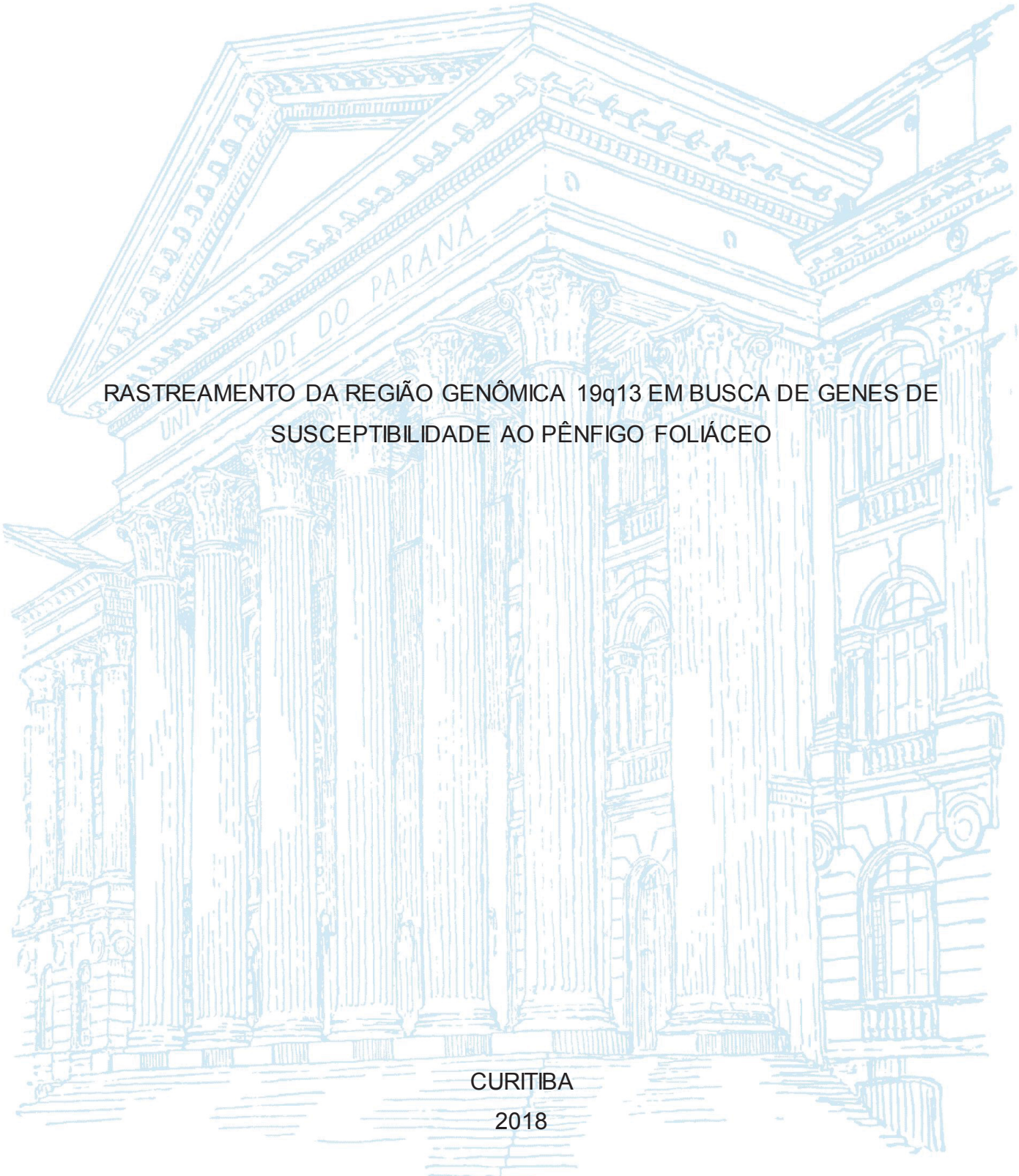
UNIVERSIDADE FEDERAL DO PARANÁ

TICIANA DELLA JUSTINA FARIAS



RASTREAMENTO DA REGIÃO GENÔMICA 19q13 EM BUSCA DE GENES DE SUSCEPTIBILIDADE AO PÊNFIGO FOLIÁCEO

CURITIBA

2018

TICIANA DELLA JUSTINA FARIAS

RASTREAMENTO DA REGIÃO GENÔMICA 19q13 EM BUSCA DE GENES DE SUSCEPTIBILIDADE AO PÊNFIGO FOLIÁCEO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Genética, Setor de Ciências Biológicas, Universidade Federal do Paraná, como requisito parcial para a obtenção de título de Doutora em Genética.

Orientadora: Profª Drª Maria Luiza Petzl-Erler

Coorientadora: Profª Drª Danielle Malheiros

CURITIBA

2018

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Rosilei Vilas Boas – CRB/9-939).

Farias, Ticiana Della Justina.
Rastreamento da região genômica 19q13 em busca de genes de susceptibilidade ao pênfigo foliáceo. / Ticiana Della Justina Farias. – Curitiba, 2018.
112 f. : il. ; 30cm.

Orientadora: Maria Luiza Petzl-Erler.
Coorientadora: Danielle Malheiros.
Tese (Doutorado) – Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

1. Pênfigo. 2. Polimorfismo (Genética). 3. Pele - Doenças. I. Título. II. Petzl-Erler, Maria Luiza. III. Malheiros, Danielle. IV. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Genética.

CDD (20. ed.) 616.51

MINISTÉRIO DA EDUCAÇÃO
SETOR CIÊNCIAS BIOLÓGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO GENÉTICA

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GENÉTICA da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **TICIANA DELLA JUSTINA FARIAS** intitulada: **Rastreamento da região genômica 19q13 em busca de genes de susceptibilidade ao pênfigo foliáceo**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

Curitiba, 25 de Setembro de 2018.

MARIA LUIZA PETZL ERLER
Presidente da Banca Examinadora

JAQUELINE CARVALHO DE OLIVEIRA
Avaliador Interno

MARCIA REGINA PINCERATI
Avaliador Externo

EDUARDO ANTÔNIO DONADI
Avaliador Externo

MARCIA HOLSBACH BELTRAME
Avaliador Interno

DANIELLE MALHEIROS FERREIRA
Avaliador Interno

AGRADECIMENTOS

Agradeço aos meus pais pelo apoio incondicional e, também, pelos excelentes exemplos de caráter, dedicação, amor e compreensão. Aos meus irmãos, cunhados e Benjamin que transformam todos encontros (físicos e virtuais) muito divertidos e cheio de carinho e amor. Amo vocês!

Ao meu marido, grande amigo e companheiro Lucas. O seu apoio, amor e compreensão são fundamentais para que eu siga atrás dos meus (e nossos) sonhos. Obrigada pelo suporte e por lidar da melhor forma com a distância e saudades durante os períodos que estive fora do Brasil.

Aos familiares e amigos residentes de Curitiba, principalmente às minhas queridas primas-irmãs Scheila e Hellen. A Cíntia pela amizade e auxílio na revisão do texto. Vocês transformaram a estadia em Curitiba tão agradável, amenizando a saudades de casa.

À minha família estendida Antunes-Wendhausen. Obrigada pelo imenso carinho e apoio.

Agradeço a grande equipe do Laboratório de Genética Humana Molecular (LGMH) pelas grandes amizades, risadas e parceria nos experimentos nestes quase 6 anos. Com receio de faltar algum nome aqui, quero agradecer principalmente aos amigos e colaboradores direto deste trabalho, Danillo e Rodrigo, e minha orientadora Luiza e coorientadora Dani. Obrigada pelo imensurável aprendizado e amizade. Sinto-me profundamente lisonjeada de participar desse grupo de pesquisa formidável!

Aos amigos e pesquisadores colaboradores de Lübeck, Alemanha. Especialmente ao prof. Enno Schmidt e Hauke Busch pela oportunidade de participar do grupo LIED, iniciando uma colaboração científica promissora. Eu tenho imenso carinho por vocês e pela cidade de Lübeck.

Aos pacientes de pênfigo e controles que participaram deste estudo e aos funcionários dos hospitais pela colaboração e acolhimento.

Aos órgãos de fomento Fundação Araucária, CNPq e CAPES pelo financiamento nos projetos de pesquisa. Aos programas da CAPES e CAPES/PDSE (Processo PDSE 88881.132221/2016-01) pela bolsa durante o período de doutorado e doutorado sanduíche, respectivamente, e a agência de intercâmbio alemã, DAAD, pelo auxílio de mobilidade e moradia no exterior.

“Ciência e a vida cotidiana não podem e não devem ser separadas”

Rosalind Franklin

RESUMO

O pênfigo foliáceo (PF) é uma doença bolhosa autoimune rara da pele caracterizada pelo desprendimento de queratinócitos e presença de autoanticorpos contra desmossomos, principalmente anti-desmogleína 1 (DSG1). O PF ocorre esporadicamente no mundo, porém sua incidência é maior em certas regiões onde o PF é endêmico (PFE). A etiologia de PF é complexa, com contribuição de fatores ambientais e genéticos parcialmente conhecidos. Polimorfismos em genes relacionados a respostas imunes foram associados ao PFE, incluindo genes localizados no complexo de receptores leucocitários (LRC, do inglês *leucocyte receptors complex*), na região 19q13.42. Além disso, vários genes do LRC estiveram diferencialmente expressos em pacientes de PFE, em diversas formas clínicas, e controles, que sustenta a hipótese de que variantes no LRC contribuem na patogênese do PF. O objetivo deste trabalho foi realizar uma profunda varredura do LRC estendido, 19q13, em busca de variantes genéticas que possam influenciar a susceptibilidade ao PF. Para isso, 229 pacientes com PFE e 194 indivíduos controles foram genotipados para 2.510 SNPs (polimorfismo de nucleotídeo único, do inglês *single nucleotide polymorphism*) do LRC estendido por microarranjo de DNA. Nove SNPs estiveram associados ao PFE ($p < 10^{-3}$). Destes, os SNPs *rs8103298* (OR = 2,50) e *rs3810265* (OR = 2,33) estão no pseudogene *RPS9P4* e quatro SNPs (*rs9304655*, *rs6509255*, *rs10412596*, *rs10439109*) estão localizados no gene do RNA longo não-codificador (lncRNA) *lnc-IGFL2-1* (1,6 <OR < 2,2). Além desses, variantes nos genes codificadores *SULT2B1* (*rs11083944*, OR = 0,49), *SCAF1* (*rs2304208*, OR = 3,04) e *RDH13* (*rs1654431*, OR = 2,34) estiveram associadas ao PFE. Na região específica do LRC, 19q13.42, três dos 176 SNPs estiveram associados ao PFE: a variante intergênica *rs461569* (OR = 1,50, $p = 0,004$), e as variantes *rs35336528* e *rs1865097* (OR = 3,39 e OR = 0,74, $p < 10^{-2}$, respectivamente) localizadas nos genes *LENG8* e *FCAR*, respectivamente. Além disso, quatro haplótipos com SNPs dos genes *KIR3DL2/3*, *LAIR2* e *LILRB1* estiveram associados ($p < 0,05$). O mapeamento fino da região próxima a variante intergênica *rs461569* (22 kb) foi realizado por sequenciamento de 48 indivíduos. Os 45 SNPs detectados pelo sequenciamento foram avaliados quanto ao desequilíbrio de ligação e anotação funcional, sendo então selecionados quatro SNPs (*rs2619223*, *rs465169*, *rs681198* e *rs7249071)* para um novo estudo caso controle com amostras de PFE (Brasil) e PF esporádico (Alemanha). Todos quatro SNPs estiveram associados com PFE do Brasil (208 pacientes e 140 controles; $p < 0,05$) e somente os SNPs *rs2619223* e *rs681198* estiveram associados com PF esporádico da Alemanha (76 pacientes e 150 controles; $p < 0,05$). Análises *in silico* indicam que essas variantes associadas ao PF possuem um papel importante na regulação de genes da região 19q13. Este trabalho apontou a relevância desta região cromossômica como uma região muito importante na patofisiologia do PF.

Palavras-chave: LRC, 19q13, pênfigo foliáceo, *rs465169*, *lnc-IGFL2*

ABSTRACT

Pemphigus foliaceus (PF) is a rare bullous autoimmune disease of the skin characterized by the detachment of keratinocytes and the presence of autoantibodies against desmosomes, particularly anti-desmoglein 1 (DSG1). PF occurs sporadically around the globe, but its incidence is higher in some regions where PF is considered endemic (EPF). PF etiology is complex, with the contribution of environmental and genetic factors that are poorly known. Variation in genes related to immune responses has been associated with PF, including genes in the leucocyte receptor complex (LRC) region, located at 19q13.42. Moreover, many genes in the LRC were differentially expressed in PF patients, in diverse clinical subtypes, and in controls, which supports the hypothesis that variation within the LRC impacts PF pathogenesis. The aim of this study was deeply screening the extended LRC, 19q13, searching for genetic variants that could influence the susceptibility to PF. A total of 229 EPF patients and 194 controls were genotyped for 2,510 single nucleotide polymorphisms (SNP) of the extended LRC by DNA microarray. Nine SNP were associated with EPF ($p < 10^{-3}$). From these, *rs8103298* (OR = 2.50) and *rs3810265* (OR = 2.33) are in the *RPS9P4* pseudogene and four (*rs9304655*, *rs6509255*, *rs10412596*, *rs10439109*) were located within the long non-coding RNA (lncRNA) *lnc-IGFL2-1* gene (1.60 < OR < 2.2). Additionally, variants in the protein-coding genes *SULT2B1* (*rs11083944*, OR = 0.49), *SCAF1* (*rs2304208*, OR = 3.04), and *RDH13* (*rs1654431*, OR = 2.34) were also associated with EPF. Within only the LRC region, 19q13.42, three of the 176 SNP were associated with EPF: the intergenic *rs465169* (OR = 1.50, $p = 0.004$), and the *rs35336528* and *rs1865097* variants (OR = 3.39 e OR = 0.74, $p < 10^{-2}$, respectively) that are located in the *LENG8* and *FCAR* genes, respectively. In addition, we identified four haplotypes with SNPs within the *KIR3DL2/3*, *LAIR2* and *LILRB1* genes associated with EPF ($p < 0.05$). Fine mapping of the intergenic region that includes *rs461569* was performed by sequencing in 48 individuals. The 45 SNPs called on sequencing were analyzed among linkage disequilibrium and functional annotation, resulting in four SNPs (*rs2619223*, *rs465169*, *rs681198*, *rs7249071)* to perform a new case control study in EPF (Brazil) and sporadic PF (Germany) samples. All four SNPs were associated with EPF from Brazil (208 patients and 140 controls; $p < 0.05$), and *rs2619223* and *rs681198* were associated with sporadic PF from Germany (76 patients and 150 controls, $p < 0.05$). In silico analysis suggested that the variants associated with PF have an important role in regulation of genes within the region 19q13. Our study corroborates the importance of this chromosomic region as a hot spot for variants influencing PF pathogenesis.

Key-words: LRC, 19q13, pemphigus foliaceus, *rs465169*, *lnc-IGFL2*

LISTA DE FIGURAS

FIGURA 1: MANIFESTAÇÃO DE BOLHAS NO PÊNFIGO FOLIÁCEO (PF). 15
FIGURA 2: DISTRIBUIÇÃO DA EXPRESSÃO DAS CADERINAS NA EPIDERME. 17
FIGURA 3: MODELO DA ATIVAÇÃO DO PROCESSO DE ACANTÓLISE NO PÊNFIGO VULGAR. 18
FIGURA 4: LOCALIZAÇÃO CROMOSSÔMICA DE GENES QUE CODIFICAM RECEPTORES DAS CÉLULAS NK E SEUS LIGANTES 23
FIGURA 5: ESQUEMA DA DISPOSIÇÃO DOS GENES NA REGIÃO GENÔMICA 19q13.42, COMPLEXO DE RECEPTORES LEUCOCITÁRIOS (LRC). 24
FIGURA 6: ORGANIZAÇÃO DOS GENES DE IMUNORECEPTORES NO COMPLEXO DE RECEPTORES LEUCOCITÁRIOS (LRC). 25
FIGURA 7: ESTRUTURA E NOMENCLATURA DE MOLÉCULAS KIR. 26
FIGURA 8: ESQUEMA DA ORGANIZAÇÃO DOS GENES *LILR* E *LAIR*. 29
FIGURA 9: FAMÍLIA DE RECEPTORES SIGLEC EM HUMANOS. 32
FIGURE 10: RESULT OF THE ASSOCIATION ANALYSIS BETWEEN 2,510 SNPS AND *FOGO SELVAGEM*, AT 19q13 GENOMIC REGION 43
FIGURA 11: ESQUEMA DA LOCALIZAÇÃO GENÔMICA E TAMANHO DOS FRAGMENTOS AMPLIFICADOS POR PCR. 55

LISTA DE TABELAS

TABLE 1: RESULTS OF THE ASSOCIATION ANALYSIS OF SNPS ASSOCIATED WITH FOGO SELVAGEM. 41

TABLE 2: RESULTS OF THE ASSOCIATION ANALYSIS OF SNPS ASSOCIATED WITH SPORADIC PEMPHIGUS FOLIACEUS. 42

TABELA 3: PROTOCOLO DE PCR COM CONCENTRAÇÃO DOS REAGENTES POR AMOSTRA. 56

TABELA 4: SEQUÊNCIA DE OLIGONUCLEOTÍDEOS USADOS PARA AMPLIFICAÇÃO DE DOIS FRAGMENTOS (A E B) POR PCR 56

TABELA 5: PROTOCOLO DE PCR PARA AMPLIFICAÇÃO DOS DOIS FRAGMENTOS A E B POR PCR 56

TABELA 6: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DOS SNPS DA REGIÃO INTERGÊNICA DO LRC, DAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO (PF) E CONTROLES (CTR). 62

TABELA 7: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DOS SNPS NOS GENES *FCAR* E *LENG8*, DAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO E CONTROLES 63

TABELA 8: RESULTADOS DE NOVA ANÁLISE DE ASSOCIAÇÃO DOS SNPS LOCALIZADOS NA REGIÃO INTERGÊNICA DO LRC, NAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO. 64

TABELA 9: RESULTADOS DA CORRELAÇÃO PARA AVALIAR O DESEQUILÍBRIO DE LIGAÇÃO ENTRE OS SNPS NA REGIÃO INTERGÊNICA DO LRC NAS AMOSTRAS DE CONTROLES 64

TABELA 10: RESULTADOS DA ANÁLISE DE ASSOCIAÇÃO DOS SNPS LOCALIZADOS NOS GENES *FCAR* E *LENG8* NAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO 65

TABELA 11: RESUMO DOS POSSÍVEIS EFEITOS FUNCIONAIS DOS SNPS LOCALIZADOS NA REGIÃO INTERGÊNICA DO LRC 66

SUMÁRIO

1 INTRODUÇÃO 13
2 REVISÃO BIBLIOGRÁFICA 14
2.1 PÊNFIGO FOLIÁCEO 14
2.1.1 Acantólise no pênfigo 16
2.1.2 Imunopatogênese do PFE 18
2.1.3 Fatores ambientais e a etiologia do PF 20
2.1.4 Susceptibilidade genética ao PF 21
2.2 REGIÃO 19q13 E O LRC 22
2.2.1 Receptores de células killer semelhantes à imunoglobulina (KIR) 25
2.2.2 Receptores de leucócitos semelhantes à imunoglobulina (LILR) 27
2.2.3 Receptores associados a leucócitos semelhantes à imunoglobulina (LAIR) 29
2.2.4 Receptores de lectinas ligantes de ácido siálico semelhantes à imunoglobulina (SIGLECs) 30
2.2.5 Receptor para IgA FcαRI (FCAR, FcαRI ou CD89) 32
2.2.6 DAP10 e DAP12 32
2.3 JUSTIFICATIVA 33
3 OBJETIVO 34
3.1 OBJETIVOS ESPECÍFICOS 34
4 CAPÍTULO I 35
4.1 ABSTRACT 35
4.2 INTRODUCTION 36
4.3 METHODS 38
4.3.1 Patient and control population samples 38
4.3.2 DNA pemphigus samples genotyping 38
4.3.3 Selection of 19q13 region markers and statistical analysis 38
4.3.4 In silico analysis 39
4.4 RESULTS 40
4.5 DISCUSSION 44
4.6 CONCLUSION 48
4.7 REFERENCES 48
5 CAPÍTULO II 53
5.1 ABSTRACT 53

6 CAPÍTULO III 54
6.1 MATERIAIS E MÉTODOS 54
6.1.1 O mapeamento fino da região intergênica 54
6.1.2 Sequenciamento da região intergênica 55
6.1.3 Análise de correlação entre *rs465169* e outros SNPs localizados na RI 57
6.1.4 Estudo de associação com variantes localizadas no LRC (19q13) 58
6.1.4.1 Caracterização da amostra utilizada no novo estudo de associação 58
6.1.4.2 Variantes selecionadas para segunda genotipagem 59
6.1.4.3 Genotipagem 59
6.1.4.4 Análise estatística do estudo caso-controle 60
6.1.4.5 Análises *in silico* para impacto funcional dos SNPs 60
6.2 RESULTADOS 60
6.2.1 LRC e região intergênica (19q13.42) 61
6.2.2 Impacto funcional dos SNPs localizados na região intergênica (19q13.42) . 65
6.3 DISCUSSÃO 67
6.3.1 Mapeamento fino da região 19q13.42 67
6.3.2 Polimorfismos nos genes *LENG8* e *FCAR* 71
7 CONCLUSÕES 73
REFERÊNCIAS BIBLIOGRÁFICAS 74
APÊNDICE A – MATERIAL SUPLEMENTAR DO CAPÍTULO I 85
APÊNDICE B – DETALHES DE MATERIAIS E MÉTODOS 95
APÊNDICE C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PACIENTES COM PÊNFIGO E GRUPO CONTROLE 97
APÊNDICE D – FICHA DE AVERIGUAÇÃO PACIENTES DE PÊNFIGO 101
APÊNDICE E – FICHA DE AVERIGUAÇÃO GRUPO CONTROLE 106
APÊNDICE F – DESCRIÇÃO AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO E CONTROLES 109
APÊNDICE G - EFEITOS FUNCIONAIS DOS SNPS LOCALIZADOS NA REGIÃO 19q13.42 110

1 INTRODUÇÃO

Pênfigo é um grupo de doenças autoimunes nas quais ocorre a formação de bolhas na epiderme e há presença de autoanticorpos, principalmente contra as desmogleínas (DSG) 1 e/ou 3. Entre os principais tipos de pênfigo, o pênfigo foliáceo (PF) é caracterizado por bolhas nas camadas mais superficiais da epiderme (estrato córneo, lâmina granular) e o autoanticorpo mais prevalente é anti-DSG1. O PF ocorre de forma esporádica no mundo todo e de forma endêmica, com incidência aumentada, em regiões geográficas limitadas. No Brasil, o PF endêmico (PFE) é encontrado nas regiões sudeste e centro-oeste do País.

O PFE é uma doença multifatorial na qual há contribuição de fatores ambientais e genéticos parcialmente conhecidos. Entre os diversos fatores ambientais, a hipótese mais aceita sugere que a exposição contínua a picadas de insetos hematófagos, em indivíduos residentes na área endêmica e geneticamente predispostos, possa atuar como gatilho inicial para o desenvolvimento de autoanticorpos que reagem à DSG1, supostamente por mimetismo molecular. Com relação ao componente genético de susceptibilidade, polimorfismos em genes relacionados ao sistema imune, tais como genes *HLA* e de citocinas já foram associados com a doença. Dentre estes, é notável o número de genes e polimorfismos associados, bem como a concentração de genes diferencialmente expressos que estão localizados na região genômica 19q13. Nesta região cromossômica, localizam-se genes que codificam receptores celulares relacionadas ao sistema imune, pertencentes ao complexo de receptores leucocitários (LRC).

A partir disso, esta região foi alvo de um rastreamento através da técnica de genotipagem por microarranjo de DNA que permitiu descoberta de novos genes e polimorfismos envolvidos na susceptibilidade ao PF.

2 REVISÃO BIBLIOGRÁFICA

2.1 PÊNFIGO FOLIÁCEO

O pênfigo (de origem grega *pemphix*, que significa bolhas) é um grupo de doenças autoimunes órgão-específico que apresenta formação de bolhas na epiderme e mucosas, provocando desconforto ao paciente. Podem levar à morte em casos não diagnosticados e não devidamente tratados, ou refratários ao tratamento.

Entre os tipos de pênfigo, destacam-se o pênfigo vulgar e o pênfigo foliáceo. O pênfigo vulgar (PV) acomete a epiderme e/ou as mucosas. PV é o tipo mais comum de pênfigo com a incidência de 0,1-0,5/100.000 indivíduos por ano, sendo maior em judeus e no Japão (revisado por BAUM et al., 2014; BYSTRYN; RUDOLPH, 2005; JOLY; LITROWSKI, 2011). O pênfigo foliáceo (PF) acomete principalmente a epiderme e, epidemiologicamente pode ocorrer no mundo inteiro de forma esporádica, conhecido também como pênfigo de Cazenave, ou de forma endêmica (PFE) em certos países. Ambas as formas de PF possuem manifestações clínicas semelhantes (revisado em BAUM et al., 2014; BYSTRYN; RUDOLPH, 2005).

