs (Table 2).

DISCUSSION

Here we report the first case-control association study searching for the influence of lncRNA polymorphisms in pemphigus. The genotype and allele frequencies of 2,080 SNPs distributed over the whole genome were obtained for 229 EPF patients and 6,681 controls without the disease. Our goal was to identify lncRNA variants which impact pemphigus pathogenesis.

On the basis of the results, we suggest that alterations in the lncRNA structure and gene expression possibly affect their function and influence pemphigus susceptibility. The findings highlight the importance of lncRNA variation in autoimmune and possibly other complex diseases. The interest in the genetic variation of lncRNAs genes is relatively recent since it has only become evident within the last few years that alterations of lncRNA impact biological processes and diseases. With the recent reports that SNPs in lncRNAs are associated with different complex diseases[28-31], questions about how those SNPs affect the lncRNA function have emerged.

We found one SNP strongly associated with EPF and suggestive associations with other five SNPs. All six SNPs are located in poorly known lncRNA genes. There are no studies of these lncRNAs and no description about their functions and interactions with proteins, DNA or RNAs, which limits our ability to infer their functional impact. However, for the five exonic SNPs, *rs7195536*, *rs6095016*, *rs6942557*, *rs1542604*, and *rs17774133*, it was possible to hypothesize that they may alter the lncRNA secondary structures. Variation in non-coding RNA nucleotide sequence has been reported to influence the RNA structure and gene regulation[43]. Indeed, in silico analysis predicted that the structure of three of analysed lncRNAs differed according to the single nucleotide variants *rs7195536* G, *rs6942557* C and *rs1542604* T. Considering that the functional effects of lncRNAs depend on their conformation and this relies on its nucleotide sequence, single nucleotide variations can impair lncRNA function by changing its primary and secondary structures.

Variation in lncRNA can also modify interactions between lncRNAs and other molecules, such as DNA, other RNAs or proteins. For example, the *rs917997 T* variant associated with celiac disease alters *lnc13* affinity to hnRNPD (heterogeneous nuclear ribonucleoprotein D), thus affecting the lncRNA-protein complex and its action in the regulation of gene expression[30].

Our in silico analysis suggests that *rs1542604*, *rs6942557*, and *rs17774133* can create or disrupt binding sites for miRNAs. This kind of interference of a lncRNA SNP has been observed in pancreatic cancer[44]. That study showed that the *rs11655237 G>A* SNP located in the exon 4 of *LINC00673* creates a miR-1231 binding site that affects the lncRNA levels in a variant-specific manner. Thus, we suggest that the SNPs predicted to create or disrupt miRNA binding sites may alter potential lncRNA-miRNA interactions, which is a plausible explanation for the observed suggestive genetic associations with EPF susceptibility.

Four of the EPF-associated and suggestively associated lncRNAs SNPs reported here - *rs7144332*, *rs1542604*, *rs6942557*, and *rs17774133* - are located in introns or alternatively spliced exons and may not be present in at least one of the mature lncRNAs. These SNPs can be located in regulatory sequences, such as promoters, enhancers or alternative splicing signal sequences and modify the expression of lncRNA genes. Variation in lncRNA genes can interfere in its own transcriptional or post-transcriptional regulation. An example is the *T* allele of the cervical cancer-associated SNP *rs920778* located in an intronic *HOTAIR* enhancer that regulates its expression[29].

The functional annotation analysis revealed that the SNP *rs7195536* located in lncRNA *AC009121.1* is an eQTL for three nearby genes: *RMI2*, *RP11485G7.5* (*AC009121.2*), and *CTD-3088G3.4* (*AC099489.2*). Thus, the expression of these genes probably is (co-)regulated by lncRNA *AC009121.1* and influenced by the *rs7195536* genotypes. Moreover, based on the high score from Regulome DB for *rs6942557* (1b; Table S2) and the lack of strong linkage disequilibrium (defined as $r^2 > 0.8$) between this variant and others (Table S3), we suggest that this variant may have a direct effect on EPF susceptibility and is a strong candidate for functional validation.

Some possibilities of how an SNP can affect the function of lncRNAs are suggested here solely based on in silico analysis and require further *in vitro* or *in vivo* functional validation. It is also important to bear in mind that some SNPs associated or suggestively associated with EPF identified here are in linkage disequilibrium ($r^2 > 0.8$) with other SNPs (Table S3), and therefore they may not be the causal variants.