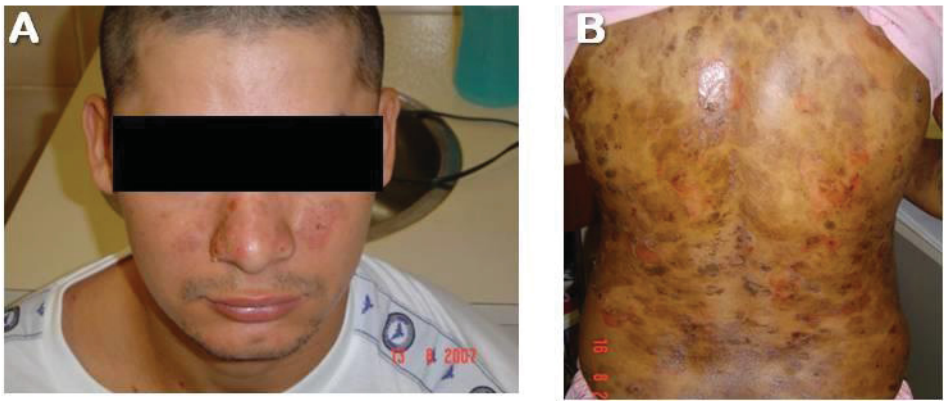
O PF esporádico possui incidência menor que 1 caso por milhão de habitantes por ano nos EUA e Europa. A prevalência do PF é muito maior nas regiões endêmicas, ocorrendo em países da América do Sul, como Colômbia, Peru e Brasil, e na África, como na Tunísia (revisado por JOLY; LITROWSKI, 2011; ORTEGA-LOAYZA et al., 2013). O PFE encontrado no Brasil também é conhecido como fogo selvagem (FS), pela sensação de ardência na pele, e ocorre predominantemente em áreas rurais dos estados do Paraná, São Paulo, Mato Grosso do Sul, Goiás e Minas Gerais (revisado em AOKI et al., 2015; DIAZ et al., 1989). Especificamente na reserva da tribo Terena em Limão Verde (MS), a prevalência de PFE é próxima de 3%, enquanto que El Bagre, na Colômbia, a prevalência é próxima a 5% (ABRÈU-VELEZ et al., 2003; JOLY; LITROWSKI, 2011; HANS-FILHO et al., 1996). Diferentemente, a incidência de PFE na Tunísia é de 6-7 casos/milhão de habitantes/ano, aumentando para 20 casos/milhão de habitantes/ano em mulheres adultas na região sul do País (MORINI et al., 1993; SELLAMI et al., 2004).

Homens e mulheres são acometidos pela doença em proporções semelhantes e o início do PFE ocorre em jovens e adultos de diversas origens étnicas entre 10 e 30 anos de idade (revisado em ABRÉU-VÉLEZ et al., 2010; CHAGAS et al., 2005;

LOMBARDI et al., 1992). A maioria dos pacientes de PFE mora em meio rural e relata alta frequência de insetos hematófagos em suas casas, principalmente o borrachudo (*Simulium nigrimanum*) (AOKI et al., 2004).

A principal característica clínica do PF é a formação de erosões e eritemas na epiderme, causadas por rupturas de bolhas superficiais frágeis. As erosões podem ser restritas em certas regiões do corpo (forma localizada) como na face, couro cabeludo e tórax, podendo progredir para diversos locais do corpo de forma generalizada (FIGURA 1) (HANS-FILHO et al., 1999; JOLY; LITROWSKI, 2011).

FIGURA 1: FORMAS DE MANIFESTAÇÃO DE BOLHAS CARACTERÍSTICAS DO PÊNFIGO FOLIÁCEO.



Fonte: acervo LGMH (2007).
Detalhe: A) Pênfigo foliáceo (PF) na forma localizada, apresentando eritema na face; B) PF na forma generalizada com bolhas em todo o corpo, em destaque a parte dorsal do tórax.

O diagnóstico de PF é baseado em diversas informações clínicas. O sinal de Nikolsky é um teste no qual se aplica uma pressão na pele aparentemente saudável para avaliar a adesão das células no tecido epidérmico. A condição histológica e de acantólise (desprendimento da camada superficial da pele) também podem ser analisadas em biópsias do paciente. E por fim, autoanticorpos anti-DSG1 são dosados no soro dos pacientes pela técnica de ELISA (BYSTRYN; RUDOLPH, 2005; JOLY; LITROWSKI, 2011).

O principal tratamento de PF é pela administração de corticosteroides como agentes anti-inflamatórios e imunossupressores, como ciclosporina e metotrexato, que inibem o sistema imunológico e a produção de anticorpos presentes em altos níveis na patologia (BRESSAN et al., 2010; SPINDLER et al., 2013). No entanto, a dose, o tipo de medicação e os efeitos adversos variam entre pacientes, por isso este tratamento não é bem aceito e, às vezes, interrompido. Outra abordagem de

tratamento vem crescendo, a utilização de anticorpo monoclonal anti-CD20, Rituximab, que promove a depleção de células B circulantes e, consequentemente, a produção de anticorpos (revisado em BYSTRYN; RUDOLPH, 2005; RUOCCO et al., 2013).

2.1.1 Acantólise no pênfigo

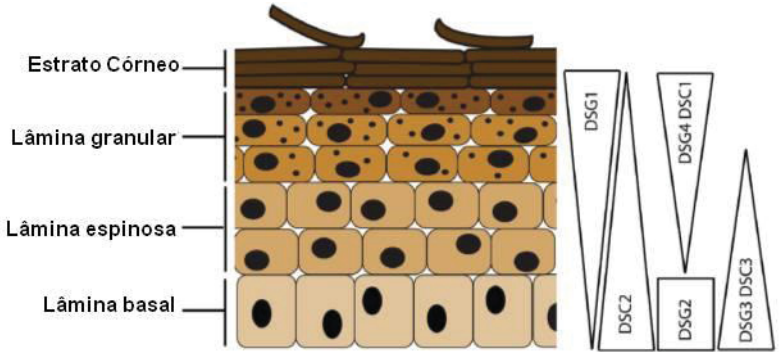
A formação de bolhas na epiderme é provocada pelo processo de acantólise, no qual ocorre a ruptura da adesão célula-célula entre os queratinócitos. A estrutura de adesão formada por proteínas e glicoproteínas pode ser desestabilizada por mutações em genes codificadores de proteínas desmossômicas ou por interação de autoanticorpos com esse complexo, comprometendo a integridade da epiderme e propiciando o aparecimento de bolhas na região (revisado em AL-JASSAR et al., 2013; BROOKE; NITOIU; KELSELL, 2012; DELVA; TUCKER; KOWALCZYK, 2009; ELLEBRECHT; PAYNE, 2014).

Entre as proteínas do desmossomo, as caderinas desmogleínas (DSG) e desmocolinas (DSC) possuem um padrão de expressão diferenciado ao longo da epiderme (FIGURA 2). O autoanticorpo predominante no soro de pacientes de PF é o anti-desmogleína 1 (anti-DSG1), sendo a DSG1 mais expressa em camadas superficiais da pele, local característico de ocorrência de bolhas (FIGURA 2). Por outro lado, no PV os principais autoanticorpos encontrados são anti-DSG3 e anti-DSG1 e as bolhas ocorrem em regiões mais profundas da epiderme e em mucosas, resultando em um quadro clínico mais grave que no PF (FIGURA 2) (AMAGAI et al., 1992; ARTEAGA et al., 2002; ISHII et al., 1997; OHATA; ISHII; FURUMURA, 2014; WARREN et al., 2003).

Apesar dos autoanticorpos reconhecerem as moléculas DSG no pênfigo, a formação de bolhas ocorre quando há presença de autoanticorpos contra ambas DSG1 e DSG3 nas camadas epiteliais onde as duas são coexpressas ou apenas contra DSG1 na região superficial da epiderme, em modelos animais (MAHONEY et al., 1999). Recentemente têm-se observado que a interação antígeno-anticorpo não é suficiente para o desprendimento epidérmico, porém, essa interação altera eventos de sinalização intracelular que podem induzir dissociação celular (VIELMUTH; WASCHKE; SPINDLER, 2015; WASCHKE et al., 2005). Além disso, o perfil de autoanticorpos possivelmente patogênicos parece não se restringir às DSGs, tanto

em PV quanto em PF (CHERNYAVSKY et al., 2007; FLORES et al., 2012; GRANDO, 2011).

FIGURA 2: DISTRIBUIÇÃO DA EXPRESSÃO DAS CADERINAS AO LONGO DAS CAMADAS DO TECIDO DA EPIDERME.

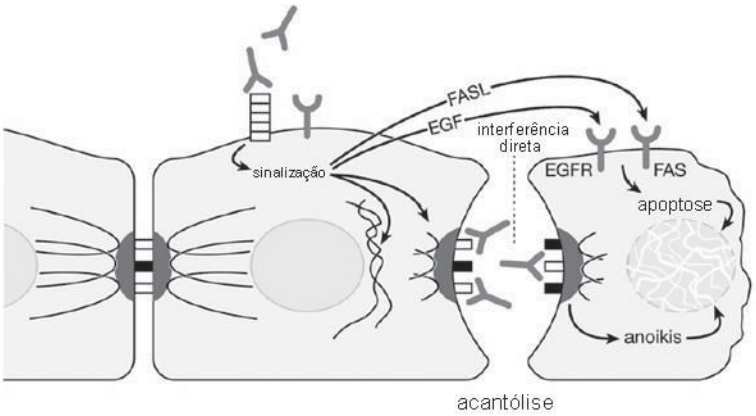


Fonte: adaptado de BROOKE; NITOIU; KELSELL (2012).
Detalhe: Os tipos (DSG1-4 e DSC1-3) e quantidades de proteínas expressas variam entre os diversos extratos da epiderme.

Grando et al. (2009) sugeriram que a formação de bolhas em PV é mais complexa que a supracitada. A hipótese seria que a junção de eventos de apoptose+acantólise (apoptólise) relacionaria vias de apoptose celular, contração das células basais e acantólise de células suprabasais na formação de bolhas em PV.

Neste contexto, a formação de bolhas ocorria através da ligação de autoanticorpos aos antígenos em PV, ativando receptores e sinais intracelulares através do fator de crescimento epidérmico (EGF, do inglês *epidermal growth factor*) que induz diversas vias de sinalização celular seguidas de cascatas de morte celular (apoptose). As proteínas celulares são clivadas e estimulam a produção de autoanticorpos secundários, seguido de acantólise e morte celular (FIGURA 3) (AMAGAI; STANLEY, 2011; BEKTAS et al., 2013; CHERNYAVSKY et al., 2007; ELLEBRECHT; PAYNE, 2014; FRUŠIĆ-ZLOTKIN et al., 2009; GRANDO et al., 2009; SCHMIDT; WASCHKE, 2009; SHARMA; MAO; PAYNE, 2007). No entanto, alguns estudos não encontraram evidências de apoptose no processo acantolítico no pênfigo (JANSE et al., 2014; SCHMIDT; WASCHKE, 2009).

FIGURA 3: MODELO ESQUEMÁTICO DA ATIVAÇÃO DO PROCESSO DE ACANTÓLISE QUE OCORRE NO PÊNFIGO VULGAR.



Fonte: adaptado de SCHMIDT; WASCHKE (2009).
Detalhe: A ligação de autoanticorpos aos receptores de queratinócitos estimula a transdução de sinais intracelulares (via EGF, FASL) que promovem a apoptose e a perda de adesão celular, resultando no processo de acantólise característico no pênfigo.

Os mecanismos moleculares da formação de anti-DSG1 e quais antígenos ativam formação desses autoanticorpos são desconhecidos no PFE, porém acredita-se que DSG1 epidérmico e antígenos exógenos compartilhem epítopos que provocam reação cruzada de anticorpos e desencadeiam quadro clínico de PFE (LI et al., 2003; AOKI et al., 2004).

2.1.2 Imunopatogênese do PFE

O pênfigo é uma doença autoimune na qual há produção de autoanticorpos e atuação de diversas respostas de células do sistema imune. Ambas as células B e T podem participar no gatilho inicial da doença, incluindo a atuação das células T ao estimular a maturação de células B durante o processo de produção de autoanticorpos (revisado em AMAGAI et al., 2000; GRANDO, 2011; PAN; ZHU; XU, 2015).

A perda da tolerância de células T tem sido demonstrada como um fator importante no desenvolvimento da doença (AMAGAI et al., 2000). As células T CD4+, que expressam moléculas HLA-DR, possivelmente reconhecem epítopos específicos do ectodomínio de DSG1 em pacientes com PFE e de DSG3 em pacientes com PV (HANS-FILHO et al., 1999; HERTL et al., 1998; VELDMAN; FELICIANI, 2008). Como resposta dessa interação, citocinas características de resposta T auxiliares 2 (Th2, Th do inglês *helper*) são produzidas, como interleucinas (IL)-4, IL-5 e IL-6, sendo a IL-4 moduladora de reposta IgG nos pacientes de PF (AOKI et al., 2004; LIN et al., 2000).

As células B evidentemente possuem um importante papel no desenvolvimento de PF, uma vez que apresentação de antígeno DSG1 pelas células B às células T, que possivelmente escaparam da seleção anti-DSG, provoca expansão clonal e autoreatividade que levaria ao desenvolvimento da autoimunidade em PF (YAMAGAMI et al., 2009). Em PFE, a resposta anti-DSG através da interação com antígeno ocorreria tanto na fase clínica quanto pré-clínica (QIAN et al., 2009). Assim, antígenos presentes nas áreas endêmicas de PFE podem induzir a produção de autoanticorpos antes do início da doença e, consequentemente, seriam importantes no início e manutenção do PF.

As relações e regulação de células T auxiliares Th1, Th2, Th17 e Treg no pênfigo não estão bem esclarecidas. O balanço da regulação entre Th17/Treg é crítico no desenvolvimento de doenças autoimunes; no pênfigo, estas células tem uma correlação negativa entre si (revisado em PAN; ZHU; XU, 2015; XU et al., 2013).

Além das células T e B, alterações nas NK podem contribuir com o aumento do risco de autoimunidade. Elas estão em menor número na corrente sanguínea em uma variedade de doenças autoimunes e inflamatórias, tais como esclerose sistêmica, diabetes tipo I, artrite reumatoide e PV (PERRICONE et al., 2008; SCHLEINITZ et al., 2010; ZAKKA et al., 2011).

As células NK também estão menor número em pacientes não tratados de PV e penfigoide bolhoso que indivíduos controle (GRANDO et al., 1989). A proliferação de NK é maior quando há peptídeos de DSG3 e altos níveis de IL-6, IL-8 e INF-γ em PV quando comparados aos controles. Isso indica que as células NK podem participar tanto na apresentação de antígeno quanto na ativação de células T CD4+ para produção de citocinas pró-inflamatórias, respectivamente (STERN et al., 2008). O papel das NK na pele ou em doenças autoimunes da pele ainda não está bem esclarecido, porém sabe-se que elas interagem com células T CD4+, levando a respostas do tipo Th2 e aumentando a expressão de IL-5 em PV (revisado em ZAKKA *et al*., 2012; von BUBNOFF *et al*., 2010). O controle fino e complexo para a ativação de células NK pode ser rompido quando há aumento no número dessas células, como observado em pacientes de PV quando comparado aos controles (TAKAHASHI et al., 2006).

PASPARAKIS, HAASE e NESTLE (2014) reuniram informações sobre a resposta imunológica na pele e como a sua desregulação pode levar à inflamação do local e à participação e/ou presença de diversas células no tecido, mostrando a

complexidade da resposta imune na pele. A infiltração de células inflamatórias, como os neutrófilos, participaria na patologia de PF e PV. Os neutrófilos participam de alguma via não usual na localização de acantólise em PF, e sabe-se que eles liberam enzimas proteolíticas que poderem danificar o tecido infiltrado (OHATA; ISHII; FURUMURA, 2014).

2.1.3 Fatores ambientais e a etiologia do PF

O PFE é uma doença de etiologia multifatorial, tantos fatores ambientais quanto genéticos influenciam no início e desenvolvimento da doença. O aumento da prevalência da doença em certas regiões do Brasil sugere que a exposição a um ou mais fatores ambientais desconhecidos nestas regiões estejam envolvidos com a patogênese do pênfigo.

Os pacientes de PFE moram, ou moraram por grande parte da vida, em regiões rurais próximos a rios e lagos e infestados por diversos insetos, entre eles borrachudos (*Simulium sp*.) (AOKI et al., 2004; DIAZ et al., 1989). Os agricultores expostos às picadas do borrachudo são 4,7 vezes mais susceptíveis a desenvolver PFE em comparação aos indivíduos não expostos (LOMBARDI et al., 1992). Além disso, alta incidência de novos casos de PFE ocorre no final da época chuvosa que coincide com as condições ótimas de reprodução e proliferação de insetos (DIAZ et al., 1989). Neste sentido, há uma hipótese de que na saliva de insetos hematófagos haveria antígenos que reagiriam de forma cruzada com a DSG1, desencadeando a produção de anticorpos anti-DSG1 e o quadro clínico em indivíduos geneticamente susceptíveis (CULTON et al., 2008; LI et al., 2003; LOMBARDI et al., 1992).

Cabe salientar que os indivíduos que vivem em regiões endêmicas de PFE são expostos a doenças tropicais (leishmaniose, doença de Chagas e oncocercose) e podem apresentar anticorpos aos patógenos das mesmas. A produção de autoanticorpos anti-DSG1 poderia ser em resposta ao antígeno de um único vetor ou aos três vetores que indiretamente levaria ao desenvolvimento de PFE (DIAZ et al., 2004).

A hipótese de mimetismo molecular se fortalece com o fato da redução de casos de PFE em alguns estados no Brasil devido à urbanização, alterando a exposição de agentes etiológicos associados ao desenvolvimento de PFE, como o borrachudo (AOKI et al., 2015; DO CHIOSSI; ROSELINO, 2001; LI et al., 2003).

Outros possíveis fatores ambientais de risco ao PF são o mercúrio, a radiação ultravioleta, infecções virais e bacterianas e solventes orgânicos (ABRÉU-VÉLEZ et al., 2010).

Em adição, infecções bacterianas ou virais já foram associadas a doenças autoimunes, como lúpus eritematoso sistêmico e artrite reumatoide (revisado em RUOCCO et al., 2014; SAMARKOS; VAIOPOULOS, 2005). Agentes infecciosos podem causar doenças autoimunes por diversos mecanismos. Um dos mecanismos mais importantes é a presença de um antígeno específico produzido ou liberado pelo patógeno que possuiria um papel central no desenvolvimento da autoimunidade, superantígenos ou mimetismo molecular do antígeno (revisado por RUOCCO et al., 2014; SAMARKOS; VAIOPOULOS, 2005).

A detecção de vírus da hepatite B (HBV), vírus da hepatite C (HCV), *H. pylori*, *Toxoplasma gondii* e citomegalovírus (CMV) é maior em soro de pacientes com doenças autoimunes bolhosas do que em controles, sugerindo um papel destas infecções no desenvolvimento destas doenças em indivíduos geneticamente susceptíveis (revisado em SAGI et al., 2011). Recentemente, foi associada a exposição a *H. pylori* e *S. stercoralis* (nematodo) em soro pacientes com PV em relação aos controles (MORTAZAVI et al., 2015). Logo, ainda não está claro como e se as infecções poderiam levar ao desenvolvimento do pênfigo.

2.1.4 Susceptibilidade genética ao PF

Diversos estudos demonstram a contribuição genética na susceptibilidade ao PF. Os primeiros polimorfismos genéticos que foram associados ao PF estão localizados no complexo principal de histocompatibilidade (MHC, do inglês *major histocompatibility complex*), principalmente os genes *HLA* (antígenos leucocitários humanos, do inglês *human leucocyte antigens*) (BROCHADO et al., 2016; PAVONI et al., 2003; PETZL-ERLER; SANTAMARIA, 1989).

Outros polimorfismos de genes que codificam moléculas participantes da resposta imune também foram associados com o PF, tais como genes que codificam citocinas (PEREIRA et al., 2004), moléculas coestimuladoras de células B (MALHEIROS; PETZL-ERLER, 2009), coreceptores e ligantes de células T (DALLA-COSTA et al., 2010), receptores de células *natural killer* (NK) semelhantes a imunoglobulina (AUGUSTO et al., 2012), receptores de células NK semelhantes a

lectina (CIPOLLA et al., 2016), receptores associados a leucócitos semelhantes a imunoglobulina (CAMARGO; AUGUSTO; PETZL-ERLER, 2016) e moléculas do complemento (BUMILLER-BINI et al., 2018; SALVIANO-SILVA; PETZL-ERLER; BOLDT, 2017).

Em um estudo de expressão gênica com microarranjos (*microarrays*) foi avaliado o padrão de expressão global em células T (CD4+) de pacientes com PFE localizado e generalizado sem tratamento, pacientes sob tratamento imunossupressor e controles sem a doença. Neste estudo, foi verificado que as regiões cromossômicas 12p13 e 19q13 concentram uma maior proporção de genes diferencialmente expressos em comparação a outras regiões genômicas (MALHEIROS et al., 2014).

Na região 19q13, foram localizados 9 (7,4 %) genes dos genes diferencialmente expressos entre pacientes com PFE sem tratamento na forma generalizada e controles e 8 (12,3 %) genes dos genes diferencialmente expressos entre pacientes de PFE com a forma generalizada e localizada, ambos sem tratamento (MALHEIROS et al., 2014).

No momento está sendo realizado o estudo de associação de polimorfismos genéticos do genoma completo (GWAS, do inglês *genome wide association studies*) pela técnica de microarranjo de DNA de SNPs (polimorfismo de nucleotídeo único, do inglês *single nucleotide polymorphism*) localizados em éxons e regiões reguladoras. Esse estudo permitirá detectar novas variantes envolvidas na susceptibilidade ao PF. Os dados referentes a região genômica 19q13, que é a região alvo de estudo desta tese, foram extraídos desse ensaio para a análise de associação caso-controle.

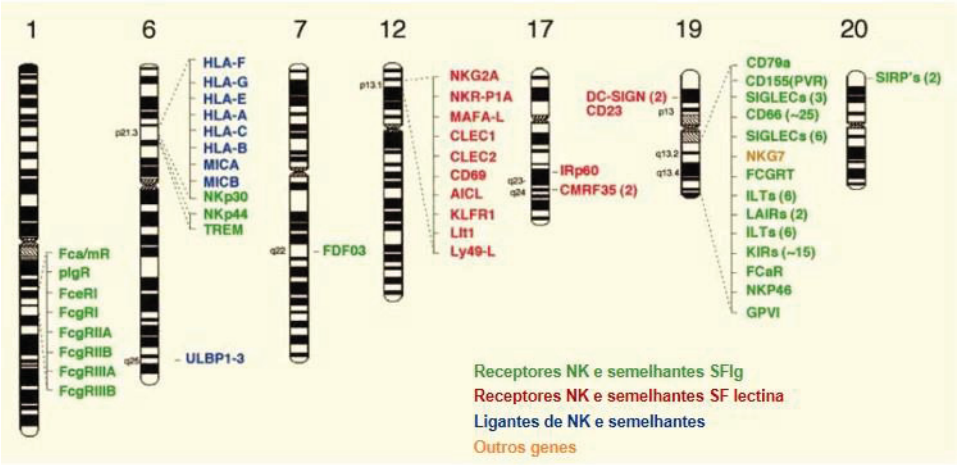
2.2 REGIÃO 19q13 E O LRC

O cromossomo 19 é um cromossomo pequeno, porém com grande concentração de genes e pseudogenes quando comparado com os demais cromossomos no genoma humano. Há também grande concentração de famílias gênicas e duplicação gênica cujos produtos possuem funções biológicas similares. Isso também é percebido no braço longo do cromossomo 19, especificamente na região 19q13. Além dos genes pertencentes ao complexo de receptores leucocitários (LRC, do inglês *leucocyte receptor complex*), nesta região estão localizados diversos fatores de transcrição tipo ZNF (do inglês *zinc finger transcription factor*), genes da

subfamília do citocromo P450 (CYP2, do inglês *cytochrome P450 subfamily II genes*) e família de protease serina KLK (do inglês *kallikrein*) (GRIMWOOD et al., 2004).

Na sub-região 19q13.4 localiza-se o LRC que contém genes que codificam moléculas pertencentes à superfamília das imunoglobulinas (SFIg) (KELLEY; WALTER; TROWSDALE, 2005; WENDE et al., 1999). Além dos receptores de NK serem codificados por genes localizados na região 19q13.4, o complexo *natural killer* (NKC, do inglês *natural killer complex*) contém genes que codificam lectinas do tipo C concentrados na região 12p13.1 (FIGURA 4) (BROWN et al., 1997; YOKOYAMA; PLOUGASTEL, 2003). O objeto de estudo deste trabalho são os genes e variantes localizados na região 19q13.

FIGURA 4: ESQUEMA DA LOCALIZAÇÃO CROMOSSÔMICA DE ALGUNS GENES QUE CODIFICAM RECEPTORES DAS CÉLULAS NK E SEUS LIGANTES EM HUMANOS.



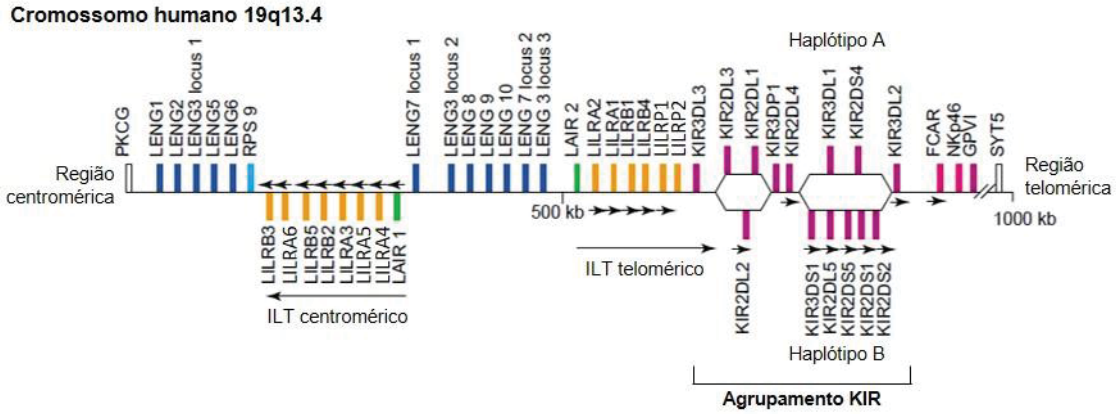
Fonte: Adaptado de TROWSDALE *et al.*, 2001.
Detalhe: Os genes pertencentes à superfamília gênica de imunoglobulina (SFIg) e da lectina (SF lectina) estão em verde e vermelho, respectivamente. Genes dos ligantes HLA de classe I estão em azul.

O LRC humano tem aproximadamente 1 Mpb (mega, 10^6 pb) de extensão e contém 45 genes que codificam moléculas de superfície celular participantes de respostas imunes. Destes, 30 codificam receptores da SFIg que, por sua vez, podem ser agrupados em famílias gênicas baseados na organização gênica, filogenia e estrutura (MARTIN et al., 2002) (FIGURA 5). Na região 19q13, além do LRC, situa-se o LRC estendido (FIGURA 6).

No LRC, 19q13.4, estão localizadas as famílias gênicas receptores de células *killer* semelhantes à imunoglobulina (KIR, do inglês *killer-cell immunoglobulin-like receptor*), receptores de leucócitos semelhantes à imunoglobulina (LILR ou ILT, do inglês *leucocyte immunoglobulin-like receptors*), os receptores associados a

leucócitos semelhantes à imunoglobulina (LAIR, do inglês *leucocyte-associated immunoglobulin-like receptors*) (FIGURA 5) (revisado em BARROW; TROWSDALE, 2008; TROWSDALE et al., 2015).

FIGURA 5: ESQUEMA DA DISPOSIÇÃO DOS GENES SITUADOS NA REGIÃO GENÔMICA 19q13.42, CORRESPONDENTE AO COMPLEXO DE RECEPTORES LEUCOCITÁRIOS (LRC).



Fonte: Adaptado de Martin (2002).
Legenda: LRC: complexo de receptores leucocitários (do inglês *leucocyte receptor complex*)

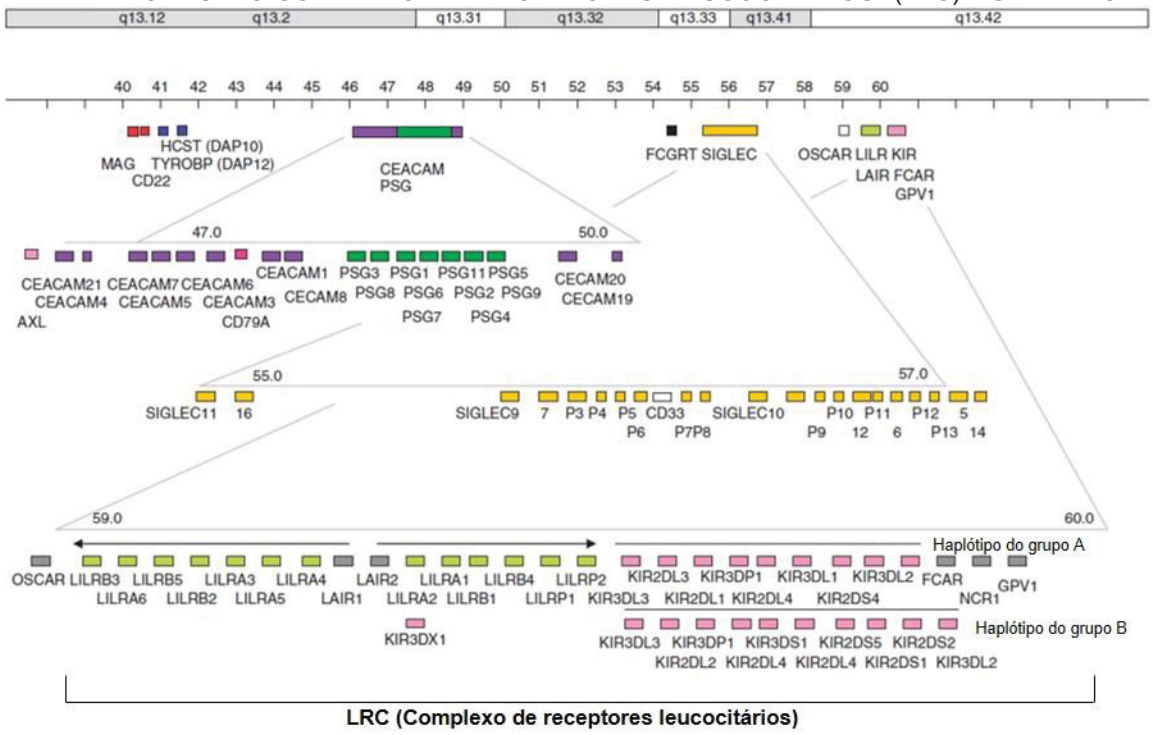
Na região centromérica ao LRC, denominada LRC estendido, localizam-se genes que codificam *lectinas ligantes ao ácido siálico semelhante à imunoglobulina* (SIGLEC, do *inglês sialic acid binding immunoglobulin-like lectins*). Nesta região, também há genes que codificam moléculas da família CD66 ou moléculas de adesão celular relacionada ao antígeno carcinoembrionário (CEACAM, do inglês *carcinoembryonic antigen-related cell adhesion molecule*) e glicoproteína específica da gestação (PSG, do inglês *pregnancy-specifc glycoprotein*) (FIGURA 6) (revisado em BARROW e TROWSDALE 2008).

Na região telomérica do LRC estão os genes para o receptor de citotoxicidade natural 1 (NCR1 ou NKp46, do *inglês natural cytotoxicity-triggering receptor 1*), receptor do segmento Fc de IgA (FCAR ou CD89, do inglês *IgA Fc fragment receptor*), gilcoproteína VI (GP6, do inglês *platelet glycoprotein VI*) cujos ligantes são os mais diversos como imunoglobulinas, hemaglutininas virais e colágeno (FIGURA 6) (revisado em TROWSDALE *et al.*, 2015; KELLY, WALTER, TROWSDALE, 2005; BARROW e TROWSDALE 2008).

Outros genes localizados no LRC codificam o receptor associado ao osteoclasto (OSCAR, do inglês *osteoclast-associated receptor*), receptor de Fc gama (FcγR, do inglês *Fc gama receptor*) e genes que codificam as moléculas adaptadoras DAP10 ou HCST (do inglês *hematopoietic cell signal transducer*) e DAP12 ou

TYROBP (do inglês *protein tyrosine kinase binding protein*) (FIGURA 6) (revisado em BARROW e TROWSDALE, 2008; TROWSDALE *et al*., 2015).

FIGURA 6: ESQUEMA DA ORGANIZAÇÃO DOS GENES DE IMUNORECEPTORES E ADAPTADORES NO COMPLEXO DE RECEPTORES LEUCOCITÁRIOS (LRC) ESTENDIDO.



Fonte: Adaptado de BARROW e TROWSDALE, 2008.
Legenda: LRC: complexo de receptores leucocitários (do inglês *leucocyte receptor complex*)

Neste trabalho investigamos se polimorfismos localizados na região LRC estendido, 19q13 (Capítulo I) e restrita ao LRC, 19q13.42 (Capítulo II) modulam a susceptibilidade ao PF.

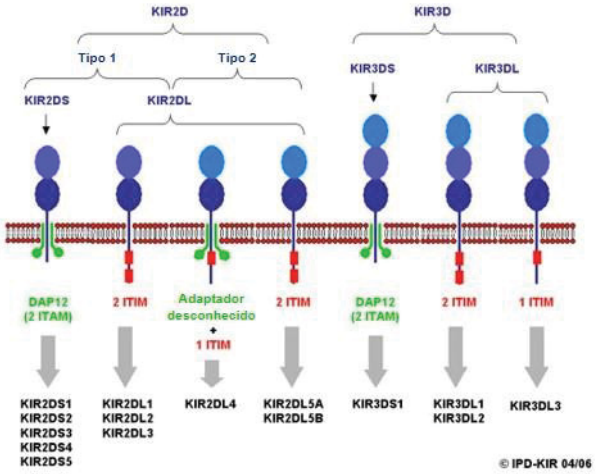
2.2.1 Receptores de células killer semelhantes à imunoglobulina (KIR)

Um dos conjuntos de receptores mais polimórficos e estudados das células NK é o KIR. A nomenclatura dos KIR está baseada em sua estrutura (FIGURA 7). Essas moléculas apresentam dois ou três domínios extracelulares (2D e 3D) que reconhecem as moléculas HLA-A, B ou C, uma porção transmembranar (TM) e uma cauda intracitoplasmática que pode ser longa e inibidora (L, *long*) com um motivo inibidor baseado em tirosina (ITIM, do inglês *immunoreceptor tyrosine-based inhibitory motif*) ou curta e ativadora (S, *short*) que interage com uma proteína adaptadora com motivo ativador baseado de tirosina (ITAM, do inglês *immunoreceptor tyrosine-based*

activation motif), como o DAP12 ou FcRγ (FIGURA 7) (LONG *et al*., 1996; ANDRE *et al*., 2001; HSU *et al*., 2002). O receptor KIR2DL4 possui tanto ITIM como ITAM em sua cauda intracitoplasmática longa, porém a sinalização emitida por este receptor é predominante ativador (KIKUCHI-MAKI *et al*., 2003).

Os principais ligantes de KIR são moléculas de HLA de classe I, e a sua interação baseia-se na hipótese da falta do próprio (do inglês *missing-self*). Se o receptor inibidor não interage com o seu ligante HLA, pois este está ausente na célula-alvo, não há envio do sinal inibidor o que permite a ativação das células NK pelo predomínio de sinalizações dos receptores ativadores (RAULET, 2006; PARHAM, 2006; VILCHES; PARHAM, 2002).

FIGURA 7: ESTRUTURA E NOMENCLATURA DE MOLÉCULAS KIR.



Fonte: Adaptado de IPD (https://www.ebi.ac.uk/ipd/kir/).

Detalhes: Os domínios extracelulares dos receptores de células killer semelhantes à imunoglobulina (KIR), KIR2D e KIR3D, estão representados na porção superior da figura e as caudas citoplasmáticas na porção inferior da figura. Caudas longas (L) e curtas (S) e o número de domínios extracelulares definem o nome do gene. Molécula ITIM (*immunotyrosine-based inhibitory motif*), em vermelho, se liga a cauda citoplasmática longa e molécula ITAM (*immunotyrosine-based activatory motif*), em verde, se liga a cauda citoplasmática curta.

A família de genes *KIR* é composta por 14 genes e 2 pseudogenes. No entanto o número de genes *KIR* varia de acordo com autor, pois não é claro distinguir o que seria gene individualmente ou alelos do mesmo gene. Por exemplo, *KIR2DL2* e *KIR2DL3* e *KIR3DS1* e *KIR3DL1* foram considerados inicialmente dois genes distintos, mas atualmente são considerados alelos do mesmo gene. Por outro lado, o *KIR2DL5* foi separado em *KIR2DL5A* e *KIR2DL5B* (MIDDLETON; GONZALEZ 2009).

Os haplótipos dos genes *KIR* são divididos em haplogrupos A e B, os quais diferem quanto ao número e a variedade de genes presentes. No entanto, os quatro

genes de moldura (*KIR3DL3*, *KIR3DP1*, *KIR2DL4* e *KIR3DL2*) estão presentes em ambos os grupos (HSU et al., 2002).

O haplogrupo A é formado por haplótipos que contêm genes que codificam os receptores inibidores KIR2DL1, KIR2DL3 e KIR3DL1 e os receptores ativadores KIR2DS4 e KIR2DL4 (gene moldura). Este haplogrupo é encontrado em diversas populações e a variação no número de cópias (CNV, do inglês *copy number variation*) é menor que a do haplogrupo B.

O haplogrupo B é formado por haplótipos com maior quantidade e diversidade de receptores ativadores (SINGH et al., 2012; KULKARNI; MARTIN; CARRINGTON, 2008; CARRINGTON; NORMAN, 2003). A diversidade deste haplogrupo também está relacionada com a variabilidade de frequências de CNV nas populações (VUKCEVIC et al., 2015) e pelos os receptores inibidores serem mais polimórficos do que os receptores ativadores (SANCHEZ-MAZAZ et al., 2011; PYO et al., 2010).

Além da alta variabilidade quanto a presença e ausência dos segmentos gênicos, alélica, número de cópias e da composição em haplótipos dos genes *KIR*, a expressão dos receptores KIR ocorre de forma variegada e aleatória entre os clones das células NK (revisado em TROWSDALE et al., 2015).

A diversidade alélica dos genes *KIR* pode influenciar a interação dos KIR com seus ligantes HLA-I, os quais também são polimórficos (revisado em TROWSDALE et al., 2015; PARHAM, 2005). Assim, essa grande variabilidade na interação entre KIR e HLA-I pode modular o limiar de ativação nas células NK e permitir que baixos níveis de interação receptor/ligante sejam suficientes para ativar estas células (CAMPBELL; PURDY, 2010; UHRBERG, 2005). A diversidade genética do conjunto de receptor-ligante KIR-HLA já foi associada à susceptibilidade a doenças virais, câncer e doenças autoimunes, incluindo o PFE (revisado em AUGUSTO e PETZL-ERLER, 2015; TROWSDALE et al., 2015; AUGUSTO et al., 2012).

2.2.2 Receptores de leucócitos semelhantes à imunoglobulina (LILR)

Os receptores de leucócitos semelhantes à imunoglobulina (LILR, ILT, LIR, CD85) são pertencentes à SFIg e altamente relacionados entre si. Seus genes são adjacentes ao loco de *KIR*. LILRs são divididos em dois grupos LILRA e LILRB. Os LILRB (LILRB1-B5) possuem funções inibidoras pela presença de motivo ITIM na cauda citoplasmática. Os LILRA possuem funções ativadoras pela associação com

moléculas adaptadoras portadoras de ITAMs. LILRA3 é o único membro da família que é solúvel, porém alguns transcritos alternativos dos demais genes *LILR* codificam moléculas solúveis que podem inibir a interação de LILR com HLA-I e, por sua vez, respostas inibidoras (TROWSDALE et al., 2015; BARROW; TROWSDALE, 2008, JONES et al., 2009).

A maioria dos membros da família possui quatro domínios semelhantes à imunoglobulina (D1, D2, D3 e D4), sendo que LILRB4 e B5 possuem apenas 2 domínios (D1 e D2) (revisado em TROWSDALE et al., 2015; CHEN et al., 2009).

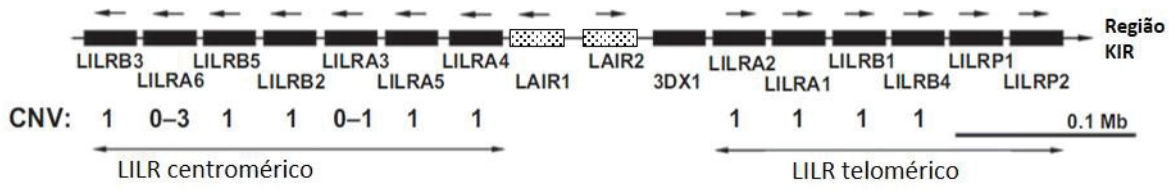
Os receptores inibidores LILRB1 e LILRB2 e receptores ativadores LILRA1 e LILRA2 e o receptor solúvel LILRA3 se ligam a moléculas de HLA de classe I. Interessantemente, LILRB1 e LILRB2 interagem fortemente com a molécula HLA-G em detrimento as outras moléculas HLA-I, participando possivelmente de respostas inibidoras importantes mediadas por HLA-G. Os receptores LILRA4-A6 e B3-B5, os quais possuem menos que 60% de identidade de sequência com os outros receptores, possuem ligantes ainda desconhecidos e não-HLA-I (revisado em TROWSDALE et al., 2015; CHEN et al., 2009).

Em humanos as moléculas de LILR são predominantemente expressas em células mieloides, embora LILRB1 também seja expresso em células B, algumas células NK e subtipos de células T. Estas moléculas de superfície regulam atividades celulares através do balanço de sinais ativadores e inibidores e reconhecem moléculas de HLA-I de acordo com a hipótese da falta do próprio de forma semelhante aos KIR (revisado em TROWSDALE et al., 2015; LOPEZ-ALVAREZ et al., 2014).

A sinalização dos receptores LILR está envolvida com o perfil de citocinas e expressão de moléculas co-estimuladoras em monócitos, macrófagos e células dendríticas que modulam a resposta adaptativa do sistema imune. A expressão de LILR é fortemente influenciada pela resposta de sinalização de outros receptores, como a ativação dos receptores *toll-like* (TLR, do inglês *toll-like receptor*) em monócitos, resultando na baixa expressão de LILRB1, LILRB4, LILRA2 e LILRA4 (revisado em TROWSDALE et al., 2015; BLEHARSKI et al., 2003; JU et al., 2004).

O complexo gênico de *LILR* consiste em 11 genes e 2 pseudogenes agrupados, indicando evento de duplicação e inversão em larga escala (FIGURA 8) (revisado em TROWSDALE et al., 2015).

FIGURA 8: ESQUEMA DA ORGANIZAÇÃO DOS GENES *LILR* E *LAIR*.



Fonte: Adaptado de TROWSDALE et al., 2015.
Detalhes: Os genes receptores de leucócitos semelhantes à imunoglobulina (LILR) e dos receptores associados a leucócitos semelhantes à imunoglobulina (LAIR) estão localizados no complexo de receptores leucocitários (LRC, do inglês *leucocyte receptor complex*). As setas menores com sentido único indicam a orientação do gene, as flechas maiores indicam a porção centromérica (esquerda) e telomérica (direita). CNV: variação no número de cópias.

A maioria dos genes *LILR* é pouco polimórfica na região codificadora. Os genes *LILRB3* e *LILRA6* têm muitos polimorfismos não sinônimos. Esses genes codificam receptores homólogos com alta similaridade de sequência dos domínios extracelular, porém sua sinalização intracelular é oposta: LILRB3 é receptor inibidor enquanto LILRA6 ativador (JONES et al., 2016; BASHIROVA et al., 2014; LOPEZ-ALVAREZ et al., 2014).

LILRB5 é o gene mais polimórfico com mais de 30 alelos. O *LILRA3* é o único gene da família *LILR* que tem alelos funcionais e não funcionais, apresentando uma deleção de 6,7 Kb que é mais frequente em japoneses (revisado em TROWSDALE et al., 2015; BARROW; TROWSDALE, 2008; HIRAYASU et al., 2006).

Os receptores LILR participam da resposta imunológica de infecções virais e bacteriana (*M. leprae*) e também reconhecem moléculas β-amiloide (Aβ) presentes na doença de Alzheimer. Polimorfismos do *LILRA3* foram associados a doenças autoimunes, porém alguns resultados necessitam ser confirmados (revisado em TROWSDALE et al., 2015). Como estes genes *LILR* estão em desequilíbrio de ligação com o loco *KIR*, há dificuldade em determinar efeitos exclusivos dos polimorfismos dos genes *LILR* na susceptibilidade a doenças (revisado em BARROW; TROWSDALE, 2008).

2.2.3 Receptores associados a leucócitos semelhantes à imunoglobulina (LAIR)

LAIR-1 é uma molécula de membrana com apenas um domínio extracelular semelhante à imunoglobulina (Ig) com dois motivos ITIMs na cauda citoplasmática. Seu homólogo, o receptor LAIR-2 possui 84% de similaridade de sequência com LAIR-1, porém é solúvel por não possuir a porção transmembranar. LAIR-1 é expresso na

superfície de quase todas as células do sistema imune, como células NK, células B e T, monócitos, com efeito forte inibidor que pode limitar a atividade de células NK e citotoxicidade de células T, por exemplo (revisado por BARROW; TROWSDALE 2008; MEYAARD et al., 1997; MEYAARD et al., 1999).

O LAIR-1 se liga a diversas moléculas de colágeno e o resultado é a possível inibição da ativação de células imunes primárias (LEBBINK *et al*., 2006). Experimentos sugerem que colágenos presentes na matriz extracelular possam regular a atividade de células imunes *in vivo* e que as células ativadas migram aos tecidos com alguma lesão ou trauma. Os ligantes de LAIR-2 também são colágenos e, como é um receptor solúvel, LAIR-2 pode agir como regulador negativo da sinalização de LAIR-1 por sequestrar o seu local de interação com colágeno (revisado por BARROW; TROWSDALE, 2008; MEYAARD, 2008).

O gene *LAIR-1* possui 10 éxons e *LAIR-2* possui 5 éxons (MEYAARD, 2008). O gene *LAIR-1* é homólogo ao gene *LAIR-2* que é transcrito na orientação oposta à sua (genes hachurados na FIGURA 8). A região LILR/LAIR passou por um evento de duplicação e inversão na região central em larga escala, logo se observa que os genes *LAIR* estão circundados pelos genes *LILR* (FIGURA 8) (revisado em BARROW; TROWSDALE, 2008; MEYAARD 2008; WENDE; VOLZ; ZIEGLER, 2000).

Polimorfismos de ambos os genes *LAIR-1* e *LAIR-2* foram associados a susceptibilidade ao PFE pelo nosso grupo de pesquisa. Um haplótipo composto por 4 SNPs do gene *LAIR-2*, associado ao aumento de seus níveis de RNAm, foi fortemente associado com o risco de desenvolver PFE (OR = 4,02, p = 0,0008) (CAMARGO; AUGUSTO; PETZL-ERLER, 2016).

2.2.4 Receptores de lectinas ligantes de ácido siálico semelhantes à imunoglobulina (SIGLECs)

Os receptores de lectinas ligantes de ácido siálico semelhantes à Ig (Siglecs) reconhecem ácido siálico como ligante, apresentando 15 tipos diferentes de receptores em humanos. Siglecs são proteínas de superfície com um único domínio do tipo IgV na porção N-terminal, seguido de número variável de domínios IgC2. A maioria dos Siglecs possui um ou mais motivos inibidor ITIM em sua cauda citoplasmática. Estes receptores são expressos predominantemente em células do

sistema hematopoiético (revisado em BARROW; TROWSDALE, 2008; McMILLAN; CROCKER, 2008).

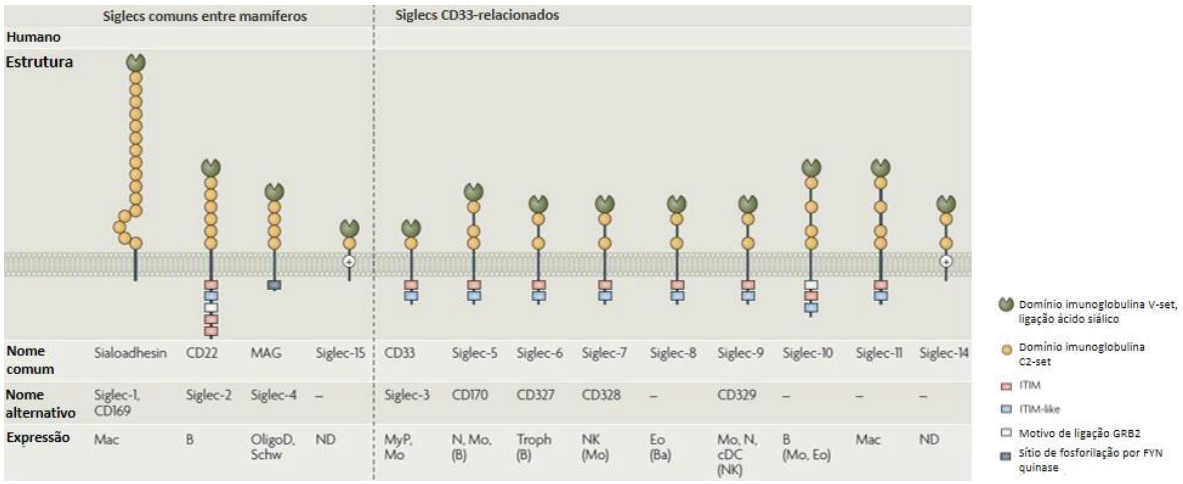
CD22 é o Siglec melhor caracterizado com 6 motivos ITIM em sua cauda citoplasmática, inibindo respostas de células B pela sua proximidade ao receptor de células B (BCR, do inglês *B-cell receptor*). Siglecs relacionados ao CD33 (CD33r) possuem dois motivos ITIM, cuja resposta inibidora mediada por CD33r ocorre através das células T e proliferação de células T primárias. Quando CD33r não possui motivos ITIM, a porção TM possui domínio para a molécula adaptadora DAP12 (revisado em CROCKER; PAULSON; VARKI, 2007; BARROW; TROWSDALE, 2008).

Os genes que codificam os Siglecs estão localizados na região estendida do LRC, 19q13, exceto os genes *SIGLEC-1* e *SIGLEC-15* que estão localizados no cromossomo 20 e 18, respectivamente.

Os Siglecs podem ser divididos em dois subgrupos baseados em similaridade de sequência gênica e conservação evolutiva. Um grupo é formado pelos genes que codificam CD22 (Siglec-2), glicoproteína associada a mielina (MAG, do inglês *myelin-associated glycoprotein* ou Siglec-4), sialoadesina (Siglec-1 ou Sn, do inglês *sialoadhesin*) e Siglec-15 (FIGURA 9). Os genes *CD22* e *MAG* estão localizados na região 19q13.12 na posição centromérica do LRC estendido, a 0,5 Mpb dos genes DAP10 e DAP12 (FIGURA 6) (revisado em McMILLAN; CROCKER, 2008; CROCKER; PAULSON; VARKI, 2007; BARROW; TROWSDALE, 2008).

O outro grupo é formado por Siglecs CD33r, que possuem similaridade de sequência com o *CD33* (50% a 90% de identidade de sequência) e de evolução rápida, envolvendo diversos mecanismos evolutivos, como duplicação gênica, embaralhamento de éxons e conversão gênica. Os genes Siglecs CD33r estão localizados na região 19q13.33, a 16 Mpb na porção telomérica dessa região (revisado em McMILLAN; CROCKER, 2008; CROCKER; PAULSON; VARKI, 2007; BARROW; TROWSDALE, 2008). Em humanos, conhecem-se nove genes Siglecs-CD33r: *CD33*, *Siglecs-5* a *-11* e *Siglec-14* (FIGURA 9) (revisado em McMILLAN; CROCKER, 2008; BARROW; TROWSDALE, 2008). SNPs do *CD33* foram associados à susceptibilidade do PFE pelo nosso grupo de pesquisa (CAMARGO, 2015). Estudos de expressão estão sendo finalizados para compreender se SNPs alteram os níveis de RNAm do gene.

FIGURA 9: FAMÍLIA DE RECEPTORES SIGLEC EM HUMANOS.



Fonte: Adaptado de CROCKER; PAULSON; VARKI, 2007.
Detalhe: Lado esquerdo o grupo com receptores sialoadesina, CD22, MAG e Siglec-15. Do lado direito os receptores Siglecs relacionados ao CD33 (CD33r).