This is the first study that identifies SNPs located in lncRNA genes that are associated with pemphigus. There are no previous reports of variation within lncRNA loci being associated with any form of pemphigus. The results of our study indicate that genetic variants of lncRNA influence pemphigus pathogenesis, and that these lncRNAs are involved in biologic processes relevant to the disease. Therefore, it is plausible that these SNPs contribute to disease manifestation due to their impact on molecular interactions. This work adds valuable information to the growing research about susceptibility and pathogenesis of pemphigus and suggests possible targets for functional validation. Besides, our findings highlight the importance of lncRNA variation in autoimmune and possibly other complex diseases.

ACKNOWLEDGMENTS

We thank all the patients and controls subjects that voluntarily have participated in this study. We also thank the staff of the Laboratório de Genética Molecular Humana - UFPR for their assistance and help when needed.

REFERENCES

1. Jones JC, Arnn J, Stachelin LA, Goldman RD. Human antibodies against desmosomes: possible causative factors in pemphigus. *Proc Natl Acad Sci USA* 1984; **81**:2781-85.

2. Diaz LA, Sampaio SAP, Rivitti EA, *et al.* Endemic pemphigus foliaceus (fogo selvagem): II. Current and historical epidemiologic studies. *J Invest Dermatol* 1989; **92**:4-12.

3. Lombardi C, Borges PC, Chaul A, *et al.* (1992). Environmental risk factors in endemic pemphigus foliaceus (Fogo selvagem). "The Cooperative Group on Fogo Selvagem Research". *J Invest Dermatol* 1992; **98**:847-850

4. Qian Y, Jeong JS, Maldonado M, *et al.* Cutting Edge: Brazilian Pemphigus Foliaceus Anti-Desmoglein 1 autoantibodies cross-react with sandy fly salivary LJM11 antigen. *J Immun* 2013; **189**:1535-39.

5. Petzl-Erler ML, Santamaria J. Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)? *Tissue Antigens* 1989; **33**:408-14.

6. Pavoni DP, Roxo VM, Marquart Filho A, Petzl-Erler ML. Dissecting the associations of endemic pemphigus foliaceus (Fogo Selvagem) with HLA-DRB1 alleles and genotypes. *Genes Immun* 2003; **4**:110-116.

7. Malheiros D, Petzl-Erler ML. Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus. *Genes Immun* 2009; **10**:547-58.

8. Pereira NF, Hansen JA, Lin MT, *et al.* Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants. *Cytokine* 2004; **28**:233-41.

9. Braun-Prado K, Petzl-Erler ML. Programmed cell death 1 gene (PDCD1) polymorphism and pemphigus foliaceus (fogo selvagem) disease susceptibility. *Genet Mol Bio* 2007; **30**:314.

10. Dalla-Costa R, Pincerati MR, Beltrame MH, *et al.* Polymorphisms in the 2q33 and 3q21 chromosome regions including T-cell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. *Hum Immunol* 2010; **71**:809-17.

11. Piovezan BZ, Petzl-Erler ML. Both qualitative and quantitative genetic variation of MHC class II molecules may influence susceptibility to autoimmune diseases: the case of endemic pemphigus foliaceus. *Hum Immunol* 2013; **74**:1134-140.

12. Augusto DG, Lobo-Alves SC, Melo MF, *et al.* Activating KIR and HLA Bw4 ligands are associated to decreased susceptibility to pemphigus foliaceus, an autoimmune blistering skin disease. *PloS One* 2012; **7**:7 :e39991. doi: 10.1371/journal.pone.0039991

13. Augusto DG, O'Connor GM, Lobo-Alves SC, *et al.* Pemphigus is associated with KIR3DL2 expression levels and provides evidence that KIR3DL2 may bind HLA-A3 and A11 in vivo. *Eur J Immunol* 2015; **45**:2052-60.

14. Camargo CM, Augusto DG, Petzl-Erler ML. Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus. *Hum Genet* 2016; **135**:233-44.

15. Cipolla GA, Park JK, de Oliveira LA, *et al.* A 3'UTR polymorphism marks differential KLRG1 mRNA levels through disruption of a miR-584-5p binding site and associates with pemphigus foliaceus susceptibility. *Biochem Biophys Acta* 2016; **1859**:1306-13.