2.2.5 Receptor para IgA FcαRI (FCAR, FcαRI ou CD89)

O gene *FCAR* que codifica o receptor para IgA FcαRI (FIGURA 6) que pode induzir respostas inibidoras e ativadoras dependendo da maneira como IgA se liga ao receptor. Na forma monomérica e sérica, IgA induz fosforilação parcial de ITAM do adaptador FcRγ, provocando resposta inibidora. Por outro lado, se o ligante desse receptor for a molécula multimérica IgA:complexo antigênico, provoca forte fosforilação de FcRγ e, em consequência, ativação celular. O receptor FcαRI possui aproximadamente 35% de similaridade de sequência com outros receptores codificados por genes presentes no LRC (revisado em TROWSDALE et al., 2015; BARROW; TROWSDALE, 2008).

2.2.6 DAP10 e DAP12

DAP10 e DAP12 são dois adaptadores fisicamente associados com imunoreceptores, envolvidos em respostas inibidoras e ativadoras. Os receptores são codificados por genes localizados na região 19q13.12 (FIGURA 6).

DAP10 (ou HCST) é uma proteína transmembranar e a região citoplasmática é responsável pelas sinalizações intracelulares. DAP10 possui 93 aminoácidos e é expresso na superfície celular associado ao receptor NKG2D, formando um complexo com 6 moléculas com um homodímero de NKG2D e 4 proteínas DAP10. NKG2D é a

subunidade de interação com o ligante e DAP10 é a subunidade de sinalização (revisado em BARROW; TROWSDALE, 2008; WU et al., 1999).

DAP12 (KARAP ou TYROPB) é uma molécula adaptadora com estrutura similar ao DAP10. DAP12 tem 133 aminoácidos com um domínio extracelular curto que forma um dímero com outra molécula DAP12. A região citoplasmática de DAP12 possui um motivo ITAM que promove a ativação celular, por liberação de citocinas e quimiocinas, fagocitose e liberação de grânulos citotóxicos, dependendo do subtipo de linfócito ativado (revisado em BARROW; TROWSDALE, 2008; TOMASELLO; VIVIER, 2005).

2.3 JUSTIFICATIVA

O complexo de receptores leucocitários (LRC) e a região do LRC estendido apresentam uma grande concentração de genes cujos produtos possuem relevantes papeis na resposta imunológica. Diversos estudos destacam que polimorfismos ou diferentes níveis de expressão desses genes, como *KIR*, *LAIR* e *CD33*, estão associados a alteração na resposta imune e com doenças autoimunes, destacando-se a susceptibilidade ao PFE. Além do mais, a análise de expressão global de células T CD4+ revelou que na região 19q13 concentram-se genes diferencialmente expressos entre PFE e controles. No entanto, a variabilidade e a função dos genes presentes nesta região genômica é pouco explorada.

Considerando a relevância de polimorfismos do LRC na susceptibilidade ao PF e doenças autoimunes e produtos gênicos atuantes na resposta imunológica, consideramos esta região cromossômica importante na busca de variabilidade genética na susceptibilidade ao PF.

3 OBJETIVO

Identificar polimorfismos genéticos associados ao pênfigo foliáceo localizados na região 19q13 em uma análise de associação caso-controle.

3.1 OBJETIVOS ESPECÍFICOS

i. Realizar o estudo de associação caso-controle dos polimorfismos localizados na região LRC estendido (19q13);
ii. Realizar o estudo de associação caso-controle dos polimorfismos localizados no LRC (19q13.42);
iii. Selecionar os possíveis SNPs causais, por mapeamento fino, e realizar nova análise caso-controle em amostras de PF endêmico e PF esporádico;
iv. Avaliar os possíveis impactos funcionais de SNPs associados a susceptibilidade do PF.

Para atender os objetivos específicos, a tese está dividida em 3 capítulos: (I) estudo de associação caso-controle dos polimorfismos da região LRC estendido (19q13), em forma de artigo científico; (II) polimorfismos do LRC (19q13.42), artigo científico publicado, e (III) descrição dos achados em desdobramentos do estudo de associação da região genômica 19q13.42.

4 CAPÍTULO I

Non-coding RNA gene variants in the extended leucocyte receptor complex (LRC) region are associated with susceptibility to endemic pemphigus foliaceus, an autoimmune skin disease.

Ticiana Della Justina Farias[1], Danillo Gardenal Augusto[1,2], Rodrigo Coutinho de Almeida[1,3], Danielle Malheiros[1], Hauke Busch[4], Enno Schmidt[4], Maria Luiza Petzl-Erler[1]

1 Laboratory of Human Molecular Genetics, Department of Genetics, Federal University of Paraná, Curitiba, Brazil.
2 Department of Neurology, University of California San Francisco, San Francisco, CA 94158, USA.
3 Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands.
4 Lübeck Institute of Experimental Dermatology (LIED), University of Lübeck, Lübeck, Germany.

4.1 ABSTRACT

Pemphigus foliaceus (PF) is a blistering autoimmune skin disease endemic in Brazil, also known as *fogo selvagem* (FS), characterized by the detachment of epidermal cells and the presence of autoantibodies against desmoglein 1. The pathophysiology of PF is multifactorial, involving environmental and genetic factors. Previously we have demonstrated associations of susceptibility to FS with both polymorphisms and differential expression levels of immune-related genes located in the extended leucocyte receptor complex (LRC) in chromosome 19q13. Until now, the polymorphism of genes in the 19q13 region (the extended LRC) has not been properly explored in complex diseases. Here we aimed to screen the extended LRC for genetic variants that may contribute to FS susceptibility, to identify coding and non-coding genes and regulatory elements that contribute to the pathogenesis of the disease. For this purpose, we performed association analysis by logistic regression with 2,510 tag single nucleotide polymorphisms (SNPs) of a 20 Mb segment within the 19q13 region, extracted from data of 229 FS patients and 194 controls genotyped by DNA microarray. Subsequently, the associated SNPs were genotyped in 76 sporadic PF patients and 150 controls from Germany. Nine SNPs were significantly associated ($p < 10^{-3}$) with variable susceptibility to FS: the *rs8103298* and *rs3810265* SNPs in the *RPS9P4* pseudogene (OR = 2.50 and OR = 2.33, respectively), four SNPs within the long non-coding RNA (lncRNA) *lnc-IGFL2-1* gene (*rs9304655*, *rs6509255*, *rs10412596*,

rs10439109; OR ≈ 2.0 for all), and SNPs in the protein-coding genes *SULT2B1* (*rs11083944*, OR = 0.49), *SCAF1* (*rs2304208*, OR = 3.04), and *RDH13* (*rs1654431*, OR = 2.34). None of these SNPs were associated with sporadic PF. These results led to the discovery of variants in four protein-coding genes, two ncRNA genes and regulatory regions which might influence the FS pathogenesis. We suggest exploring the extended LRC region in other immune-related diseases as well.

Key-words: 19q13, pemphigus foliaceus, *RPS9P4*, *lnc-IGFL2-1*, lncRNA

4.2 INTRODUCTION

Pemphigus is a group of blistering autoimmune skin diseases characterized by keratinocyte detachment (acantholysis) that result in flaccid blisters and by presence of autoantibodies against desmosomal cadherins. In pemphigus foliaceus (PF), the blisters are located at superficial layers of the epidermis (granular layers) and the major autoantigen is desmoglein 1 (DSG1) (ARTEAGA et al., 2002; MAHONEY et al., 1999). PF is endemic in Central and Southeastern Brazil where it is popularly known as *fogo selvagem* (FS, which means “wild fire”). PF is also endemic in neighboring countries as Colombia, Peru, and Bolivia (BYSTRYN; RUDOLPH, 2015; ORTEGA-LOAYZA et al., 2013) and Tunisia (MORINI et al., 1993; SELLAMI et al., 2004). The sporadic PF form is rare and can be observed across the globe. In Germany, PF prevalence were estimated to be around 5 patients/million in 2014 (reviewed by van BEEK; ZILLIKENS; SCHMIDT, 2018; HÜBNER et al., 2016). Conversely, FS prevalence is up to 3% in some regions (HANS-FILHO et al., 1996). Both the endemic and sporadic forms exhibit similar clinical, immunological and histological characteristics (reviewed by HANS-FILHO et al., 2018; BAUM et al., 2014).

PF is a complex disease triggered by ill-defined environmental factors. Salivary antigens from hematophagous insets may trigger the FS, possibly by inducing cross-reactive autoantibody response against DSG1 in genetically susceptible individuals (LOMBARDI et al., 1992; reviewed by QIAN et al., 2016). The involvement of infectious triggers is an alternative hypothesis that also remains to be investigated (reviewed by RUOCCO et al., 2014).

Variants of numerous genes influence susceptibility to FS. The first genetic associations reported were variants within the *human leucocyte antigen* (*HLA*) genes,

with strong associations with alleles of *HLA-DRB1* and *HLA-DQB1* (BROCHADO et al., 2016; PAVONI et al., 2003; PETZL-ERLER; SANTAMARIA, 1989). Our group has demonstrated several other genetic associations of FS with genes involved in immune responses (BUMILLER-BINI et al., 2018; SALVIANO-SILVA; PETZL-ERLER; BOLDT, 2017; CIPOLLA et al., 2016; PIOVEZAN; PETZL-ERLER, 2013; DALLA-COSTA et al., 2010; MALHEIROS; PETZL-ERLER, 2009; PEREIRA et al., 2004).

The focus of this study is to analyze variants within the extended leucocyte receptor complex (LRC) for associations with FS. The LRC is located at the 19q13.42 genomic region and contains more than 45 genes, including some immunoglobulin superfamily (IgSF) receptor genes that encode the polymorphic killer-cell immunoglobulin-like receptors (KIR), the leucocyte immunoglobulin-like receptors (LILR) and the leucocyte-associated immunoglobulin-like receptors (LAIR) (WENDE et al., 1999). Further in the centromeric region, in the so-called extended LRC, there are additional genes that encode receptors expressed in different immune system cells, such as the sialic acid binding immunoglobulin-like lectin (SIGLEC) family and the adaptor molecules DAP10 and DAP12, interspersed with genes of apparently unrelated functions (reviewed by BARROW; TROWSDALE, 2008; WENDE; VOLZ; ZIEGLER, 2000).

Studies from our group pointed out to the importance of the cytogenetic region 19q13 to pemphigus susceptibility, with strong associations of *KIR* (AUGUSTO et al., 2012, 2015), *LAIR1* and *LAIR2* (CAMARGO; AUGUSTO; PETZL-ERLER, 2016), and of SIGLEC family-member *CD33* (CAMARGO et al., unpublished results) polymorphisms. Screening of the full LRC revealed new variants associated with FS which could influence expression of LRC genes (FARIAS et al., 2018). Apart from these genetic associations, our group also performed a genome-wide expression profiling with over 50,000 probes. This study revealed that genes in 19q13 (*CD33, NALP12, NKG7, LILRB2*) were differentially expressed in CD4+ lymphocytes when comparing patients and controls, as well as different clinical forms of FS (MALHEIROS et al., 2014). These previous results indicate that variation within the LRC is relevant for FS susceptibility and motivated a systematic search for additional susceptibility loci at the extended LRC.

We hypothesized that variants in the extended LRC region that influence *fogo selvagem* susceptibility remain unknown. Variation in this region could impact gene function and expression. In this way, we analyzed 2,510 SNPs in the extended LRC in

FS and controls, to search for genetic associations with differential susceptibility to pemphigus. The associated variants were subsequently analyzed in sporadic PF patients and controls from Germany.

4.3 METHODS

4.3.1 Patient and control population samples

The FS sample was comprised by 229 unrelated patients with diagnosis based on clinical, histopathological and/or immunohistochemical criteria, recruited at hospitals from endemic areas, as described in Cipolla et al. (2016). The control sample encompassed 194 individuals without known history of autoimmune diseases who lived in endemic areas. Both patients and control samples were of predominantly European ancestry. A second case-control study was performed with 76 sporadic PF patients and 150 controls from Biobank PopGen, Germany matched for age and gender. All individuals signed informed consent. This study was performed according to Brazilian and German federal laws and approved by the Human Research Ethics Committees of the Federal University of Paraná and of the University of Lübeck (Biobank PopGen, number 2017-018).

4.3.2 DNA pemphigus samples genotyping

DNA obtained from peripheral blood was isolated by phenol-chloroform-isoamyl alcohol protocol (SAMBROOK; RUSSELL 2001). Genotyping was performed by microarray DNA Human Infinium® CoreExome-24 Beadchip (Illumina®), according to manufacturer instructions. The German samples were genotyped by iPLEX® MassARRAY® system (Agena Bioscience, USA), at *Insitut für Klinische Molekularbiologie* (IKMB), *Christian-Albrechts-Universität*, Kiel, Germany.

4.3.3 Selection of 19q13 region markers and statistical analysis

From the DNA microarray data, we extracted 8,669 SNPs within the 19q13 region (chr19: 35,700,000 – 55,600,000, GRCh37/hg19), considering MAG and GP6 genes as the extended LRC flanking genes. After the standard quality controls

(ANDERSON et al., 2010), we filtered SNPs by minor allele frequency higher than 2% (MAF > 2%) and by SNPs in Hardy-Weinberg equilibrium in the control sample ($p > 0.01$), and then performed the case-control study with the 2,510 remaining SNPs.

Association analysis was performed by logistic regression with the two first components from principal component analysis (PCA) as covariant, to correct for possible population stratification. Since the effect of genetic variants on disease susceptibility may not differ between the heterozygote and one of the homozygote genotypes, we applied the dominant and recessive models of the less frequent allele in the population (the minor allele) in addition to the additive model.

Seven SNPs (*rs3810265, rs9304655, rs10412596, rs10439109, rs11083944, rs2304208, rs1654431*) were genotyped in the German cohort. For SNP *rs8103298* no genotype results were obtained and the SNP *rs6509255* was excluded from the second genotyping due to strong LD ($r^2 = 0.97$) with *rs9304655*.

The Hardy-Weinberg equilibrium test and estimates of linkage disequilibrium (LD) were performed with the PLINK v1.9 software (CHANG et al., 2015). Odds ratio (OR) and the 95% confidence interval (CI) were calculated with the PLINK v1.9 software and VassarStats (VassarStats: Website for Statistical Computation). The *p* values of 0.001 and 0.05 were considered as the significance limit for the FS and sporadic PF associations, respectively.

4.3.4 In silico analysis

SNP and gene annotations were obtained from UCSC (HINRICHS et al., 2016) and Ensembl (AKEN et al., 2016; MCLAREN et al., 2016; ZERBINO et al., 2107) genome browsers. To visualize the SNP distribution along chromosome 19, the additive and dominant association results were evaluated by the LocusZoom (PRUIM et al., 2010) online tool. The possible structural and regulatory impact of the genetic variants and their eQTL (expression Quantitative Trait Loci) effect were observed with HaploReg v. 4.1 (WARD; KELLIS, 2012) online tool, Genotype-Tissue Expression (GTEx) Portal (CARITHERS et al., 2015) and the Ensembl database. LncRNA gene annotation was obtained at LincSNP 2.0 (NING et al., 2017), LNCipedia.org v. 5.0 (VOLDERS et al., 2015), and LncRBase (CHAKRABORTY et al., 2014) online tools.

4.4 RESULTS

The screening of the extended LRC revealed nine SNPs associated with FS (FIGURE 10, TABLE 1). We found two variants *rs8103298* and *rs3810265* within ribosomal protein S9 pseudogene 4 (*RPS9P4*) associated with increased susceptibility to FS in the additive and dominant models. These SNPs were in moderate linkage disequilibrium (LD) with each other (r^2 = 0.69). In addition, four SNPs (*rs9304655*, *rs6509255*, *rs10412596*, and *rs10439109*) associated with FS are within *lnc-IGFL2-1* (AC007785.1, ENSG00000267922), a lncRNA to the *IGF-like family member 2 (IGFL2)* gene. Detailed SNPs annotation can be found in Table S1 (APÊNDICE A). The *rs9304655* is in strong LD with *rs6509255* (r^2 = 0.97) and moderate LD with *rs10439109* and *rs10412596* (r^2 = 0.60 and r^2 = 0.48, respectively) (FIGURE 10).

The intronic *rs11083944* SNP of *sulfotransferase family 2B member 1* (*SULT2B1*) and the *rs2304208* SNP within exon 9 of *SR-related CTD associated factor 1* (*SCAF1*) genes were associated with FS in the dominant model. Conversely, the *rs1654431* SNP within intron 5 of *retinol dehydrogenase 1* (*RDH13*) gene was associated with FS in the recessive model (TABLE 1). The SNP *rs2304208* (*SCAF1*) also associated with FS presented moderate LD (r^2 = 0.48) with *rs3810265* (*RPS9P4*) in our sample (FIGURE 10). None of the seven SNPs (*rs3810265, rs9304655, rs10412596, rs10439109, rs11083944, rs2304208, rs1654431*) were associated with sporadic PF (TABLE 2). The allelic and genotypic frequencies are in Table S2 (APÊNDICE A).

In silico analysis suggest that all variants associated with FS are within regulatory regions, as transcription factor (TF) binding sites (Tables S3 and S4 – APÊNDICE A), regulatory DNA motifs or sites of chromatin state modification (described later in the Discussion section). According to the GTEx Portal, the SNPs within *lnc-IGFL2-1* have an eQTL effect on the *lnc-IGFL1-1*, *lnc-IGFL2-1*, *lnc-IGLF3-1,* and *IGLF2-AS1* genes and on the *IGFLP1* pseudogene in sun-exposed and not sun-exposed skin (Figures S1-S4, APÊNDICE A). In addition, the *rs1654431* SNP within the *RDH13* gene is an eQTL with a negative effect on the expression of the *RDH13*, *NLRP2* and *lnc-NLRP2-2* genes in whole blood, according to the GTEx Portal (CARITHERS et al., 2015) (Figure S5, APÊNDICE A) (TABLE S5 and S6 – APÊNDICE A).

TABLE 1: RESULTS OF THE ASSOCIATION ANALYSIS BY THE ADDITIVE, DOMINANT AND RECESSIVE MODELS SHOWING THE SNPS ASSOCIATED WITH FOGO SELVAGEM.

SNP	Reference allele	model	OR	CI 95%	*p*	Gene
rs8103298	C	additive	2.50	1.62 – 3.85	0.00004	
		dominant	2.72	1.71 – 4.34	0.00003	*RPS9P4*
		recessive	2.84	0.55 – 14.54	0.21050	
rs3810265	A	additive	2.33	1.42 – 3.82	0.0008	
		dominant	2.47	1.47 – 4.13	0.0006	*RPS9P4*
		recessive	2.25	0.23 – 22.10	0.4865	
rs9304655	A	additive	1.65	1.25 – 2.19	0.0005	
		dominant	2.02	1.34 – 3.04	0.0007	*lnc-IGFL2-1* / upstream of *IGFL2*
		recessive	1.80	1.06 – 3.05	0.0302	
rs6509255	C	additive	1.63	1.22 – 2.17	0.0009	
		dominant	2.00	1.33 – 3.00	0.0009	*lnc-IGFL2-1* / upstream of *IGFL2*
		recessive	1.70	0.99 – 2.94	0.0546	
rs10412596	T	additive	1.73	1.26 – 2.37	0.0008	
		dominant	1.95	1.32 – 2.89	0.0009	*lnc-IGFL2-1* / downstream of IGFL2-AS1
		recessive	2.03	0.94 - 4.42	0.0731	
rs10439109	A	additive	1.49	1.13 – 1.98	0.0053	
		dominant	2.17	1.38 – 3.41	0.0009	*lnc-IGFL2-1* / downstream of *lnc-IGFL3-1*
		recessive	1.30	0.82 – 2.07	0.2681	
rs11083944	G	additive	0.62	0.47 - 0.84	0.0014	
		dominant	0.49	0.32 - 0.75	0.0007	*SULT2B1*
		recessive	0.63	0.38 – 1.06	0.0847	
rs2304208	C	additive	2.72	1.49 – 4.98	0.0012	
		dominant	3.04	1.60 – 5.79	0.0007	*SCAF1*
		recessive	2.41	0.25 – 23.50	0.4484	
rs1654431	A	additive	1.39	1.05 – 1.83	0.0221	
		dominant	1.11	0.72 – 1.71	0.6369	*RDH13*
		recessive	2.34	1.42 – 3.84	0.0008	

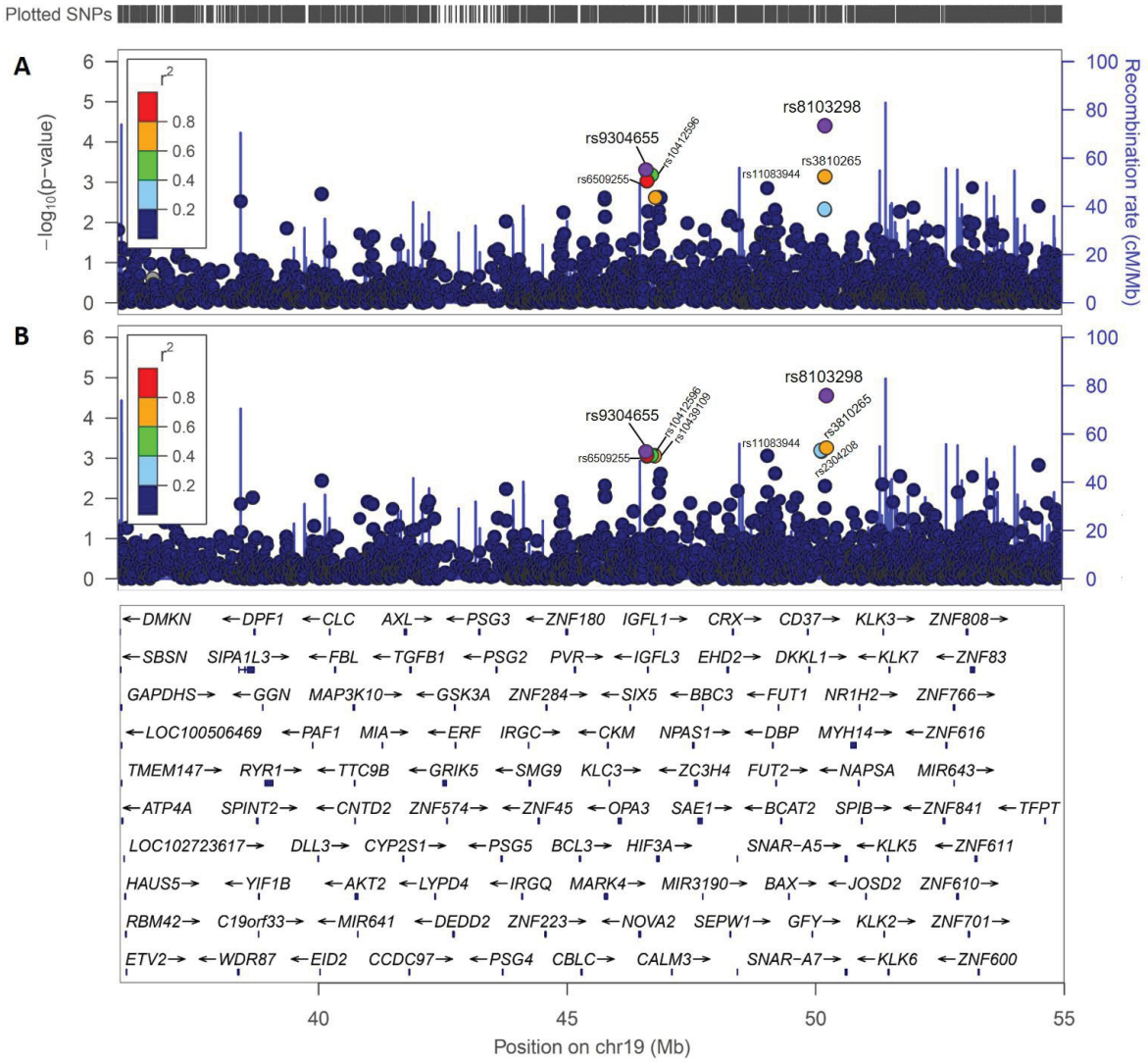
SNP: single nucleotide polymorphism, OR: odds ratio; CI: confidential interval, *p*: *p*-value. In bold the *p*-values $< 10^{-3}$ and in bold and italic *p*-values $< 10^{-4}$.

TABLE 2: RESULTS OF THE ASSOCIATION ANALYSIS BY THE ADDITIVE, DOMINANT AND RECESSIVE MODELS SHOWING THE SNPS ASSOCIATED WITH SPORADIC PEMPHIGUS FOLIACEUS.

Gene / SNP	Reference allele	model	OR	CI 95%	p
lnc-IGFL2-1					
rs9304655	A	additive	0.92	0.56 – 1.53	0.798
		dominant	1.01	0.52 – 1.97	1.000
		recessive	0.60	0.16 – 2.21	0.563
rs10412596	T	additive	0.81	0.46 – 1.43	0.573
		dominant	0.91	0.49 – 1.72	0.873
		recessive	NC	NC	NC
rs10439109	A	additive	1.31	0.86 – 1.99	0.236
		dominant	1.50	0.79 – 2.86	0.215
		recessive	1.37	0.62 – 3.02	0.436
SULT2B1					
rs11083944	G	additive	1.02	0.63 – 1.65	1.000
		dominant	1.95	0.88 – 4.34	0.098
		recessive	0.50	0.21 – 1.18	0.108
RDH13					
rs1654431	A	additive	1.02	0.64 – 1.62	1.000
		dominant	1.34	0.63 – 2.87	0.449
		recessive	0.78	0.36 – 1.67	0.519

SNP: single nucleotide polymorphism, OR: odds ratio; CI: confidential interval, *p*: *p*-value. NC: not calculated. The *rs3810265* (*RPS9P4*) and *rs2304208* (*SCAF1*) SNPs are not shown because they were monomorphic in German samples.

FIGURE 10: RESULT OF THE ASSOCIATION ANALYSIS ACCORDING TO THE ADDITIVE (A) AND DOMINANT (B) MODELS BETWEEN 2,510 SNPS AND *FOGO SELVAGEM*, SHOWING THE DISTRIBUTION OF SNPS AND THE LOCALIZATION OF THE GENES ALONG THE 19q13 GENOMIC REGION (GRCH37/HG19).



Legend: The purple dots are the two SNPs with lowest p-value (*rs8103298* for *lnc-IGFL2-1* gene and *rs9304655* for *RPS9P4* pseudogene). The red, orange and light blue dots are SNP presenting linkage disequilibrium evaluated by correlation analysis (r^2) with the nearest purple SNP. Mb: mega base (10^6 bp). SNP single nucleotide polymorphisms. Adapted from LocusZoom online tool (PRUIM et al., 2010).

4.5 DISCUSSION

In former studies, we demonstrated strong associations between FS susceptibility and polymorphisms and differential expression levels from several genes within the extended leucocyte receptor complex (LRC) (MALHEIROS et al., 2014; AUGUSTO et al., 2012, 2015; CAMARGO; AUGUSTO; PETZL-ERLER, 2016; FARIAS et al., 2018, CAMARGO et al., unpublished results). Here, we report associations between FS and polymorphisms along the extended LRC, at the 19q13 cytogenetic location. The results show that the extended LRC bears functional elements whose variants have an impact on FS pathogenesis. That impact may result from the potential effect of the genetic variants on transcription in several tissues and cell types.

The strongest association of FS is with *RPS9P4* pseudogene variants (*rs8103298 C* and *rs3810265 A*) which are within transcription factor (TF) binding sites and histone methylation marks (H3K4me1/3), potentially changing the DNA accessibility and the transcription levels in skin cells, according to the HaploReg online tool (WARD; KELLIS, 2012) (Table S3- S4, APÊNDICE A).

Both variants are also annotated within the transcript of uncertain coding potential (TUCPs) TCONS_l2_00013175 (HINRICHS et al., 2016) (Table S1, APÊNDICE A). This transcript participates of a gene co-expression module enriched with lincRNAs (long intergenic non-coding RNAs), TFs, and microRNAs that may regulate the expression of protein-coding genes whose altered levels may contribute to prostate cancer pathogenesis (CUI et al., 2015). In fact, the interaction of non-coding RNAs transcribed from pseudogenes with TF and other ncRNAs may form complexes that regulate transcription (GRANDER and JOHNSSON 2015; GROEN, CAPRARO, and MORRIS 2014). Moreover, transcripts of pseudogenes could regulate inflammatory conditions that also characterize FS pathogenesis, through the nuclear factor kappa B (NF-κB) pathway (RAPICAVOLI et al., 2013). Thus, the variants of the *RPS9P4* pseudogene transcript could influence the function of this ncRNA on gene regulation, contributing to FS pathogenesis.

We found four variants (*rs9304655 A*, *rs6509255 C*, *rs10412596 T*, *rs10439109 A*) within *lnc-IGFL2-1* associated with FS which decrease the expression levels of *lnc-IGFL1* and of *IGFLs* genes in the skin (Table S5, Figure S1-S5, CARITHERS et al., 2015). The *IGFL*-related *lnc-IGFL2-1* may regulate the expression of neighboring *IGFL*

genes. The insulin-like growth factor (IGF)-like (IGFL) family comprises four genes (*IGFL1-4*) and two pseudogenes (*IGFLP1-2*) clustered within a 35 kb segment at the cytogenetic location 19q13. Soluble IGFL proteins show structural similarity with IGF proteins and play critical roles in cellular energy metabolism and in cell proliferation. Thus, the IGFL proteins might act as secreted growth regulators (EMTAGE et al., 2006).

The *IGFL2* antisense RNA 1 *(IGFL2-AS1)* negatively regulates the expression of *IGFL1*. The decreased expression of *IGFL2-AS1* and the consequent upregulation of *IGFL1* reduces cell-cell adhesion and may reduce migration of breast cancer cells (TRACY et al., 2018). Increased levels of *IGFL1* were also observed for psoriasis and are a hallmark for that autoimmune skin disease (GUO et al., 2014). Additionally, mouse T cells express the IGFL1 receptor (IGFL1R), indicating cross-talk between keratinocytes and cells of the immune system in inflammatory skin conditions (LOBITO et al., 2011). Therefore, high *IGFL1* expression levels, due to low *IGFL2-AS1* levels associated with FS, might facilitate the development of inflammatory skin conditions. It is conceivable that the involvement of IGFL1 in keratinocyte proliferation, increased NF-κB signaling, and stimulation of pro-inflammatory cytokine expression (GUO et al., 2014) may facilitate the development of FS as well.

Nevertheless, the *lnc-IGFL2-1* SNPs are in strong LD ($r^2 > 0.8$) with 40 other SNPs along a 55 kilobases (kb)-length genome segment that includes SNPs within the *IGFL2* gene (European populations from the 1000 genomes project, MACHIELA; CHANOCK, 2015; WARD; KELLIS, 2012). The SNPs associated with FS are tag SNPs for variants within *IGFL2*. This gene could influence FS susceptibility since its expression level is down-regulated in injured skin of psoriasis patients and is negatively modulated after mitochondrial depletion in human skin fibroblasts (LOBITO et al., 2011; IVANOV et al., 2011). On the other hand, *IGFL2* is overexpressed in Sézary syndrome, a rare, aggressive CD4+ cutaneous T-cell lymphoma (BOOKEN et al., 2007). Other than *IGFL1/2* differential levels in skin inflammatory conditions, *IGFL3* levels did not differ between injured and non-injured skin of psoriasis patients (LOBITO et al., 2011). However, the FS risk *rs10412596 C/C* genotype decreases the *IGFL3* mRNA levels in sun-exposed skin tissue (CARITHERS et al., 2015), suggesting a possible effect of this gene in FS skin pathogenesis. In addition to the possible effect of *lnc-IGFL* and *IGFL* variants on gene regulation, the SNPs associated with FS might modify TFs binding motifs (MEF2, GATA2, and others), according to the HaploReg

online tool (WARD; KELLIS, 2012). Therefore, it may be speculated that altered *IGFL* or *lnc-IGFL* expression levels might modulate skin pathogenesis.

We found other three SNPs associated with differential susceptibility to FS within the *SULT2B1*, *SCAF1* and *RDH13* protein-coding genes.

Sulfotransferase enzymes catalyze the sulfate conjugation of many hormones, neurotransmitters, drugs, and xenobiotic compounds (JI et al., 2007; YANAI et al., 2004). Two isoforms of human hydroxysteroid sulfotransferase (SULT) are encoded by the *SULT2B1* gene: SULT2B1a and SULT2B1b (HER et al., 1998). Only SULT2B1b expression is observed in normal human skin and in epidermal keratinocytes (HIGASHI et al., 2004; KOHJITANI et al., 2006). The SULT2B1 protein plays a key role in homeostasis and barrier function of the human skin. The lack of SULT2B1 due to loss-of-function mutations causes an autosomal-recessive congenital ichthyosis phenotype via increased proliferation of keratinocytes, thickening of epithelial layers, and altered epidermal cholesterol metabolism (HEINZ et al., 2017).

Skin tissue homeostasis seems to be related to the *rs11083944 G/G* and *G/A* genotypes of *SULT2B1* that are associated with decreased susceptibility to FS. That association was not observed for sporadic PF. However, this result is inconclusive because in the control sample of sporadic PF the *rs11083944* genotype frequencies were not in HWE ($p = 0.004$).

The *rs11083944* SNP is an eQTL for *SULT2B1* expression and other genes, according to GTEx Portal (Table S4, CARITHERS et al., 2015). In addition, the *rs11083944 G* allele affects the binding site of transcriptional enhancer factor dimer TEF-3 (TEAD4) and Eomesodermin (EOMES) (Table S3 – APÊNDICE A). TEAD4 promotes cell proliferation, migration and induce epithelial-mesenchymal transition (ZHANG et al., 2009). EOMES is involved in the differentiation of CD8+ T cells and NK and innate lymphoid cells (ILCs) during immune response which regulates lytic effector and LRC genes (ATREYA et al., 2007; ZHANG et al., 2018). Consequently, the *rs11083944* may disturb skin homeostasis and immune cell differentiation, therefore facilitating FS pathogenesis.

The SR-Related CTD Associated Factor 1, SCAF1, protein interacts with the C-terminal domain (CTD) of the large subunit of RNA polymerase II (POLR2A) and participates in the pre-mRNA splicing (KATSAROU et al., 2005; SCORILAS et al., 2001). SCAF1 has a wide gene regulatory function and is highly expressed in normal skin tissues, according to the Expression Atlas webtool (PETRYSZAK et al., 2016).

Besides resulting in the benign methionine to threonine amino acid replacement at position p.1146, the *SCAF1* SNP *rs2304208* may influence gene regulation. The *rs2304208* SNP is within binding motifs for the TFs HNF4 and ZIC and marks chromatin states in skin cells, according to the HaploReg online tool (WARD; KELLIS, 2012). Therefore, the polymorphism may modulate gene regulation and contribute to FS development.

Finally, the *rs1654431 A/A* genotype associated with FS decreases *RDH13* gene expression in sun-exposed skin and increases its levels in whole blood, according to the GTEx Portal (Table S4; CARITHERS et al., 2015). The *RDH13* gene encodes mitochondrial dehydrogenase/reductase retinol dehydrogenase 13 enzyme, which catalyzes retinoids, protecting the organelle against oxidative stress (BELYAEVA et al., 2008). An alternative annotation has shown that the *rs1654431* SNP is within a *lnc-NLRP2* intron (Table S1, APÊNDICE A). This variant also has a positive eQTL effect on both *NLRP2* and *lnc-NLRP2* genes expression in whole blood and esophagus mucosa (Table S4, CARITHERS et al., 2015). The *NLRP2* gene at 19q13.4, encodes the protein NLRP2 (NACHT, LRR and PYD domains-containing protein 2) which acts as a modulator of the NF-κB and pro-caspase-1 pathway, when stimulated by inflammatory mediators (cytokines, among others), what is common in autoimmune responses (BRUEY et al., 2004). The regulatory impact of this variant is reinforced since it alters more than 30 TFs binding motifs (LBP1 and PAX5), according to the HaploReg online tool (WARD; KELLIS, 2012). This variant might regulate gene expression, especially of the *NLRP2* gene that potentially plays a pro-inflammatory role on FS pathogenesis.

No genetic associations with SNPs *rs3810265, rs9304655, rs10412596, rs10439109, rs11083944, rs2304208, rs1654431* were found in German PF samples. Actually, *rs3810265* (*RPS9P4*) and *rs2304208* (*SCAF1*) were monomorphic in the German samples. However, lack of association these extend LRC variants cannot be explained by differing allele frequencies. Although sporadic and endemic PF share clinical, immunologic and histologic characteristics (HANS-FILHO, et al., 2018; BYSTRYN; RUDOLPH, 2005), the epidemiology differs and environmental triggers might be different. Therefore, the restriction of genetic associations to FS could result from the clearly different environmental conditions to which the Brazilian FS patients living in endemic areas and the PF patients living in Germany are exposed.

Unfortunately, the environmental triggers are poorly known. The results of the present study provide additional motivation for investigation in that field.

4.6 CONCLUSION

This study revealed that still unexplored genes and polymorphisms in the extended LRC region at the cytogenetic location 19q13 influence the susceptibility to FS, but not to sporadic PF. We found nine previously undiscovered markers within region 19q13 which mainly have an effect on gene regulation. Some of the genes whose expression may be modified by the associated variants may be important to maintain the upper layers of keratinocytes in the epidermis. Other may be involved in immune responses. These results corroborate the importance of the 19q13 region in susceptibility to FS and probably other inflammatory skin diseases, indicating the need of further functional studies.

4.7 REFERENCES

1. AKEN, B. L. et al. The Ensembl gene annotation system. Database, v. 2016, p. baw093, 2016.
2. ANDERSON, C. A. et al. Data quality control in genetic case-control association studies. Nat Protoc, v. 5, p. 1564-73, 2010.
3. ARTEAGA, L. A. et al. A Subset of Pemphigus Foliaceus Patients Exhibits Pathogenic Autoantibodies Against Both Desmoglein-1 and Desmoglein-3. Journal of Investigative Dermatology, v. 118, n. 5, p. 806–811, 2002.
4. ATREYA, I. et al. The T-box transcription factor eomesodermin controls CD8 T cell activity and lymph node metastasis in human colorectal cancer. Gut, v. 56, n. 11, p. 1572-8, 2007.
5. AUGUSTO, D. G. et al. Activating KIR and HLA Bw4 Ligands Are Associated to Decreased Susceptibility to Pemphigus Foliaceus, an Autoimmune Blistering Skin Disease. PLoS ONE, v. 7, n. 7, p. e39991, 2012.
6. AUGUSTO, D. G. et al. Pemphigus is associated with KIR3DL2 expression levels and provides evidence that KIR3DL2 may bind HLA-A3 and A11 in vivo. European Journal of Immunology, v. 45, n. 7, p. 2052–2060, 2015.
7. BARROW, A. D.; TROWSDALE, J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. Immunological reviews, v. 224, p. 98–123, 2008.
8. BELYAEVA, O. V. et al. Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. FEBS Journal, v. 275, n. 1, p. 138–147, 2008.
9. BOOKEN, N. et al. Sézary syndrome is a unique cutaneous T-cell lymphoma as identified by an expanded gene signature including diagnostic marker molecules CDO1 and DNM3. Leukemia, v. 22, n. 2, p. 2405044, 2007.

10. BROCHADO, M. et al. Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region. Journal of Autoimmunity, v. 72, p. 19–24, 2016.
11. BRUEY, J.M. et al. PAN1/NALP2/PYPAF2, an inducible inflammatory mediator that regulates NF-kappaB and caspase-1 activation in macrophages. The Journal of Biological Chemistry, v. 279, n. 50, p. 51897-907, 2004.
12. BUMILLER-BINI, V. et al. Sparking Fire Under the Skin? Answers From the Association of Complement Genes With Pemphigus Foliaceus. Frontiers in Immunology, v. 9, p. 695, 2018.
13. BYSTRYN, J.-C. C.; RUDOLPH, J. L. Pemphigus. Lancet (London, England), v. 366, n. 9479, p. 61–73, 2005.
14. CAMARGO, C.; AUGUSTO, D. G.; PETZL-ERLER, M. Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus. Human Genetics, v. 135, n. 2, p. 233–244, 2016.
15. CARITHERS, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreservation and Biobanking, v. 13, n. 5, p 311–319, 2015.
16. CHAKRABORTY, S. et al. LncRBase: an enriched resource for lncRNA information. PLoS One, v. 9, n. 9, p. e108010, 2014.
17. CHANG, C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience, v. 4, p.7, 2015.
18. CIPOLLA, G. A. 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. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, v. 1859, n. 10, p. 1306–1313, 2016.
19. DALLA-COSTA, R. et al. Polymorphisms in the 2q33 and 3q21 chromosome regions including T-cell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. Human Immunology, v. 71, n. 8, p. 809–817, 2010.
20. EMTAGE, P. et al. IGFL: A secreted family with conserved cysteine residues and similarities to the IGF superfamily. Genomics, v. 88, n. 4, p. 513-20, 2006.
21. FARIAS, T.D.J. et al. Screening the full leucocyte receptor complex genomic region revealed associations with pemphigus that might be explained by gene regulation. Immunology, 2018. doi: 10.1111/imm.13003.
22. GRANDÉR, D.; JOHNSSON, P. Pseudogene-Expressed RNAs: Emerging Roles in Gene Regulation and Disease. Current Topics in Microbiology and Immunology, v. 394, p. 111-26, 2016.
23. GROEN, J. N.; CAPRARO, D.; MORRIS, K.V. The emerging role of pseudogene expressed non-coding RNAs in cellular functions. The International Journal of Biochemistry & Cell Biology, v. 54, p. 350-5, 2014.
24. GUO P. et al. Gene expression profile based classification models of psoriasis. Genomics, v. 103, n. 1, p. 48-55, 2014.
25. HANS-FILHO, G. et al. Fogo selvagem: endemic pemphigus foliaceus. Anais Brasileiros de Dermatologia, v. 93, n. 5, p. 638-50, 2018.
26. HANS-FILHO, G. et al., An active focus of high prevalence of fogo selvagem on an Amerindian reservation in Brazil. Journal of Investigative Dermatology, v. 107, p. 68–75, 1996.
27. HEINZ, L. et al. Mutations in SULT2B1 Cause Autosomal-Recessive Congenital Ichthyosis in Humans. The American Journal of Human Genetics, v. 100, p. 926–939, 2017.

28. HER, C. Human hydroxysteroid sulfotransferase SULT2B1: two enzymes encoded by a single chromosome 19 gene. Genomics, v. 53, n. 3, p. 284-95, 1998.
29. HIGASHI, Y. et al. Expression of cholesterol sulfotransferase (SULT2B1b) in human skin and primary cultures of human epidermal keratinocytes. The Journal of Investigative Dermatology, v. 122, n. 5, p. 1207-13, 2004.
30. HINRICHS, A. S. et al. UCSC Data Integrator and Variant Annotation Integrator. Bioinformatics, v. 32, n. 9, p. 1430-2, 2016.
31. HÜBNER, F. et al. Prevalence and Age Distribution of Pemphigus and Pemphigoid Diseases in Germany. Journal of Investigative Dermatology, v. 136, n. 12, p. 2495-8, 2016.
32. IVANOV, N. I. et al. Radiation response and regulation of apoptosis induced by a combination of TRAIL and CHX in cells lacking mitochondrial DNA: A role for NF-κB–STAT3-directed gene expression. Experimental Cell Research, v. 317, n. 11, p. 1548-66, 2011.
33. JI, Y. et al. Human hydroxysteroid sulfotransferase SULT2B1 pharmacogenomics: gene sequence variation and functional genomics. The Journal of Pharmacology and Experimental Therapeutics, v. 322, n. 2, p. 529-40, 2007.
34. KATSAROU, M. E. et al. Expression of the C-terminal domain of novel human SR-A1 protein: interaction with the CTD domain of RNA polymerase II. Biochemical and Biophysical Research Communications, v. 334, n. 1, p. 61-8, 2005.
35. KOHJITANI, A. et al. Cloning, characterization and tissue expression of rat SULT2B1a and SULT2B1b steroid/sterol sulfotransferase isoforms: divergence of the rat SULT2B1 gene structure from orthologous human and mouse genes. Gene, v. 367, p. 66-73, 2006.
36. LOBITO, A. A. et al. Murine Insulin Growth Factor-like (IGFL) and Human IGFL1 Proteins Are Induced in Inflammatory Skin Conditions and Bind to a Novel Tumor Necrosis Factor Receptor Family Member, IGFLR1. Journal of Biological Chemistry, v. 286, n. 21, p. 18969–18981, 2011.
37. LOMBARDI, C. et al. Environmental risk factors in endemic pemphigus foliaceus (Fogo selvagem). "The Cooperative Group on Fogo Selvagem Research". The Journal of investigative dermatology, v. 98, n. 6, p. 847–50, 1992.
38. MACHIELA, M. J.; CHANOCK, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics, v. 31, n. 21, p. 3555–3557, 2015.
39. MAHONEY, M. G. et al. Explanations for the clinical and microscopic localization of lesions in pemphigus foliaceus and vulgaris. Journal of Clinical Investigation, v. 103, n. 4, p. 461–468, 1999.
40. MALHEIROS, D. et al. Genome-wide gene expression profiling reveals unsuspected molecular alterations in pemphigus foliaceus. Immunology, v. 143, n. 3, p. 381–395, 2014.
41. MALHEIROS, D.; PETZL-ERLER, M. Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus. Genes and Immunity, v. 10, n. 6, p. 547, 2009.
42. MCLAREN, W. et al. The Ensembl Variant Effect Predictor. Genome Biology, v. 17, n. 1, p. 122, 2016.
43. MORINI, J. et al. Pemphigus foliaceus in young women. An endemic focus in the Sousse area of Tunisia. Archives of dermatology, v. 129, n. 1, p. 69–73, 1993.
44. NING, S. et al., LincSNP 2.0: an updated database for linking disease-associated SNPs to human long non-coding RNAs and their TFBSs. Nucleic Acids Research, v. 45, n. D1, p. D74-D78, 2017

45. ORTEGA-LOAYZA, A. et al. Endemic pemphigus foliaceus in the Peruvian Amazon. Clinical and Experimental Dermatology, v. 38, n. 6, p. 594–600, 2013.
46. PAVONI, D. et al. Dissecting the associations of endemic Pemphigus Foliaceus (Fogo Selvagem) with HLA-DRB1 alleles and genotypes. Genes and Immunity, v. 4, n. 2, p. 110, 2003.
47. PEREIRA, N. et al. Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants. Cytokine, v. 28, n. 6, p. 233–241, 2004.
48. PETRYSZAK, R. et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Research, v. 44, n. D1, p. D746–D752, 2016.
49. PETZL-ERLER, M.; SANTAMARIA, J. Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)? Tissue Antigens, v. 33, n. 3, p. 408–414, 1989.
50. PIOVEZAN, B.; PETZL-ERLER, M. Both qualitative and quantitative genetic variation of MHC class II molecules may influence susceptibility to autoimmune diseases: The case of endemic pemphigus foliaceus. Human Immunology, v. 74, n. 9, p. 1134–1140, 2013.
51. PRUIM, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics, v.26, p. 2336-7, 2010.
52. QIAN, Y. et al. Non-infectious environmental antigens as a trigger for the initiation of an autoimmune skin disease. Autoimmunity Reviews, v. 15, n. 9, p. 923–30, 2016.
53. RAPICAVOLI, N. A. et al., A mammalian pseudogene lncRNA at the interface of inflammation and anti-inflammatory therapeutics, Elife, v. 2, p. e00762, 2013.
54. RUOCCO, E. et al. Viruses and Pemphigus: An Intriguing Never-Ending Story. Dermatology, v. 229, n. 4, p. 310–5, 2014.
55. SALVIANO-SILVA, A.; PETZL-ERLER, M.; BOLDT, A. CD59 polymorphisms are associated with gene expression and different sexual susceptibility to pemphigus foliaceus. Autoimmunity, v. 50, n. 6, p. 377-85, 2017.
56. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual. CSHL Press, New York, 1448 p. 2001.
57. SCORILAS, A. et al. Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of pre-mRNA splicing factors: overexpression in aggressive ovarian cancer. British Journal of Cancer, v. 85, n. 2, p. 190-8, 2001.
58. SELLAMI, M. et al. Anti-desmoglein 1 antibodies in Tunisian healthy subjects: arguments for the role of environmental factors in the occurrence of Tunisian pemphigus foliaceus. Clinical & Experimental Immunology, v. 137, n. 1, p. 195–200, 2004.
59. TRACY K. M. et al. Selective expression of long non-coding RNAs in a breast cancer cell progression model. Journal of Cellular Physiology, v. 233, n. 2, p. 1291-9, 2018.
60. van Beek N, Zillikens D, Schmidt E. Diagnosis of autoimmune bullous diseases. J Dtsch Dermatol Ges. 2018. 16(9):1077-1091.
61. VOLDERS, P. J. et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Research, v. 43, n. 8, p. 4363-4, 2015.
62. WARD, L. D.; KELLIS, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Research, v. 40, n. D1, p. D930–D934, 2012.

63. WENDE, H. et al. Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. Mammalian genome: official journal of the International Mammalian Genome Society, v. 10, n. 2, p. 154–60, 1999.
64. WENDE, H.; VOLZ, A.; ZIEGLER, A. Extensive gene duplications and a large inversion characterize the human leukocyte receptor cluster. Immunogenetics, v. 51, n. 8–9, p. 703–713, 2000.
65. YANAI, H. et al. Expression of cholesterol sulfotransferase (SULT2B1b) in human platelets. Circulation, v. 109, n. 1, p. 92-6, 2004.
66. ZERBINO, D. R. et al. Ensembl 2018. Nucleic Acids Research, v. 46, n. D1, p. D754-D761, 2018.
67. ZHANG, H. et al. TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition. Journal of Biological Chemistry, v. 284, n. 20, p. 13355-62, 2009.
68. ZHANG, J. et al., T-bet and Eomes govern differentiation and function of mouse and human NK cells and ILC1. European Journal of Immunology, v. 48, n. 5, p. 738-50, 2018.

5 CAPÍTULO II

Screening the full leukocyte receptor complex (LRC) region revealed associations with pemphigus that might be explained by gene regulation[1]

Ticiana Della Justina Farias[1], Danillo G. Augusto[1,2], Rodrigo Coutinho de Almeida[3], Danielle Malheiros[1], Maria Luiza Petzl-Erler[1*]

[1] Laboratory of Human Molecular Genetics, Department of Genetics, Federal University of Paraná, Curitiba, Brazil; [2] Biological Sciences, Universidade Estadual de Santa Cruz, Ilhéus, Brazil; [3] Department of Biomedical Data Sciences, Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

5.1 ABSTRACT

Pemphigus foliaceus (PF) is a blistering autoimmune skin disease rare in most of the world but endemic in certain regions of Brazil. PF is characterized by the detachment of epidermal cells and the presence of autoantibodies against desmoglein 1. In previous studies, we have shown that genetic polymorphisms and variable expression levels of certain leukocyte receptor complex (LRC) genes were associated with PF. However, the role of the LRC on PF susceptibility remained to be investigated. Here, we analyzed 527 tag single nucleotide polymorphisms (SNP) distributed within the 1.5 Mb LRC. After quality control, a total of 176 SNPs were analyzed in 229 PF patients and 194 controls. Three SNPs were associated with differential susceptibility to PF. The intergenic variant *rs465169* (OR = 1.50, $p = 0.004$) is located in a region that might regulate several immune-related genes, including *VSTM1*, *LILRB1/2*, *LAIR1/2*, *LILRA3/4*, and *LENG8*. The *rs35336528* (OR = 3.44; $p = 0.009$) and *rs1865097* (OR = 0.57; $p = 0.005$) SNPs in *LENG8* and *FCAR* genes, respectively, were also associated with PF. Moreover, we found four haplotypes with SNPs within the *KIR3DL2/3*, *LAIR2* and *LILRB1* genes associated with PF ($p < 0.05$), which corroborate previously reported associations. Thus, our results confirm the importance of the LRC for differential susceptibility to PF and reveal new markers that might influence expression levels of several LRC genes, as well as candidates for further functional studies.

Key-words: LRC, 19q13.42, pemphigus foliaceus, *rs465169*, *FCAR, LENG8*

---

[1] Este capítulo é composto por um artigo publicado pela revista científica ***Immunology*** em setembro de 2018 (doi: 10.1111/imm.13003).