16. Penny GD, Kay GF, Sheardown SA, *et al.* Requirement for Xist in X chromosome inactivation. *Nature* 1996; **379**:131–37.

17. Wang KC, Chang HY. Molecular mechanisms of long noncoding RNAs. *Mol Cel* 2011; **43**:904–914.

18. Maurano MT, Humbert R, Rynes E, *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* 2012; **337**: 1190-95.

19. Harrow J, Frankish A, Gonzalez JM, *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012; **22**: 1760-74.

20. Derrien T, Johnson R, Bussotti G, *et al.* The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Res* 2012; **22**:1775-89.

21. Chen L-L. Linking long noncoding RNA localization and function. Trends Biochem Sci 2016; 41:761-72.

22. Rinn JL, Chang HY. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 2012; **81**:145–166.

23. Salviano-Silva A, Lobo-Alves SC, Almeida RC, *et al.* Besides Pathology: Long Non-Coding RNA in Cell and Tissue Homeostasis. *Noncoding RNA* 2018; **4**(1):3. doi: 10.3390/ncrna4010003.

24. Gutschner T, Hämmerle M, Eißmann M, *et al.* The non-coding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells. *Cancer Res* 2012; **73**:1180-89.

25. Huarte M, Guttman M, Feldser D, *et al.* A large intergenic noncoding RNA induced byp53 mediates global gene repression in the p53 response. *Cell* 2010; **142**:409-19.

26. Lu L, Zhu G, Zhang C, *et al.* Association of large noncoding RNA hotair expression and downstream intergenic CpG island methylation with survival in breast cancer. *Breast Cancer Res Treat* 2012; **136**:875-83.

27. Kogo R, Shimamura T, Mimori K, *et al.* Long noncoding RNA hotair regulates polycomb-dependent chromatin modification and is associated with poor prognosis in colorectal cancers. *Cancer Res* 2011; **71**:6320-326.

28. Rautanen A, Pirinen M, Mills TC, *et al.* Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. *Am J Hum Genet* 2016; **98**:1-9.

29. Guo L, Lu X, Zheng L, *et al.* Association of Long Non-Coding RNA HOTAIR Polymorphisms with Cervical Cancer Risk in a Chinese Population. *PLoS One* 2016; **11**:7 e0160039. doi: 10.1371/journal.pone.0160039

30. Rubio A, Fernandez-Jimenez N, Kratchmarov R, *et al.* A long noncoding RNA associated with susceptibility to celiac disease. *Science* 2016; **352**:91-95.

31. Kumar V, Westra H-J, Karjalainen J, *et al.* Human Disease-Associated Genetic Variation Impacts Large Intergenic Non-Coding RNA Expression. *PLoS Genet* 2013; **9**:1e1003201. doi: 10.1371/journal.pgen.1003201

32. Kehdy FSG, Gouveia MH, Machado M, *et al*. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proc Natl Acad Sci USA* 2015; **112**:8696-701.

33. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual, CSHL Press, New York, 2001.

34. Anderson CA, Pettersson FH, Clarke GM, *et al*. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564-73.

35. Gong J, Liu W, Zhang J, Miao X, Guo A-Y. lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res* 2015; **43**(D1): D181-D186.

36. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genetics* 2006, **38**:904-909.

37. Chang CC, Chow CC, Tellier LC, *et al*. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015, **4**:7.

38. Lorenz R, Bernhart SH, Höner zu Siederdissen C, *et al*. ViennaRNA Package 2.0. *Algorithms Mol. Biol*.2011; **6**:26.

39. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012; **40** (DI), D930–D934.

40. Boyle AP, Hong EL, Hariharan M, *et al*. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790-97.

41. Miao Y-R, Liu W, Zhang Q, Guo A-Y. lncRNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucl Acids Res* 2017; **46**(DI):D276-D280.

42. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* 2014; **42**:(DI), D58–D73.

43. Wan Y, Qu K, Zhang QC, *et al*, Landascape and variation of RNA secondary structure across the human transcriptome. *Nature* 2014; **505**: 706-709.