6 CAPÍTULO III

Mapeamento fino da região LRC na susceptibilidade ao pênfigo foliáceo

Neste capítulo serão abordados os desdobramentos do estudo de associação apresentado no capítulo II. Diversas variantes presentes no LRC, na região 19q13.42, foram associadas ao PF endêmico, destacando-se a variante intergênica *rs461569* localizada em uma região reguladora e entre os genes *CACNG6* e *VSTM1*. No presente capítulo, ao utilizar a expressão "região intergênica (RI)", estaremos nos referindo a essa região. Para buscar se a variante *rs461569* é a variante causal da associação, foi realizado o mapeamento fino dessa região. No presente capítulo, constam os resultados do mapeamento fino da RI e da validação do estudo de associação com variantes nos genes *LENG8* e *FCAR* presentes no LRC, em coorte independente de PF esporádico.

6.1 MATERIAIS E MÉTODOS

6.1.1 O mapeamento fino da região intergênica

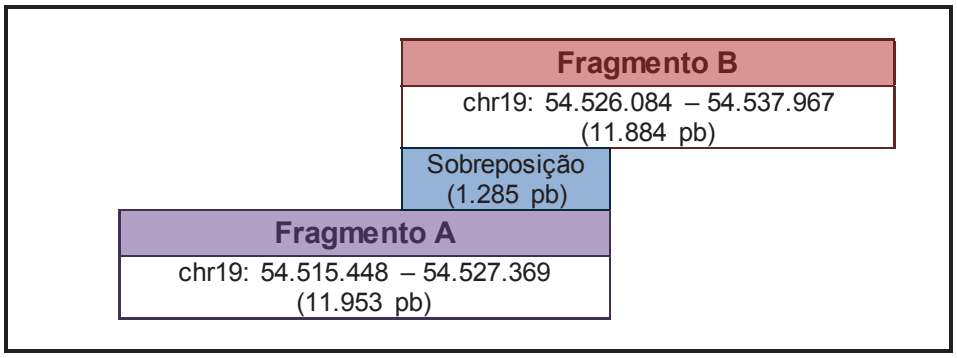
No estudo de associação da região 19q13.42 (Capítulo II), foi identificado um SNP intergênico localizado em uma região com potencial função reguladora de diversos genes do LRC. A variante *rs465169* está localizada entre e quase equidistante (11 kilobases (kb, 10^3 bases)) dos SNPs flanqueadores mais próximos (*rs459247* no gene *CACNG6* e *rs4239594* no gene *VSTM1*) que foram genotipados pelo microarranjo de DNA (FIGURA 11). Logo, há cerca de 22 kb de extensão entre os genes *CACNG6* e *VSTM1* sem informação de variabilidade genética exceto a variante *rs465169*. Por isso, decidiu-se pela realização de um mapeamento fino na RI em busca da descrição das variantes de sequência a fim de facilitar a identificação do possível SNP causal dessa associação genética. O seguimento deste projeto foi realizado com a colaboração com os professores MD PhD Enno Schmidt e PhD Hauke Busch do *Lübecker Institut für Experimentelle Dermatologie* (LIED), Universidade de Lübeck, Alemanha, em dois períodos de 5 meses em 2017 e 2018.

6.1.2 Sequenciamento da região intergênica

A região de 22 kb (chr19: 54.515.448 – 54.537.967, GRCh37/hg19) foi selecionada para o sequenciamento pelo método de sequenciamento paralelo massivo de nova geração (NGS, do inglês *next generation sequencing*) de uma região alvo (*targeted resequencing*).

Amostras de DNA de 48 indivíduos foram selecionadas para o sequenciamento, sendo 38 indivíduos saudáveis do banco de dados Painel do LGMH e 10 pacientes de PF com genótipo *rs465169 A/A*, com intuito de avaliar a presença de haplótipos preferenciais em amostras com este genótipo. Foram amplificados por PCR (reação em cadeia da polimerase, do inglês *polimerase chain reaction*) dois produtos de 12 kb de comprimento cada, com sobreposição de 1.285 pb (FIGURA 12). A PCR foi realizada com o uso da enzima polimerase para fragmentos longos GoTaq® Long PCR Master Mix Promega Corp. de acordo com o protocolo apresentado na TABELA 3. A sequência dos oligonucleotídeos utilizados para PCR estão na TABELA 4. As condições de ciclagem da PCR para ambos fragmentos estão na TABELA 5.

FIGURA 11: ESQUEMA DA LOCALIZAÇÃO GENÔMICA NO CROMOSSOMO 19 (CHR19) E TAMANHO DOS FRAGMENTOS A E B AMPLIFICADOS POR PCR PARA SEQUENCIAMENTO. EM AZUL ESTÁ INDICADA A SOBREPOSIÇÃO EM PARES DE BASES (PB) ENTRE OS FRAGMENTOS.



SNP	rs459247	rs465169		rs4239594
Posição no Chr19	54.515.874	54.526.972		54.537.079
Anotação	3'UTR *CACNG6*	intergênico		3' UTR *VSTM1*

Fonte: autora (2018)

Detalhes: Os SNPs presentes no chip microarranjo de DNA estão localizados na versão GRCh37/hg19 do genoma, refletindo sua posição dos fragmentos (A e B) amplificados por PCR.

TABELA 3: PROTOCOLO DE PCR COM CONCENTRAÇÃO DOS REAGENTES POR AMOSTRA.

Reagentes	Concentração Final
MasterMix	1X
Oligo_Fwd	0,5 µM
Oligo_Rev	0,5 µM
água	-
DNA	20 ng

Legenda: PCR (reação em cadeia da polimerase, do inglês *polimerase chain reaction*).

TABELA 4: SEQUÊNCIA DE OLIGONUCLEOTÍDEOS USADOS PARA AMPLIFICAÇÃO DE DOIS FRAGMENTOS (A E B) POR PCR

Fragmento B	
Forward (Fwd)	5' GGCTGCAATAAGTAAGCCAAG 3'
Reverse (Rev)	5' AAAGGCCCAACCTCTCAGAT 3'
Fragmento A	
Forward (Fwd)	5'ACGCGTGAGACTTCTCTAAGC 3'
Reverse (Rev)	5' GGATGACGCCTTTCGTTC 3'

Legenda: PCR (reação em cadeia da polimerase, do inglês *polimerase chain reaction*).

TABELA 5: PROTOCOLO DE PCR PARA AMPLIFICAÇÃO DOS DOIS FRAGMENTOS A E B POR.

Temperatura		Tempo	Número de ciclos
Fragmento B	Fragmento A		
94°C	94°C	2 min	1
94°C	94°C	15 seg	5
69°C	67°C	6 min	
94°C	94°C	15 seg	
66°C	65,7°C	6 min	30
68°C	68°C	2 min	

Legenda: PCR (reação em cadeia da polimerase, do inglês *polimerase chain reaction*).

Os produtos de PCR foram purificados com kit DNA Clean & Concentrator™-5 (Zymo Research Corp., EUA) de acordo com as informações do fabricante. Para avaliar a qualidade de pureza dos produtos de PCR, a densidade óptica (D.O.) pela incidência de luz U.V. foi obtida com auxílio do espectrofotômetro NanoPhotometer® (IMPLEN). Os critérios de qualidade considerados foram razão 260/280 acima de 1,7 e a razão 260/320 maior que 1,0.

Antes de iniciar o protocolo de sequenciamento, foi necessário medir a concentração de DNA dos produtos de PCR através do ensaio fluorescente de Qubit® dsDNA HS (*High Sensitivity*) (Invitrogen™, EUA), de acordo com o protocolo sugerido pelo fabricante.

As amostras foram diluídas na concentração de 0,2 ng/µl no volume de 5 µl para o preparo de biblioteca pelos kits Nextera® XT e Nextera® XT Index Kit v2 Set A

(Illumina Inc., EUA). O sequenciamento das amostras ocorreu nos dois sentidos (*paired-end)* utilizando o ensaio Reagent Nano Kit v2 para 300 ciclos no equipamento MiSeq® (Illumina Inc., EUA) localizado no LIED, Universidade de Lübeck, Alemanha.

Os dados gerados pelo sequenciamento foram avaliados por ferramentas de bioinformática pelos colaboradores Axel Künstner e Hauke Busch do grupo de pesquisa Biologia de Sistemas, LIED, Alemanha.

A região do genoma de referência (chr19: 54.515.448 – 54.537.967; GRCh37/hg19) foi extraída do banco de dados Ensembl (AKEN et al., 2016), adicionando-se mais 1000 pb às regiões flanqueadoras (chr19: 54.514.448 – 54.538.967, GRCh37/hg19).

Os dados do sequenciamento foram triados pela ferramenta Trimmomatic (BOLGER; LOHSE; USADEL, 2014) e os *reads* foram mapeados usando-se a ferramenta Burrows-Wheeler Alignment (bwa) (LI; DURBIN, 2009). Do arquivo de saída, os arquivos SAM (Sequence Alignment/Map format) foram filtrados, os pares de *reads* foram agrupados e as duplicações foram marcadas pela ferramenta Picard (http://broadinstitute.github.io/picard/). O alinhamento de inserção/deleção (indel) e a recalibração da qualidade da chamada de bases, dos SNPs descritos no dbSNP versão 138, foram realizados com o auxílio do algoritmo GATK (https://software.broadinstitute.org/gatk/). A identificação de SNPs e indels foi realizada com auxílio da ferramenta VarScan (KOBOLDT et al., 2009), considerando cobertura mínima de 10 vezes e valor de $p = 0,01$ como limiar. Os SNPs foram anotados de acordo com o banco de dados Seattle Seq Annotation versão 138 (NG et al., 2009).

6.1.3 Análise de correlação entre *rs465169* e outros SNPs localizados na RI

A partir dos dados do sequenciamento dos indivíduos controles, foi realizada a análise de desequilíbrio de ligação (DL) para verificar a correlação (r^2) entre os 45 SNPs identificados pelo sequenciamento e a variante *rs465169* associada ao PF com auxílio do software PLINK v1.9 (CHANG et al., 2015).

Além da análise de DL, levou-se em consideração a ocorrência de haplótipos preferenciais nas amostras de PF por visualização direta, uma vez que todos os pacientes eram homozigotos para o alelo *rs465169 A* associado a doença.

Os valores de DL, análises dos haplótipos em amostras de PF e possível função dos SNPs foram compilados com intuito de escolher quais variantes seriam potencialmente as causais da associação com PF. Dos SNPs anotados na região sequenciada, foram destacados SNPs com DL intermediário com *rs465169* ($0,6 < r^2 < 0,9$) com possível impacto funcional. Foram desconsiderados SNPs com forte DL com *rs465169* ($r^2 > 0,9$), pois seria possível inferir genótipos sem realizar genotipagens individuais. Os SNPs com fraco DL com *rs465169* ($r^2 < 0,6$) não seriam informativos na procura de SNP causal de associação. Adotando-se estes critérios, foram selecionados 5 SNPs (*rs2619223*, *rs465169*, *rs681198*, *rs7249071* e *rs4806488*) para novo estudo caso-controle.

6.1.4 Estudo de associação com variantes localizadas no LRC (19q13), incluindo a região 19q13.42

6.1.4.1 Caracterização da amostra utilizada no novo estudo de associação

Um segundo estudo caso-controle foi realizado com 208 casos de PFE e 140 controles do banco de dados do LGHM, dominado “Brasil”. A maioria dos quais, 188 casos de PFE e 133 controles, estavam presentes no ensaio de microarranjo de DNA realizado anteriormente (ver seção Métodos – Capítulos I e II) (Apêndice C). Além das amostras presentes no microarranjo, foram incluídos 13 casos de PFE e 5 controles cujo RNA mensageiro (RNAm) estaria disponível para ensaios de expressão gênica e outros 7 casos de PFE e 2 controles para os quais há dados de expressão gênica publicados por Malheiros et al. (2014). A descrição dessa amostra encontra-se no Apêndice G (TABELA S1 e FIGURA S1). Esta amostra foi utilizada para genotipar os SNPs selecionados a partir do mapeamento fino da RI e ausentes no chip do microarranjo, incluindo *rs461569*. No entanto, para os demais SNP associados ao PF endêmico descritos no capítulo II, esta genotipagem servirá como validação técnica.

Para fins de comparação entre as duas formas de PF – endêmico e esporádico – foram genotipados 76 pacientes PF esporádico e 150 controles pareados por idade e sexo e provenientes do Biobanco PopGen, Alemanha (mais detalhes TABELA S1 e FIGURA S2 – APÊNDICE F). Nós denominamos esta amostra como “Alemanha”. Este estudo foi aprovado pelo Comitê de Ética em pesquisa sobre seres humanos da UFPR

e pelo comitê de ética da Universidade de Lübeck, Alemanha (Biobanco PopGen, número 2017-018).

6.1.4.2 Variantes selecionadas para segunda genotipagem

Foram selecionados 5 SNPs localizados na RI próximos ao SNP *rs465169*, supracitados no tópico 6.1.3. Dois SNPs associados localizados nos genes *LENG8* e *FCAR* associados previamente ao PF (Capítulo II) foram incluídos neste estudo. O SNP (*rs76692276*), também localizado no gene *LENG8*, foi incluído em virtude de seu possível impacto funcional na susceptibilidade ao PF.

6.1.4.3 Genotipagem

A genotipagem de múltiplos SNPs (multiplex) foi realizada pela plataforma iPLEX® do sistema MassARRAY® (Agena Bioscience, San Diego, CA, EUA) no Insitut für Klinische Molekularbiologie (IKMB), Christian-Albrechts-Universität, Kiel, Alemanha. Os dados brutos da genotipagem foram analisados com auxílio do software SpectroTYPER v4.0 (Agena Bioscience, San Diego, CA). O SNP rs76692276 foi genotipado pelo ensaio hidrólise de sonda TaqMan® (Applied Biosystems™, EUA).

Os resultados da genotipagem dos SNPs obtidos pela plataforma iPLEX foram conferidos em todas as amostras. Para cada SNP, verificou-se:

a) o padrão de agrupamento das amostras nos grupos de homozigotos e heterozigotos;
b) qualidade e amplitude do pico de cada alelo;
c) conferência da genotipagem por visualização direta, considerando o padrão entre amostras e entre SNPs;
d) comparação de resultados obtidos pela técnica iPLEX com outro método (microarranjo de DNA e sequenciamento).

Foram excluídos resultados nos quais o pico do oligonucletídeo de extensão possuía maior amplitude que a dos alelos específicos, ou com picos duvidosos e/ou semelhantes ao ruído (*background*) do experimento. Além disso, foram excluídos resultados nos quais as frequências alélicas obtidas pelo iPLEx diferiram muito do

esperado, seja pelo teste de EHW e por comparação com frequência do conjunto da população Europeia (CEU, TSI, FIN, GBR, IBS) do projeto 1000 genomas por consulta ao banco de dados Ensembl (http://www.ensembl.org).

6.1.4.4 Análise estatística do estudo caso-controle

As frequências alélicas e genotípicas foram obtidas por contagem direta e o equilíbrio de Hardy-Weinberg (EHW) foi testado nos grupos de pacientes e controles pelo software PLINK v1.9 (CHANG et al., 2015). A análise de associação foi realizada pelo cálculo da *odds ratio* (OR) nos modelos aditivo, dominante e recessivo com intervalo de confiança (IC) de 95% e valor de $p = 0,05$ como limite de significância com auxílio do software PLINK v1.9 (CHANG et al., 2015) e VassarStats (VassarStats: Website for Statistical Computation). A significância estatística foi avaliada pelo teste exato de Fisher. O DL entre SNPs foi calculado com auxílio do software PLINK v1.9 (CHANG et al., 2015).

Foram comparadas amostras de pacientes e controles nas seguintes situações (a) Amostras do Brasil: PFE e controles do LGMH/UFPR e (b) Amostras da Alemanha, Alemanha: PF esporádico e controles do Biobanco PopGen.

6.1.4.5 Análises *in silico* para impacto funcional dos SNPs

As análises *in silico* dos SNPs associados no Capítulo II estão descritas no respectivo capítulo. Logo, aqui será descrito apenas o possível impacto dos SNPs presentes no mapeamento fino da RI flaqueadora do SNP *rs465169* em busca da(s) variante(s) causal(is) e do SNP *rs76692276* do gene *LENG8*.

Os possíveis impactos funcionais das variantes associadas com PF endêmico e esporádico foram avaliados com auxílio das ferramentas online SNiPA, RegulomeDB, GTEx, HaploReg (ARNOLD et al., 2015; BOYLE et al., 2012; CONSORTIUM et al., 2015; WARD; KELLIS, 2012). Foi avaliado o DL entre os SNPs presentes na região intergênica com SNPs do conjunto da população europeia (CEU, TSI, FIN, GBR, IBS) do projeto 1000 Genomas pela ferramenta *online* LDLink (MACHIELA; CHANOCK, 2015).

6.2 RESULTADOS

6.2.1 LRC e região intergênica (19q13.42)

As frequências genotípicas dos SNPs estão em EHW em todas amostras analisadas ($p > 0,05$), exceto para o SNP *rs2619223* em controles da Alemanha (TABELA 6) e *rs1865097* nas amostras de PF esporádico e controles da Alemanha ($p = 0,037$ e $p = 0,001$, respectivamente) (TABELA 7). As frequências alélicas e genotípicas desses quatro SNPs estão apresentadas nas TABELAS 6 e 7. Não foi possível obter dados de genotipagem do SNP *rs4806488*.

TABELA 6: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DOS SNPS LOCALIZADOS NA REGIÃO INTERGÊNICA DO LRC (19q13.42), DAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (PFE) DO BRASIL E ESPORÁDICO (PF) DA ALEMANHA E CONTROLES (CTR).

	Brasil		**Alemanha**	
	PFE (F%)	CTR (F%)	PF n (F%)	CTR n (F%)
rs2619223	n=201	n=135	n=49	n=103
G	210 (52,2)	112 (41,5)	59 (60,2)	90 (43,7)
G/G	54 (26,9)	23 (17,0)	16 (32,7)	25 (24,3)
G/C	102 (50,7)	66 (48,9)	27 (55,1)	40 (38,8)
C/C	45 (22,4)	46 (34,1)	6 (12,2)	38 (36,9)
p-EHW	0,888	1,000	0,381	0,044
rs465169	n=183	n=124	n=48	n=100
A	184 (50,3)	96 (38,7)	50 (52,1)	88 (44,0)
A/A	48 (26,2)	22 (17,7)	12 (25,0)	24 (24,0)
A/G	88 (48,1)	52 (41,9)	26 (54,2)	40 (40,0)
G/G	47 (25,7)	50 (40,3)	10 (20,8)	36 (36,0)
p-EHW	0,660	0,190	0,772	0,068
rs681198	n=198	n=134	n=51	n=118
G	202 (51,0)	112 (41,8)	58 (56,9)	99 (41,9)
G/G	49 (24,7)	24 (17,9)	14 (27,5)	26 (22,0)
G/A	104 (52,5)	64 (47,8)	30 (58,8)	47 (39,8)
A/A	45 (22,7)	46 (34,3)	7 (13,7)	45 (38,1)
p-EHW	0,570	0,860	0,252	0,058
rs7249071	n=200	n=136	n=64	n=137
A	127 (31,8)	65 (23,9)	42 (32,8)	75 (27,4)
A/A	18 (9,0)	6 (4,4)	4 (6,3)	11 (8,0)
A/T	91 (45,5)	53 (39,0)	34 (53,1)	53 (38,7)
T/T	91 (45,5)	77 (56,6)	26 (40,6)	73 (53,3)
p-EHW	0,620	0,490	0,157	0,830

SNP: *single nucleotide polymorphisms*; *p*-EHW: valor de *p* no teste de Equilíbrio de Hardy-Weinberg. Em negrito os valores de $p < 0,05$

TABELA 7: FREQUÊNCIAS ALÉLICAS E GENOTÍPICAS DOS SNPS LOCALIZADOS NOS GENES *FCAR* E *LENG8*, 19Q13.42, DOS GRUPOS DE AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (PFE) DO BRASIL E ESPORÁDICO (PF) DA ALEMANHA E CONTROLES (CTR).

	Brasil		**Alemanha**	
	PFE n (F %)	CTR n (F %)	PF n (F %)	CTR n (F %)
FCAR				
rs1865097	n=203	n=137	n=66	n=139
A	138 (34,0)	101 (36,9)	51 (38,6)	112 (40,3)
A/A	27 (13,3)	16 (11,7)	14 (21,2)	33 (23,7)
A/G	84 (41,4)	69 (50,4)	23 (34,8)	46 (33,1)
G/G	92 (45,3)	52(38,0)	29 (43,9)	60 (43,2)
p-EHW	0,270	0,460	0,037	**0,001**
LENG8				
rs35336528	n=202	n=135	n=52	n=105
G	17 (4,2)	5 (1,9)	5 (4,8)	7 (3,3)
G/G	0 (0,0)	0 (0,0)	0 (0,0)	1 (1,0)
G/A	17 (8,4)	5 (3,7)	5 (9,6)	5 (4,8)
A/A	185 (91,6)	130 (96,3)	47 (90,4)	99 (94,3)
p-EHW	NC	NC	NC	NC
rs76692276	n=192	n=110	n=70	n=120
C	8 (2,1)	6 (2,7)	8 (5,7)	3 (1,3)
C/C	0 (0,0)	0(0,0)	0 (0,0)	0 (0,0)
C/T	8 (4,2)	6 (5,5)	8 (11,4)	3 (92,5)
T/T	184 (95,8)	104 (94,5)	62 (88,6)	117 (97,5)
p-EHW	NC	NC	NC	NC

SNP: *single nucleotide polymorphisms*. *p*-EHW: valor de *p* no teste de Equilíbrio de Hardy-Weinberg. NC: EWH não calculado para MAF < 5%. Em negrito os valores de $p < 0,05$.

As 4 variantes da RI foram associadas ao PF endêmico na amostra do Brasil (OR > 1,5, $p < 0,05$) e somente os SNPs *rs2619223* e *rs681198* foram associados com PF esporádico na amostra da Alemanha (OR = 1,95 e OR = 1,82, $p < 0,05$, respectivamente), nos modelos aditivo e dominante (TABELA 8). Como os SNPs da RI estão próximos fisicamente entre si, avaliamos o DL entre eles. Os SNPs *rs681198* e *rs2619223* estão em moderado DL na amostra controle do Brasil (r^2 =0,77) e em moderado DL na amostra da Alemanha (r^2 = 0,75) (TABELA 9).

Os SNPs *rs1865097* (*FCAR*) e *rs35336528* (*LENG8*) não foram associados em amostra alguma. O SNP *rs76692276* (*LENG8*) está associado a maior susceptibilidade ao PF esporádico nos modelos aditivo e dominante (OR = 4,79 e OR = 5,03, $p = 0,02$, respectivamente) (TABELA 10). Os SNPs no gene *LENG8* não estão em DL ($r^2 < 0,2$).

TABELA 8: RESULTADOS DE NOVA ANÁLISE DE ASSOCIAÇÃO ENTRE PF E SNPS LOCALIZADOS NA REGIÃO INTERGÊNICA (RI) DO LRC, EM 19q13.42, NAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (PFE) DO BRASIL E ESPORÁDICO (PF) DA ALEMANHA.

		Brasil			**Alemanha**		
		OR	IC 95%	p	OR	IC 95%	p
rs2619223							
Aditivo	G	1,54	1,13 – 2,11	**0,007**	1,95	1,20 – 3,18	**0,010**
Dominante	G+	1,79	1,10 – 2,91	**0,018**	4,19	1,63 – 10,76	**0,002**
Recessivo	G/G	1,79	1,04 – 3,09	**0,036**	1,51	0,72 – 3,20	0,277
rs465169							
Aditivo	A	1,60	1,15 – 2,22	**0,005**	1,38	0,85 – 2,25	0,214
Dominante	A+	1,96	1,20 – 3,19	**0,007**	2,14	0,95 – 4,79	0,087
Recessivo	A/A	1,65	0,94 – 2,90	0,082	1,06	0,48 – 2,35	0,894
rs681198							
Aditivo	G	1,45	1,06 – 1,98	**0,022**	1,82	1,14 – 2,92	**0,013**
Dominante	G+	1,78	1,09 – 2,89	**0,020**	3,87	1,61 – 9,34	**0,002**
Recessivo	G/G	1,51	0,87 – 2,60	0,140	1,34	0,63 – 2,84	0,555
rs7249071							
Aditivo	A	1,48	1,05 – 2,10	**0,030**	1,30	0,82 – 2,04	0,289
Dominante	A+	1,56	1,01 – 2,42	**0,045**	1,67	0,91 – 3,04	0,099
Recessivo	A/A	2,14	0,83 – 5,55	0,109	0,76	0,23 – 2,50	0,779

SNP: *single nucleotide polymorphisms*, OR: *odds ratio*, IC: intervalo de confiança. Em negrito valor de $p < 0,05$.
Modelo dominante e recessivo refere-se ao efeito dominante/recessivo dos alelos de menor frequência neste estudo: *rs2619223 G*, *rs465169 A*, *rs681198 G*, *rs7249071 A*.

TABELA 9: RESULTADOS DA CORRELAÇÃO (R^2) PARA AVALIAR O DESEQUILÍBRIO DE LIGAÇÃO ENTRE OS SNPS NA REGIÃO INTERGÊNICA DO LRC (19q13.42) NAS AMOSTRAS DE CONTROLES DE PF ENDÊMICO (BRASIL) E ESPORÁDICO (ALEMANHA).

		Brasil	**Alemanha**
SNP1	SNP2	r^2	r^2
rs465169	*rs2619223*	**0,80**	0,62
rs465169	*rs681198*	0,72	0,58
rs465169	*rs7249071*	0,44	0,45
rs681198	*rs2619223*	0,77	0,75
rs7249071	*rs2619223*	0,35	0,34
rs7249071	*rs681198*	0,31	0,39

SNP: *single nucleotide polymorphism*. Em negrito valores $r^2 > 0,80$.

TABELA 10: RESULTADOS DA ANÁLISE DE ASSOCIAÇÃO DOS SNPS LOCALIZADOS NOS GENES *FCAR* E *LENG8* DO LRC, 19q13.42, NAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (PFE) DO BRASIL E ESPORÁDICO (PF) DA ALEMANHA.

		Brasil			**Alemanha**		
		OR	IC 95%	p	OR	IC 95%	p
FCAR							
rs1865097							
Aditivo	A	0,88	0,64 – 1,22	0,462	0,93	0,61 – 1,43	0,829
Dominante	A+	0,74	0,47 – 1,15	0,178	0,97	0,54 – 1,75	0,917
Recessivo	A/A	1,16	0,60 – 2,25	0,659	0,86	0,43 – 1,76	0,687
LENG8							
rs35336528							
Aditivo	G	2,33	0,85 – 6,39	0,121	1,47	0,45 – 4,73	0,541
Dominante	G+	2,39	0,86 – 6,64	0,115	1,76	0,51 – 6,05	0,507
Recessivo	G/G	NC	NC	NC	1,51	0,06 – 37,64	1,000
rs76692276							
Aditivo	C	0,76	0,26 – 2,22	0,588	4,79	1,25 – 18,36	**0,022**
Dominante	C+	1,01	0,34 – 2,99	1,000	5,03	1,29 – 19,65	**0,020**
Recessivo	C/C	NC	NC	NC	NC	NC	NC

SNP: *single nucleotide polymorphism*, OR: *odds ratio*, IC: intervalo de confiança. NC: não calculado, pois a frequência do genótipo homozigoto era baixa. Em negrito valor de $p < 0,05$.
Modelo dominante e recessivo refere-se ao efeito dominante/recessivo dos alelos de menor frequência: *rs1865097 A*, *rs35336528 G*, *rs76692276 C*.

6.2.2 Impacto funcional dos SNPs localizados na região intergênica (19q13.42)

Diversos SNPs na RI associados ao PF, seja endêmico ou esporádico, poderiam causar alterações na regulação gênica, pois estão em sítios de ligação de fator de transcrição (CTCF), proteínas reguladoras de transcrição (GATA, PAX-4, etc.) e sensíveis a DNase (*rs2619223*) em células sanguíneas e tecido epitelial, principalmente (TABELA 11; dados completos na TABELA S1 – S3, APÊNDICE G). Além disso, os SNPs avaliados neste trabalho estão em forte DL com outros SNPs ($r^2 > 0,9$) nas populações eurodescendentes, indicando que eles podem marcar associação de um bloco de SNPs (haplótipo) importante na regulação de genes que por isso poderia contribuir na patogênese do PF. Por exemplo, os SNPs *rs2619223* e *rs681198* estão em forte desequilíbrio de ligação ($r^2 > 0,90$) com diversos outros SNPs nas populações europeias e eurodescendentes do banco de dados 1000 genomas (MACHIELA; CHANOCK, 2015), formando um bloco de SNPs em um segmento de 7,423 kb. Ambos SNPs estão em moderado DL com o SNP *rs465169* na amostra brasileira ($r^2 \sim 0,75$) quanto na alemã ($r^2 \sim 0,60$) (TABELA 9).

TABELA 11: RESUMO DOS POSSÍVEIS EFEITOS FUNCIONAIS DOS SNPS *rs4806488*, *rs465169*, *rs7249071*, *rs681198*, *rs2619223* LOCALIZADOS NA REGIÃO INTERGÊNICA (RI) DO LRC, EM 19q13.42.

SNP	Posição (cromossomo 19)	Alelos	MAF	Distância (pb)	DL (r2)	SNiPA		Regulome DB Score	HaploReg						
						Efeito regulador	Genes potenc. regulados		Marca de histonas Promotor	Marca de histonas Intensificador	DNAse	Ligação proteína	Motivos modificados	eQTL	GENCODE genes
rs4806488	54526420	T, C	0.37	-552	0.70	Potencial		5		Células sanguíneas, tecido epitelial, muscular, digestivo e outros	músculo		TEF-1, Zbtb3		10kb 3' do CACNG6
rs465169	54526972	G, A	0.37	0		ENCODE distal promotor (intensificador)	LENG8, LILRA3/5/6, LAIR1, LILRB2, NLRP12, OSCAR, VSTM1	4	Células sanguíneas, pulmão,		Diversas células e tecidos		Maf	CACNG6	11kb 3' do CACNG6
rs7249071	54527521	T, A	0.27	549	< 0.2	sítio de ligação CTCF		2b	Células sanguíneas, outros			ZNF263	EWSR1, GATA, HDAC2, PAX-4, Pbx3		12kb 3' do CACNG6
rs681198	54527948	A, G	0.40	976	0.77	Intensificador, sítio de ligação CTCF		4		Células sanguíneas, cérvix	Células sanguíneas cérvix		Ascl2, HEN1		12kb 3' do CACNG6
rs2619223	54528930	C, G	0.39	1958	0.71	-		5			Tecido epitelial		CHD2, Ik-1, PU.1, Pax-4, RREB-1, SP1, TFII, outros		13kb 3' do CACNG6

Efeitos reguladores preditos dos SNPs (*single nucleotide polymorphism*) *rs4806488*, *rs465169*, *rs7249071*, *rs681198*, *rs2619223* presentes na região intergênica (RI) associada com pênfigo foliáceo. Nesta tabela os é possível ver o desequilíbrio de ligação (DL) entre os SNPs com SNP *rs465169* de acordo com os resultados do sequenciamento de nova geração (NGS) realizado neste trabalho. Além disso, é mostrada a posição genômica, distância deles com o SNP *rs461569* em pares de bases (pb) e predição de seus efeitos em elementos reguladores. Os dados foram obtidos pelas ferramentas *online*: SNiPA, RegulomeDB e HaploReg.

Posição: localização em pares de base de cada SNP no cromossomo (chr) 19 (versão do genoma GRCh37/hg19)

eQTL: expression quantitative trait loci

GENCODE: Enciclopédia dos genes e suas variantes

6.3 DISCUSSÃO

6.3.1 Mapeamento fino da região 19q13.42

O sequenciamento e análise da região flanqueadora do SNP *rs465169* (RI) possibilitou a seleção de 5 SNPs que foram genotipados em dois grupos amostrais: uma amostra de PF endêmico (Brasil) e uma amostra de PF esporádico (Alemanha) e as respectivas amostras controle. Os 5 SNPs foram: *rs2619223*, *rs681198*, *rs7249071*, *rs4806488*, incluindo o *rs465169*. No entanto, não foi possível utilizar os dados gerados da genotipagem do SNP *rs4806488*, devido à baixa qualidade dos dados de genotipagem.

Interessantemente, o alelo *rs461569 A* que está associado ao aumento da susceptibilidade ao PF endêmico (Capítulo II) não foi associado ao PF esporádico, apesar de estar próximo da significância (OR = 2,14, p = 0,087). Não está excluída a possibilidade que o SNP *rs465169* possa ter também um efeito na susceptibilidade ao PF esporádico, entretanto, a confirmação dependerá de análise de uma amostra maior.

Dois dos quatro SNPs genotipados, *rs2619223* e *rs681198*, que estão em DL nas duas populações estudadas ($r^2 > 0,75$), foram associados com aumento da susceptibilidade ao PFE (Brasil) e PF esporádico (Alemanha), indicando serem essas variantes os fatores genéticos de susceptibilidade compartilhados pelas duas formas da doença nas duas populações. Os dois SNPs aparentemente conferem uma maior susceptibilidade ao PF esporádico (OR ~ 4,0, $p < 0,02$) quando comparado aos seus efeitos na susceptibilidade ao PF endêmico (OR ~ 1,8, $p < 0,02$). Com base nestes resultados, os SNPs *rs2619223* e/ou *rs681198* são fortes candidatos a serem as variantes causais da associação da RI com o PF, juntamente com SNP *rs465169*. No entanto, estes SNPs estão em forte DL entre com outros, indicando que a RI contém variantes importantes na susceptibilidade ao PF. Ambos SNPs, *rs2619223* e *rs681198*, foram associados tanto no modelo dominante quanto no aditivo, porém em PF esporádico o valor de OR foi consideravelmente maior no modelo dominante em comparação ao modelo aditivo. Isso demonstra que há um efeito de dominância incompleta na susceptibilidade ao

PF: homozigotos para o alelo menos comum são mais suscetíveis que os heterozigotos, principalmente no PF esporádico.

O SNP *rs681198* está situado em um sítio de ligação para o CTCF e de alteração de conformação de cromatina em células sanguíneas, de acordo com HaploReg (WARD; KELLIS, 2012). CTCF (*CCCTC-binding factor*) é um fator de transcrição (FT) importante em diversos processos biológicos que pode agir como repressor bem como ativador em regiões reguladoras, dependendo do promotor e contexto biológico. Esse FT participa em vias de modificações na estrutura de cromatina, integridade dos cromossomos e liga-se em regiões reguladoras de longo alcance em mamíferos (CANELA et al., 2017; FILIPPOVA et al., 1996; FU et al., 2008; OHLSSON; RENKAWITZ; LOBANENKOV, 2001). O papel de regulador de expressão gênica em dimensão cromossômica, coloca este FT como importante proteína de interação na região reguladora que RI está alocado, como intensificador de promotor distal do projeto ENCODE anotado para o SNP *rs465169* (TABELA 11).

CTCF é altamente expresso na pele exposta e não-exposta ao sol de acordo com o Portal GTEx (CONSORTIUM et al., 2015). A proteína está presente em diversos tipos celulares do tecido epitelial, sendo a maior expressão por melanócitos e células de Langerhans. Dentre as células da epiderme, 60% dos níveis de RNAm *CTCF* foram encontrados em queratinócitos, de acordo com Human Protein Atlas (UHLÉN et al., 2015; THUL et al., 2017). Embora CTCF seja amplamente expresso em diferentes tecidos e contextos, ele participa na regulação de genes do complexo de diferenciação epidérmico (EDC, do inglês *epidermal differentiation complex*) em queratinócitos de ratos (POTERLOWICZ et al., 2017). EDC contém 61 genes localizados no cromossomo 3 em ratos e no cromossomo 1 em humanos, cujos produtos participam no controle de diferenciação epidérmica e na integridade da barreira física (DE STRONG et al., 2010; SEGRE, 2006). Logo, a CTCF regula genes importantes na manutenção e diferenciação das células da epiderme (KIM et al., 2015). Mutação na região de ligação do CTCF altera a sua afinidade à região reguladora em câncer de pele melanoma (POULOS et al., 2016), no entanto, se o alelo da variante associada ao PF aumenta ou diminui ação do CTCF é uma questão que necessita de investigação.

A variante *rs681198 G* alteraria o motivo de ligação de proteína ASCL2 (*Achaete-Scute Family BHLH Transcription Factor 2*) (WARD; KELLIS, 2012). ASCL2 é altamente expressa na pele exposta e não-exposta ao sol, de acordo com GTEx (CONSORTIUM et al., 2015), especialmente pelas células epiteliais presentes no estrato granular, camada mais superficial da epiderme que acomete no PF (MORIYAMA et al., 2008). Além disso, o produto do gene *ASCL2* poderia participar na regulação da expressão de genes do LRC, pois é coexpresso com *LAIR2*, *KIR2DL1* e *LILRA6* de acordo com o GeneMANIA (WARDE-FARLEY et al., 2010; MALLON et al., 2013; WU et al., 2007). Maiores níveis de ASCL2 são encontrados em monócitos de pacientes com a condição autoimune síndrome de Sjögren do que controles (KIM et al., 2017). Além disso, ASCL2 possui um importante papel em células T foliculares auxiliares (Thf), inibindo a expressão de genes de resposta Th1 e Th17 (LIU et al., 2014). O perfil de expressão gênica mostrou diferença dos níveis de RNAm de diversos genes relacionados com resposta de células Th17 entre biópsias da lesão e da região perilesional do mesmo paciente de PF (Malheiros, comunicação pessoal). As citocinas da resposta Th17 contribuem tanto ao início quanto na manutenção da inflamação da pele (CHO et al., 2012). A maior expressão de ASCL2 também aumenta os níveis de citocinas pró-inflamatórias, como IL-6 e TNFα (KIM et al., 2017). Logo, esta proteína tem um papel importante na regulação desses genes de resposta pró-inflamatória e na diferenciação celular epitelial da camada superficial, importantes em doença da pele como o PF.

O SNP *rs2619223* G altera motivos de ligação de diversas proteínas com função reguladora, como: PU.1/SPI1, CHD2 e RREB-1. O fator de transcrição PU.1 (SPI1, *spleen focus forming virus proviral integration oncogene*) é altamente expresso em células sanguíneas (CONSORTIUM et al., 2015) e é coexpresso com genes do LRC, tais como *LILRB1*, *LILRA2 LILRA4*, *LAIR1/2*, *OSCAR* em diversos tecidos e doenças de acordo com GeneMANIA (WARDE-FARLEY et al., 2010). PU.1 interfere na função de outros FT e regula genes que participam nos processos de diferenciação e ativação de macrófagos e células B (DEKOTER; SINGH, 2000; OIKAWA et al., 1999; ROSENBAUER et al., 2005). As células B são importantes na resposta autoimune de pênfigo, seja na produção de autoanticorpos e na ativação

de respostas pelas células T CD4+. A depleção dessas células é o principal alvo de terapias de nova geração no tratamento do pênfigo (HEBERT; JOLY, 2017; MUSETTE; BOUAZIZ, 2018).

Além disso, *rs2619223* situa-se em sítio de ligação de proteínas que remodelam a cromatina, tais como helicase CHD2 (*Chromodomain-helicase-DNA-binding protein 2*) (CARVILL et al., 2013) e RREB1 (*ras-responsive element binding protein 1*). RREB1 se liga ao promotor do gene *HLA-G*, reprimindo a expressão da proteína imunomoduladora HLA-G, ao recrutar proteínas do complexo CTBP (*C-terminal binding protein*) que remodelam a estrutura da cromatina (FLAJOLLET et al., 2009). No epitélio, *rs2619223* localiza-se em uma região sensível à DNase, indicando que este sítio está com a cromatina mais aberta, logo, este fragmento de DNA está mais acessível aos processos de transcrição e regulação, de acordo com a ferramenta online HaploReg (WARD; KELLIS, 2012).

Os resultados do estudo de associação, aliados às observações sobre o possível efeito funcional dos SNPs associados, mostram que a região reguladora tem um papel importante na susceptibilidade ao PF. No momento, estamos aprofundando a análise dos possíveis mecanismos que pudessem explicar os efeitos dos alelos/genótipos associados ao PF. Para tanto, estão em andamento análises de bioinformática do possível efeito eQTL (do inglês *expression quantitative trait loci*) dos alelos/genótipos dos quatro SNPs na expressão de genes do projeto ENCODE (TABELA S1 – S3 no APÊNDICE G), que serão integrados com informações sobre genes na região 19q13 diferencialmente expressos em PF (MALHEIROS et al., 2014). Em paralelo, está sendo realizado sequenciamento de RNA total de células T CD4+ de pacientes de PFE e controles do qual será possível extrair dados de expressão de genes do LRC potencialmente regulados pela esta RI. Esses trabalhos em andamento estão sendo desenvolvidos em colaboração com o LIED, Alemanha, e seus resultados deverão compor o manuscrito de mapeamento fino da RI, 19q13.42.

6.3.2 Polimorfismos nos genes *LENG8* e *FCAR*

Os SNP *rs1865097* (*FCAR)* e *rs35336528* (*LENG8*) que foram associados previamente ao PF endêmico (Capítulo II), não estiveram associados com PF em nenhuma das duas amostras (TABELA 10). Esta foi a primeira vez que analisamos a susceptibilidade do SNP *rs76692276 (LENG8*) ao PF. O alelo *rs76692276 C* (*LENG8*) foi associado com susceptibilidade ao PF esporádico. Para o alelo *rs35336528 G* do gene *LENG8* fora observada associação com maior susceptibilidade ao PFE nos modelos aditivo e dominante (OR = 3,39 e OR = 3,44, respectivamente $p < 0,01$). Por outro lado, o genótipo *rs1865097 A/A* (*FCAR*) aparentemente diminuía a susceptibilidade ao PFE (OR = 0,54, $p = 0,005$) (Capítulo II). Considerando que a amostra de PF endêmico de Curitiba analisada neste trabalho é uma subamostra da amostra utilizada no microarranjo de DNA (ver seção 6.1.4.1), essa análise serviria com validação técnica. Porém, os resultados obtidos no Capítulo II não foram replicados para estes SNPs. Possíveis causas dessa discrepância serão discutidas a seguir.

O alelo de menor frequência do SNP *rs35336528* do gene *LENG8* foi pouco frequente nas amostras de PF endêmico e esporádico, bem como na população total e de europeus do projeto 1000 genomas (MAF < 5%) (AKEN et al., 2016). Por causa da baixa MAF, a inclusão/exclusão de poucos indivíduos da amostra de pacientes e de controles resultou em alteração das frequências alélicas e genotípicas do SNP *rs35336528* da subamostra da amostra de PF endêmico (Brasil) o que por sua vez resultou em alteração do resultado da análise de associação. Portanto, é possível que a associação inicial tenha sido espúria.

Por outro lado, o alelo *rs76692276 C* (*LENG8*) foi associado ao aumento da susceptibilidade ao PF esporádico nos modelos aditivo e dominante (OR = 4,79 e OR = 5,03, $p = 0,02$), porém não com o PFE. Entretanto, assim como comentado acima para o SNP *rs35336528,* a MAF é baixa em nas populações deste estudo, assim como em populações do projeto 1000 genomas. O SNP *rs76692276* está numa região reguladora onde se ligam diversos fatores de transcrição (como PU.1), na qual há alteração da conformação de cromatina e região sensível à DNase em todos os tecidos e tipos celulares descritos na ferramenta online HaploReg (WARD;

KELLIS, 2012). Essa variante também está localizada em intensificador distal associado ao promotor que potencialmente regula a expressão dos genes *LILRA3* e *LENG9*, de acordo com o projeto ENCODE (ARNOLD et al., 2015). A possível associação com pênfigo, assim como o impacto funcional dessa variante necessitam ser melhor explorados.

O SNP *rs1865097* do gene *FCAR* não foi associado com PF endêmico nem esporádico. As frequências genotípicas deste SNP não estão em EHW na amostra controle alemã (p = 0,0004), observando-se um excesso de homozigotos G/G. As genotipagens desse SNP foram conferidos manualmente, conforme descrito na secção 6.2.3 (Capítulo III) o que diminui a possibilidade de erro na definição de genótipos. Porém, não é descartada a possibilidade da limitação da técnica ou baixa qualidade das amostras de DNA desse grupo. Este desvio da distribuição das frequências nos controles pode levar ao um viés na análise de associação, de maneira que os resultados das análises de associação não foram conclusivos.

7 CONCLUSÕES

O rastreamento e análise de variantes presentes na região do complexo de receptores leucocitários (LRC, do inglês *leucocyte receptor complex*) estendido (19q13), incluindo a sub-região do LRC (19q13.42), possibilitou a identificação de variantes que modulam a susceptibilidade ao PF endêmico e esporádico que dificilmente seriam apontadas em estudos com genes candidatos.

Na região do LRC estendido, variantes de genes de RNA não-codificadores (ncRNA), na região que contém genes *IGFL*s e pseudogene *RSP9P4*, e de genes codificadores de proteínas (*SULT21B*, *SCAF1* e *RDH13*) foram associados a susceptibilidade diferencial ao PF endêmico e não na forma esporádica.

Na região do LRC, três novas variantes foram associadas ao PFE, duas nos genes *FCAR* e *LENG8* e uma variante *rs461569* entre os genes *CACNG6* e *VSTM1*. Análise de haplótipos confirmou associações com PFE de variantes de *KIR* e *LAIR2*, previamente estudados por nosso grupo.

As duas formas de PF compartilham susceptibilidade genética na região intergênica (RI), 19q13.42. Dois dos cinco SNP avaliados na RI, *rs2619223* e *rs681198*, foram associados tanto com PF endêmico quanto com PF esporádico, indicando duas variantes candidatas para serem exploradas em análises funcionais.

Por fim, este trabalho confirmou que na região cromossômica 19q13 encontram-se diversos locos de interesse para estudos funcionais, cujas variantes potencialmente influenciam a regulação da expressão gênica, assim contribuindo para a patogênese do pênfigo.

REFERÊNCIAS BIBLIOGRÁFICAS

ABRÈU-VELEZ, A. et al. A unique form of endemic pemphigus in northern Colombia. **Journal of the American Academy of Dermatology**, v. 49, n. 4, p. 599–608, 2003.

ABRÉU-VÉLEZ, A. M. et al. Endemic pemphigus foliaceus over a century: Part I. **North American journal of medical sciences**, v. 2, n. 2, p. 51–9, 2010.

AKEN, B. L. et al. The Ensembl gene annotation system. **Database**, v. 2016, p. baw093, 2016. Disponível em: <http://www.ensembl.org/index.html> Acesso em agosto de 2018.

AL-JASSAR, C. et al. Mechanistic Basis of Desmosome-Targeted Diseases. **Journal of Molecular Biology**, v. 425, n. 21, p. 4006–4022, 2013.

AMAGAI, M. et al. Autoantibodies against the amino-terminal cadherin-like binding domain of pemphigus vulgaris antigen are pathogenic. **Journal of Clinical Investigation**, v. 90, n. 3, p. 919–926, 1992.

AMAGAI, M. et al. Use of autoantigen-knockout mice in developing an active autoimmune disease model for pemphigus. **Journal of Clinical Investigation**, v. 105, n. 5, p. 625–631, 2000

AMAGAI, M.; STANLEY, J. R. Desmoglein as a Target in Skin Disease and Beyond. **Journal of Investigative Dermatology**, v. 132, n. 3, p. 776–784, 2011.

ANDRE, P. et al. New nomenclature for MHC receptors. **Nat Immunol,** v. 2, n. 8, p. 661, 2001.

AOKI, V. et al. Environmental risk factors in endemic pemphigus foliaceus (fogo selvagem). **The journal of investigative dermatology. Symposium proceedings**, v. 9, n. 1, p. 34–40, 2004.

AOKI, V. et al. Update on fogo selvagem, an endemic form of pemphigus foliaceus. **The Journal of dermatology**, v. 42, n. 1, p. 18–26, 2015.

ARNOLD M. et al. SNiPA: an interactive, genetic variant-centered annotation browser. **Bioinformatics**, v. 31, p. 1334-6, 2015. Disponível em: http://snipa.helmholtz-muenchen.de/ Acesso em janeiro de 2018.

ARTEAGA, L. A. et al. A Subset of Pemphigus Foliaceus Patients Exhibits Pathogenic Autoantibodies Against Both Desmoglein-1 and Desmoglein-3. **Journal of Investigative Dermatology**, v. 118, n. 5, p. 806–811, 2002.

AUGUSTO, D. G. et al. Activating KIR and HLA Bw4 Ligands Are Associated to Decreased Susceptibility to Pemphigus Foliaceus, an Autoimmune Blistering Skin Disease. **PLoS ONE**, v. 7, n. 7, p. e39991, 2012.

AUGUSTO, D. G.; PETZL-ERLER, M. L. KIR and HLA under pressure: evidences of coevolution across worldwide populations. **Hum Genet**, v. 134, n. 9, p. 929-40, 2015.

BARROW, A. D.; TROWSDALE, J. The extended human leukocyte receptor complex: diverse ways of modulating immune responses. **Immunol Rev**, n. 224, p. 98-123, 2008.

BASHIROVA, A. A. et al. Diversity of the human LILRB3/A6 locus encoding a myeloid inhibitory and activating receptor pair. **Immunogenetics**, v. 66, n. 1, p. 1-8, 2014.

BAUM, S. et al. Diagnosis and classification of autoimmune blistering diseases. **Autoimmunity Reviews**, v. 13, n. 4–5, p. 482–489, 2014.

BEKTAS, M. et al. A Pathophysiologic Role for Epidermal Growth Factor Receptor in Pemphigus Acantholysis. **Journal of Biological Chemistry**, v. 288, n. 13, p. 9447–9456, 2013.

BLEHARSKI, J. R. et al. Use of genetic profiling in leprosy to discriminate clinical forms of the disease. **Science**, v. 301, n. 5639, p 1527-30, 2003.

BOLGER, A. M.; LOHSE, M.; USADEL, B. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014.

BOYLE, A. P. et al. Annotation of functional variation in personal genomes using RegulomeDB. **Genome Research**, v. 22, n. 9, p. 1790–1797, 2012. Disponível em http://www.regulomedb.org/. Acesso em março de 2018.

BRESSAN, A. L. et al. [Immunosuppressive agents in Dermatology]. **Anais brasileiros de dermatologia**, v. 85, n. 1, p. 9–22, 2010.

BROCHADO, M. et al. Differential HLA class I and class II associations in pemphigus foliaceus and pemphigus vulgaris patients from a prevalent Southeastern Brazilian region. **Journal of Autoimmunity**, v. 72, p. 19–24, 2016.

BROOKE, M. A.; NITOIU, D.; KELSELL, D. P. Cell–cell connectivity: desmosomes and disease. **The Journal of Pathology**, v. 226, n. 2, p. 158–171, 2012.

BROWN, M. G. et al. The natural killer gene complex: a genetic basis for understanding natural killer cell function and innate immunity. **Immunological Reviews**, v. 155, n. 1, p. 53–65, 1997.

BYSTRYN, J.-C. C.; RUDOLPH, J. L. Pemphigus. **Lancet (London, England)**, v. 366, n. 9479, p. 61–73, 2005.

CAMARGO, C. M. Expressão gênica e polimorfismos de *CD33, LAIR1* e *LAIR2* em populações e em pênfigo foliáceo. 163 f. Dissertação (Mestrado em Genética) – Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, 2015.

CAMARGO, C.; AUGUSTO, D. G.; PETZL-ERLER, M. Differential gene expression levels might explain association of LAIR2 polymorphisms with pemphigus. **Human Genetics**, v. 135, n. 2, p. 233–244, 2016.

CAMPBELL, K. S.; PURDY, A. K. Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations. **Immunology**, v. 132, n. 3, p. 315-25, 2011.

CANELA, A. et al. Genome Organization Drives Chromosome Fragility. **Cell**, v. 170, n. 3, p. 507- 521.e18, 2017.

CARRINGTON M.; NORMAN, P. The KIR Gene Cluster [Internet]. Bethesda (MD): National Center for Biotechnology Information (US), 2003. Ebook. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK10135/. Acesso em setembro 2018.

CHAGAS, A. et al. Situação do pênfigo foliáceo endêmico em Mato Grosso do Sul, Brasil, 1990-1999. **Revista Latino-Americana de Enfermagem**, v. 13, n. 2, p. 274–276, 2005.