44. Zheng J, Huang X, Tan W, *et al.* Pancreatic cancer risk variant in LINC00673 creates a miR-1231 binding site and interferes with PTPN11 degradation. *Nat Genet* 2016; **48**: 747-57.

Table 1. Long non-coding RNA SNPs associated ($P < 2.4 \times 10^{-5}$) and suggestively associated ($0.001 > P > 2.4 \times 10^{-5}$) with endemic pemphigus foliaceus.

CHR	lncRNA	aliases	Partially overlapped genes	SNP	Minor alelle	MAF	OR	95% CI	*P*
14	*AL110292.1*	lnc-FOXG1-22 RP11-384J4.2	*MIR4307HG*	*rs7144332* T>G	*T*	0.38	1.627	1.33 -1.99	2.82×10^{-6}
20	*lnc-PREX1-7:1*	XLOC_013774 linc-SULF2-7; TCONS_00028423	*none*	*rs6095016* G>A	*A*	0.26	0.597	0.46 - 0.77	5.45×10^{-5}
16	*AC009121.1*	lnc-PRM1-2 RP11-485G7.6	*RMI2, AC009121.2*	*rs7195536* A>G	*G*	0.26	0.578	0.44 - 0.76	7.52×10^{-5}
2	*AC133785.1*	lnc-FAM168B-1 XLOC_002322 linc-CFC1B	*ARHGEF4*	*rs1542604* C>T	*T*	0.15	0.555	0.40 - 0.77	4.63×10^{-4}
7	*LINC01176*	XLOC_006035 linc-GARS-3; lnc-ZNRF2-4	*AC006978.2*	*rs6942557* T>C	*C*	0.10	1.678	1.24 - 2.27	7.64×10^{-4}
2	*LINC01119*	lnc-SOCS5-4 XLOC_001458 linc-TTC7A-2	*LOC107985880*	*rs17774133* C>T	*T*	0.19	1.488	1.17 - 1.90	9.77×10^{-4}

CHR: chromosome, MAF: minor allele frequency in the control sample of the present study, OR: odds ratio, CI: confidence interval, *P*: *P* value.

Table 2. Prediction of the SNPs' effects on interactions between the long non-coding RNAs and microRNAs.

lncRNA / variant	miRNA	Potential SNP Effect	Alignment Length (in bases)
AC133785.1/ rs1542604 T	hsa-miR-4775	gain	13 (62.2%)
	hsa-miR-550b-3p	gain	16 (75%)
Linc01176 / rs6942557 C	hsa-miR-4294	gain	7 (100%)
	hsa-miR-4650-3p	loss	7 (100%)
Linc01119/ rs17774133 T	hsa-miR-3154	gain	22 (68.2%)
	has-miR-4520a-3p	gain	9 (100%)
	hsa-miR-3202	loss	7 (100%)
	hsa-miR-4804-5p	loss	14 (78.6%)
	hsa –miR-184	loss	21 (76.2%)
	hsa-miR-6125	loss	14 (71.4%)

Gain: represents addition of a miRNA binding site by the minor variant of the SNP. Loss: represents removal of a miRNA binding site by the minor variant of the SNP.

SUPPORTING INFORMATION

Figure S1. EPF associated and suggestively associated SNPs' position in the lncRNAs primary transcripts.

Table S1. List of single nucleotide polymorphisms (SNPs) analyzed for possible association with endemic pemphigus foliaceus.

Table S2. Predicted regulatory functional annotations for EPF-associated and suggestively associated lncRNA SNPs.

Table S3. Single nucleotide polymorphisms (SNPs) in strong linkage disequilibrium ($r^2 > 0.8$) with lncRNA SNPs associated and suggestively associated with endemic pemphigus foliaceus, in the 1000 genomes European populations.

FIGURE LEGENDS

Figure 1. Manhattan plot with results of the association analysis between endemic pemphigus foliaceus and SNPs mapped in long ncRNA genes.

Horizontal line: threshold for suggestively significant associations ($P = 0.001$). Spots represent the SNPs.

Figure 2. Predicted secondary structures for both alleles of lncRNAs in which EPF-associated or suggestively associated SNPs are present in their mature sequence. The lncRNAs *AC009121.1, Linc011761,* and *AC133785.1* have different secondary structures predicted to each allele. Vienna Package – RNAfold Web Server.