CHANG, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. **GigaScience**, v. 4, n. 1, p. 1–16, 2015.

CHEN, Y. et al. Crystal structure of myeloid cell activating receptor leukocyte Ig-like receptor A2 (LILRA2/ILT1/LIR-7) domain swapped dimer: molecular basis for its non-binding to MHC complexes. **J Mol Biol**, v. 386, n. 3, p. 841-53, 2009.

CHERNYAVSKY, A. I. et al. Desmoglein Versus Non-desmoglein Signaling in Pemphigus acantholysis characterization of novel signaling pathways downstream of pemphigus vulgaris antigens. **Journal of Biological Chemistry**, v. 282, n. 18, p. 13804–13812, 2007.

CIPOLLA, G. A. 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. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, v. 1859, n. 10, p. 1306–1313, 2016.

CONSORTIUM, T. et al. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. **Science**, v. 348, n. 6235, p. 648–660, 2015. Disponível em https://gtexportal.org/home/ Acessado em agosto de 2018.

CROCKER, P. R.; PAULSON, J. C.; VARKI, A. Siglecs and their roles in the immune system. **Nat Rev Immunol**, v. 7, n. 4, p. 255-66, 2007.

CULTON, D. A. et al. Advances in pemphigus and its endemic pemphigus foliaceus (Fogo Selvagem) phenotype: a paradigm of human autoimmunity. **Journal of autoimmunity**, v. 31, n. 4, p. 311–24, 2008.

DALLA-COSTA, R. et al. Polymorphisms in the 2q33 and 3q21 chromosome regions including T-cell coreceptor and ligand genes may influence susceptibility to pemphigus foliaceus. **Human Immunology**, v. 71, n. 8, p. 809–817, 2010.

DE STRONG, C. et al. A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. **Human Molecular Genetics**, v. 19, n. 8, p. 1453–1460, 2010.

DELVA, E.; TUCKER, D. K.; KOWALCZYK, A. P. The Desmosome. **Cold Spring Harbor Perspectives in Biology**, v. 1, n. 2, p. a002543, 2009.

DIAZ, L. A. et al. Anti-desmoglein-1 antibodies in onchocerciasis, leishmaniasis and Chagas disease suggest a possible etiological link to Fogo selvagem. **The Journal of investigative dermatology**, v. 123, n. 6, p. 1045–51, 2004.

DIAZ, L. et al. Endemic pemphigus foliaceus (Fogo Selvagem): II. Current and historic epidemiologic studies. **The Journal of investigative dermatology**, v. 92, n. 1, p. 4–12, 1989.

DO CHIOSSI, M.; ROSELINO, A. F. Endemic Pemphigus foliaceus ("Fogo selvagem"): a series from the Northeastern region of the State of São Paulo, Brazil, 1973-1998. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 43, n. 2, p. 59–62, 2001.

ELLEBRECHT, C. T.; PAYNE, A. S. Plakophilins, Desmogleins, and Pemphigus: The Tail Wagging the Dog. **Journal of Investigative Dermatology**, v. 134, n. 4, p. 874–876, 2014.

FILIPPOVA, G. et al. An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. **Molecular and Cellular Biology**, v. 16, n. 6, p. 2802–2813, 1996.

FLORES, G. et al. IgG autoantibody response against keratinocyte cadherins in endemic pemphigus foliaceus (fogo selvagem). **The Journal of investigative dermatology**, v. 132, n. 11, p. 2573–80, 2012.

FRUŠIĆ-ZLOTKIN, M. et al. Apoptotic mechanism in pemphigus autoimmunoglobulins-induced acantholysis—possible involvement of the EGF receptor. **Autoimmunity**, v. 39, n. 7, p. 563–575, 2009.

FU, Y. et al. The Insulator Binding Protein CTCF Positions 20 Nucleosomes around Its Binding Sites across the Human Genome. **PLoS Genetics**, v. 4, n. 7, p. e1000138, 2008.

GRANDO, S. A. et al. Apoptolysis: a novel mechanism of skin blistering in pemphigus vulgaris linking the apoptotic pathways to basal cell shrinkage and suprabasal acantholysis. **Experimental Dermatology**, v. 18, n. 9, p. 764–770, 2009.

GRANDO, S. A. Pemphigus autoimmunity: Hypotheses and realities. **Autoimmunity**, v. 45, n. 1, p. 7–35, 2011.

GRANDO, S. et al. The effect of experimental haemocarbofiltration upon activity of mononuclear cells from normal and autoimmune patients. **Immunology**, v. 66, n. 1, p. 138–42, 1989.

GRIMWOOD, J. et al. The DNA sequence and biology of human chromosome 19. **Nature**, v. 428, n. 6982, p. nature02399, 2004.

GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science, v. 348, n. 6235, p. 648-60, 2015.

HANS-FILHO, G. et al. An active focus of high prevalence of fogo selvagem on an Amerindian reservation in Brazil. Cooperative Group on Fogo Selvagem Research. **The Journal of investigative dermatology**, v. 107, n. 1, p. 68–75, 1996.

HANS-FILHO, G. et al. Endemic pemphigus foliaceus (fogo selvagem) -1998. The Cooperative Group on Fogo Selvagem Research. **Clinics in dermatology**, v. 17, n. 2, p. 225–35; discussion 105-6, 1999.

HEBERT, V.; JOLY, P. Rituximab in pemphigus. **Immunotherapy**, v. 10, n. 1, p. 27–37, 2017.

HIRAYASU, K. et al. Long-term persistence of both functional and non-functional alleles at the leukocyte immunoglobulin-like receptor A3 (LILRA3) locus suggests balancing selection. Hum **Genet**, v. 119, n. 4, p. 436-43, 2006.

HSU, K. C. et al. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. **Immunol Rev**, v. 190, p. 40-52, 2002.

IPD – Immuno Polymmorphism Database. Disponível em http://www.ebi.ac.uk/ipd/kir/introduction.html. Acesso em setembro 2018.

ISHII, K. et al. Characterization of autoantibodies in pemphigus using antigen-specific enzyme-linked immunosorbent assays with baculovirus-expressed

recombinant desmogleins. **Journal of immunology (Baltimore, Md. : 1950)**, v. 159, n. 4, p. 2010–7, 1997.

JANSE, I. C. et al. No Evidence of Apoptotic Cells in Pemphigus Acantholysis. **Journal of Investigative Dermatology**, v. 134, n. 7, p. 2039–2041, 2014.

JOLY, P.; LITROWSKI, N. Pemphigus group (vulgaris, vegetans, foliaceus, herpetiformis, brasiliensis). **Clinics in Dermatology**, v. 29, n. 4, p. 432–436, 2011.

JONES, D. C. et al. Alternative mRNA splicing creates transcripts encoding soluble proteins from most LILR genes. **Eur J Immunol**, v. 39, n. 11, p. 3195-206, 2009.

JU, X. S. et al. Immunoglobulin-like transcripts ILT2, ILT3 and ILT7 are expressed by human dendritic cells and down-regulated following activation. **Gene**, v. 331, p. 159-64, 2004.

KELLEY, J.; WALTER, L.; TROWSDALE, J. Comparative Genomics of Natural Killer Cell Receptor Gene Clusters. **PLoS Genetics**, v. 1, n. 2, p. e27, 2005.

KIKUCHI-MAKI, A. et al. KIR2DL4 is an IL-2-regulated NK cell receptor that exhibits limited expression in humans but triggers strong IFN-gamma production. **J Immunol**, v. 171, n. 7, p. 3415-25, 2003.

KOBOLDT, D. C. et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. **Bioinformatics**, v. 25, n. 17, p. 2283–2285, 2009.

KULKARNI, S.; MARTIN, M. P.; CARRINGTON, M. The Yin and Yang of HLA and KIR in human disease. **Semin Immunol**, v. 20, n. 6, p. 343-52, 2008.

LEBBINK, R. J. et al. Collagens are functional, high affinity ligands for the inhibitory immune receptor LAIR-1. J Exp Med, v. 203, n. 6, p. 1419-25, 2006.

LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows–Wheeler transform. **Bioinformatics**, v. 25, n. 14, p. 1754–1760, 2009.

LI, N. et al. The role of intramolecular epitope spreading in the pathogenesis of endemic pemphigus foliaceus (fogo selvagem). **The Journal of experimental medicine**, v. 197, n. 11, p. 1501–10, 2003.

LIN, M.-S. et al. Desmoglein-1–specific T lymphocytes from patients with endemic pemphigus foliaceus (fogo selvagem). **Journal of Clinical Investigation**, v. 105, n. 2, p. 207–213, 2000.

LOMBARDI, C. et al. Environmental risk factors in endemic pemphigus foliaceus (Fogo selvagem). "The Cooperative Group on Fogo Selvagem Research". **The Journal of investigative dermatology**, v. 98, n. 6, p. 847–50, 1992.

LONG, E. O.; COLONNA, M.; LANIER, L. L. Inhibitory MHC class I receptors on NK and T cells: a standard nomenclature. **Immunol Today**, v. 17, n. 2, p. 100, 1996.

LÓPEZ-ÁLVAREZ, M.R. et al. Copy number and nucleotide variation of the LILR family of myelomonocytic cell activating and inhibitory receptors. **Immunogenetics**, v. 66, n. 2, p. 73-83, 2014.

MACHIELA, M. J.; CHANOCK, S. J. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. **Bioinformatics**, v. 31, n. 21, p. 3555–3557, 2015.

MAHONEY, M. G. et al. Explanations for the clinical and microscopic localization of lesions in pemphigus foliaceus and vulgaris. **Journal of Clinical Investigation**, v. 103, n. 4, p. 461–468, 1999.

MALHEIROS, D. et al. Genome-wide gene expression profiling reveals unsuspected molecular alterations in pemphigus foliaceus. **Immunology**, v. 143, n. 3, p. 381–395, 2014.

MALHEIROS, D.; PETZL-ERLER, M. Individual and epistatic effects of genetic polymorphisms of B-cell co-stimulatory molecules on susceptibility to pemphigus foliaceus. **Genes and Immunity**, v. 10, n. 6, p. 547, 2009.

McMILLAN, S. J.; CROCKER, P. R. CD33-related sialic-acid- binding immunoglobulin-like lectins in health and disease. **Carbohydr Res**, v. 343, n. 12, p. 2050-6, 2008.

MEYAARD, L. et al Leukocyte-associated Ig-like receptor-1 functions as an inhibitory receptor on cytotoxic T cells. **J Immunol**, v. 162, n. 10, p. 5800-4, 1999.

MEYAARD, L. et al. LAIR-1, a novel inhibitory receptor expressed on human mononuclear leukocytes. **Immunity**, v. 7, n. 2, p. 283-90, 1997.

MEYAARD, L. The inhibitory collagen receptor LAIR-1 (CD305). **J Leukoc Biol**, v. 83, n. 4, p. 799-803, 2008.

MIDDLETON, D.; GONZELEZ, F. The extensive polymorphism of KIR genes. **Immunology**, v. 129, n. 1, p. 8-19, 2010.

MORINI, J. et al. Pemphigus foliaceus in young women. An endemic focus in the Sousse area of Tunisia. **Archives of dermatology**, v. 129, n. 1, p. 69–73, 1993.

MORIYAMA, M. et al. Multiple Roles of Notch Signaling in the Regulation of Epidermal Development. **Developmental Cell**, v. 14, n. 4, p. 594–604, 2008.

MORTAZAVI, H. et al. Frequency of seropositivity against infectious agents amongst pemphigus vulgaris patients: a case-control study on Strongyloides

stercoralis, Helicobacter pylori, Toxoplasma gondii, Leishmania major, and Epstein-Barr virus. **International journal of dermatology**, v. 54, n. 11, p. e458-65, 2015.

MUSETTE, P.; BOUAZIZ, J. B Cell Modulation Strategies in Autoimmune Diseases: New Concepts. **Frontiers in Immunology**, v. 9, p. 622, 2018.

NG, S. B. et al. Targeted capture and massively parallel sequencing of 12 human exomes. **Nature**, v. 461, n. 7261, p. 272, 2009.

OHATA, C.; ISHII, N.; FURUMURA, M. Locations of acantholysis in pemphigus vulgaris and pemphigus foliaceus. **Journal of Cutaneous Pathology**, v. 41, n. 11, p. 880–889, 2014.

OHLSSON, R.; RENKAWITZ, R.; LOBANENKOV, V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. **Trends in Genetics**, v. 17, n. 9, p. 520–527, 2001.

ORTEGA-LOAYZA, A. et al. Endemic pemphigus foliaceus in the Peruvian Amazon. **Clinical and Experimental Dermatology**, v. 38, n. 6, p. 594–600, 2013.

PAN, M.; ZHU, H.; XU, R. Immune cellular regulation on autoantibody production in pemphigus. **The Journal of Dermatology**, v. 42, n. 1, p. 11–17, 2015.

PARHAM, P. Taking license with natural killer cell maturation and repertoire development. **Immunol Rev**. v. 214, p. 155-60, 2006.

PAVONI, D. et al. Dissecting the associations of endemic Pemphigus Foliaceus (Fogo Selvagem) with HLA-DRB1 alleles and genotypes. **Genes and Immunity**, v. 4, n. 2, p. 110, 2003.

PEREIRA, N. et al. Cytokine gene polymorphisms in endemic pemphigus foliaceus: a possible role for IL6 variants. **Cytokine**, v. 28, n. 6, p. 233–241, 2004.

PERRICONE, R. et al. NK cells in autoimmunity: A two-edged weapon of the immune system. **Autoimmunity Reviews**, v. 7, n. 5, p. 384–390, 2008.

PETZL-ERLER, M.; SANTAMARIA, J. Are HLA class II genes controlling susceptibility and resistance to Brazilian pemphigus foliaceus (fogo selvagem)? **Tissue Antigens**, v. 33, n. 3, p. 408–414, 1989.

POTERLOWICZ, K. et al. 5C analysis of the Epidermal Differentiation Complex locus reveals distinct chromatin interaction networks between gene-rich and gene-poor TADs in skin epithelial cells. **PLOS Genetics**, v. 13, n. 9, p. e1006966, 2017.

PYO, CW. et al. Different patterns of evolution in the centromeric and telomeric regions of group A and B haplotypes of the human killer cell Ig-like receptor locus. **PLoS One**, v. 5, n. 12, p. e15115, 2010.

QIAN, Y. et al. Antigen selection of anti-DSG1 autoantibodies during and before the onset of endemic pemphigus foliaceus. **Journal of Investigative Dermatology**, v. 129, n. 12, p. 2823–2834, 2009.

RAULET, D.H. Missing self recognition and self tolerance of natural killer (NK) cells. **Semin Immunol**, v. 18, n. 3, p. 145-50, 2006.

RUOCCO, E. et al. Pemphigus: Associations and management guidelines: Facts and controversies. **Clinics in Dermatology**, v. 31, n. 4, p. 382–390, 2013.

RUOCCO, E. et al. Viruses and Pemphigus: an intriguing never-ending story. **Dermatology**, v. 0, n. 0, p. 310–315, 2014.

SAGI, L. et al. Autoimmune bullous diseases: The spectrum of infectious agent antibodies and review of the literature. **Autoimmunity Reviews**, v. 10, n. 9, p. 527–535, 2011.

SAMARKOS, M.; VAIOPOULOS, G. The role of infections in the pathogenesis of autoimmune diseases. **Current drug targets. Inflammation and allergy**, v. 4, n. 1, p. 99–103, 2005.

SANCHEZ-MAZAS, A. et al. Immunogenetics as a tool in anthropological studies. **Immunology**, v. 133, n. 2, p. 143-64, 2011.

SCHLEINITZ, N. et al. Natural killer cells in human autoimmune diseases. **Immunology**, v. 131, n. 4, p. 451–458, 2010.

SCHMIDT, E.; WASCHKE, J. Apoptosis in pemphigus. **Autoimmunity Reviews**, v. 8, n. 7, p. 533–537, 2009.

SEGRE, J. A. Epidermal barrier formation and recovery in skin disorders. **Journal of Clinical Investigation**, v. 116, n. 5, p. 1150–1158, 2006.

SELLAMI, M. et al. Anti-desmoglein 1 antibodies in Tunisian healthy subjects: arguments for the role of environmental factors in the occurrence of Tunisian pemphigus foliaceus. **Clinical & Experimental Immunology**, v. 137, n. 1, p. 195–200, 2004.

SHARMA, P.; MAO, X.; PAYNE, A. S. Beyond steric hindrance: The role of adhesion signaling pathways in the pathogenesis of pemphigus. **Journal of Dermatological Science**, v. 48, n. 1, p. 1–14, 2007.

SINGH, K. M. et al. KIR genotypic diversity can track ancestries in heterogeneous populations: a potential confounder for disease association studies. **Immunogenetics**, v. 64, n. 2, p. 97-109, 2012.

SPINDLER, V. et al. Peptide-mediated desmoglein 3 crosslinking prevents pemphigus vulgaris autoantibody-induced skin blistering. **Journal of Clinical Investigation**, v. 123, n. 2, p. 800–811, 2013.

STERN, J. et al. Possible role of natural killer cells in pemphigus vulgaris − preliminary observations. **Clinical & Experimental Immunology**, v. 152, n. 3, p. 472–481, 2008.

TAKAHASHI, H. et al. T helper type 2-biased natural killer cell phenotype in patients with pemphigus vulgaris. **Journal of Investigative Dermatology**, v. 127, n. 2, p. 324–330, 2006.

THUL, P. J. et al. A subcellular map of the human proteome. **Science**, v. 356, n. 6340, p. pii: eaal3321, 2017.

TOMASELLO, E.; VIVIER, E. KARAP/DAP12/TYROBP: three names and a multiplicity of biological functions. **Eur J Immunol**, v. 35, n. 6, p. 1670-7, 2005.

TROWSDALE, J. et al. Surveillance of cell and tissue perturbation by receptors in the LRC. **Immunol Rev,** v. 267, n. 1, p. 117-36, 2015.

UHLÉN, M. et al. Tissue-based map of the human proteome. **Science**, v. 347, n. 6220, p. 1260419, 2015.

UHRBERG, M. Shaping the human NK cell repertoire: an epigenetic glance at KIR gene regulation. **Mol Immunol**, v. 42, n. 4, p. 471-5, 2005.

VIELMUTH, F.; WASCHKE, J.; SPINDLER, V. Loss of desmoglein binding is not sufficient for keratinocyte dissociation in pemphigus. **Journal of Investigative Dermatology**, v. 135, n. 12, p. 3068–3077, 2015.

VILCHES, C.; PARHAM, P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. **Annu Rev Immunol**, v. 20, p. 217-51, 2002.

VUKCEVIC, D. Imputation of KIR Types from SNP Variation Data. **Am J Hum Genet**, v. 97, n. 4, p. 593-607, 2015.

WARD, L. D.; KELLIS, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. **Nucleic Acids Research**, v. 40, n. D1, p. D930–D934, 2012. Disponível em < https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php> Acesso em agosto de 2018.

WARDE-FARLEY, D, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. **Nucleic Acids Res,** v. 38, Web Server issue: W214-20, 2010.

WARREN, S. et al. The role of subclass switching in the pathogenesis of endemic pemphigus foliaceus. **Journal of Investigative Dermatology**, v. 120, n. 1, p. 1–5, 2003.

WASCHKE, J. et al. Pemphigus foliaceus IgG causes dissociation of desmoglein 1–containing junctions without blocking desmoglein 1 transinteraction. **Journal of Clinical Investigation**, v. 115, n. 11, p. 3157–3165, 2005.

WENDE, H. et al. Organization of the leukocyte receptor cluster (LRC) on human chromosome 19q13.4. **Mamm Genome**, v. 10, n. 2, p. 154–60, 1999.

WENDE, H.; VOLZ, A.; ZIEGLER, A. Extensive gene duplications and a large inversion characterize the human leukocyte receptor cluster. **Immunogenetics**, v. 51, n. 8-9, p. 703-13, 2000.

WU, J. et al. An activating immunoreceptor complex formed by NKG2D and DAP10. **Science**, v. 285, n. 5428, p.730-2, 1999.

YOKOYAMA, W. M.; PLOUGASTEL, B. F. Immune functions encoded by the natural killer gene complex. **Nature Reviews Immunology**, v. 3, n. 4, p. nri1055, 2003.

ZAKKA, L. et al. The role of natural killer cells in autoimmune blistering diseases. **Autoimmunity**, v. 45, n. 1, p. 44–54, 2011.

APÊNDICE A – MATERIAL SUPLEMENTAR DO CAPÍTULO I

TABLE S1: ANNOTATION OF THE SNPS SIGNIFICANTLY ASSOCIATED WITH DIFFERENTIAL SUSCEPTIBILITY TO *FOGO SELVAGEM*, ACCORDING TO ONLINE DATABASES

	Annotation	SNP position	GeneCards ID	Ensembl ID	Aliases
rs8103298					
rs3810265					
	RPS9P4	non-coding transcript exon variant	RPS9P4	ENSG00000243829	AC011495.1 TCONS_l2_00013175 CTB-33G10.1
rs9304655					
rs6509255					
	IGFL2	upstream	IGFL2	ENSG00000204866	
	lnc-IGFL2-1	intron variant	ENSG00000267922	ENSG00000267922	AC007785.1
rs10412596					
	IGFL2-AS1	downstream	IGFL2-AS1	ENSG00000268621	
	lnc-IGFL2-1	intron variant	ENSG00000267922	ENSG00000267922	AC007785.1 TCONS_00027056
rs10439109					
	lnc-IGFL2-1	intron variant	ENSG00000267922	ENSG00000267922	AC007785.1
	lnc-IGFL3-1	downstream	LOC105372424	ENSG00000269729	AC006262.2
rs1654431					
	RDH13	intron variant	RDH13	ENSG00000160439	
	GP6	upstream variant	GP6	ENSG00000088053	
	lnc-RDH13-5	intron variant	ENSG00000267149	ENSG00000267149	AC011476.2
	lnc-NLRP2-2	intron variant	ENSG00000267265	ENSG00000267265	AC011476.3

Long non-coding (lnc) RNA genes annotation was obtained from LNCipedia VOLDERS et al., 2015, GeneCards (BELINKY et al., 2013), Ensembl (AKEN et al., 2016) gene annotation databases.

TABLE S2: ALLELE AND GENOTYPE FREQUENCIES OF SNPA ASSOCIATED WITH *FOGO SELVAGEM* AND SPORADIC PEMPHIGUS FOLIACEUS AMONG PATIENTS AND CONTROLS. (continua)

		Brazil		Germany	
		FS (F%)	CT (F%)	PF (F%)	CT (F%)
rs8103298					
	G	45 (20.5)	18 (9.3)	NC	NC
	GG	7 (3.2)	2 (1.0)	NC	NC
	GA	76 (34.5)	32 (16.6)	NC	NC
	AA	137 (62.3)	159 (82.4)	NC	NC
p (HWE)		0.533	0.671	NC	NC
rs3810265					
	A	34 (15.2)	14 (7.0)	1 (0.7)	0 (0.0)
	AA	3 (1.4)	1 (0.5)	0 (0.0)	0 (0.0)
	AG	61 (27.7)	25 (13.0)	1 (1.4)	0 (0.0)
	GG	156 (70.9)	167 (86.5)	69 (98.6)	140 (100.0)
p (HWE)		0.431	1.000	NC	NC
rs9304655					
	A	105 (47.3)	18 (9.3)	29 (30.2)	83 (31.9)
	AA	49 (22.1)	2 (1.0)	3 (6.3)	13 (10.0)
	AG	112 (50.5)	32 (16.6)	23 (47.9)	57 (43.8)
	GG	61 (27.5)	159 (82.4)	22 (45.8)	60 (46.2)
p (HWE)		0.894	0.525	0.499	1.000
rs6509255					
	C	103 (46.2)	66 (34.1)	NC	NC
	CC	45 (20.3)	24 (12.5)	NC	NC
	CT	115 (51.8)	83 (43.2)	NC	NC
	TT	62 (27.9)	85 (44.3)	NC	NC
p (HWE)		0.590	0.630	NC	NC
rs10412596					
	T	71 (32.2)	41 (21.0)	20 (14.3)	47 (17.0)
	TT	22 (10.0)	9 (4.7)	0 (0.0)	5 (3.6)
	TC	97 (44.3)	63 (32.6)	20 (28.6)	37 (26.8)
	CC	100 (45.7)	121 (62.7)	50 (71.4)	96 (69.6)
p (HWE)		0.878	0.829	0.337	0.547
rs10439109					
	A	121 (54.3)	85 (43.8)	60 (45.5)	106 (39.0)
	AA	60 (26.9)	40 (20.7)	12 (18.2)	19 (14.0)
	AG	122 (54.7)	89 (46.1)	36 (54.5)	68 (50.0)
	GG	41 (18.4)	64 (33.2)	18 (27.3)	49 (36.0)
p (HWE)		0.140	0.382	0.468	0.594
rs11083944					
	G	76 (34.2)	91 (46.9)	45 (47.9)	107 (47.3)
	GG	31 (14.0)	42 (21.8)	8 (17.0)	33 (29.2)
	GA	89 (40.3)	97 (50.3)	29 (61.7)	41 (36.3)
	AA	101 (45.7)	54 (28.0)	10 (21.3)	39 (34.5)
p (HWE)		0.134	1.000	0.150	0.004
rs2304208					
	C	24 (10.5)	8 (3.9)	1 (1.0)	0 (0.0)
	CC	3 (1.3)	1 (0.5)	0 (0.0)	0 (0.0)
	CT	41 (18.5)	13 (6.7)	1 (2.0)	0 (0.0)
	TT	179 (80.3)	179 (92.7)	50 (98.0)	93 (100.0)
p (HWE)		0.718	0.246	NC	NC

TABLE S2: ALLELE AND GENOTYPE FREQUENCIES OF SNPA ASSOCIATED WITH *FOGO SELVAGEM* AND SPORADIC PEMPHIGUS FOLIACEUS AMONG PATIENTS AND CONTROLS. (conclusão)

		Brazil		**Germany**	
		FS (F%)	CT (F%)	PF (F%)	CT (F%)
rs1654431					
	A	111 (50.9)	81 (42.6)	52 (50.0)	114 (49.6)
	AA	63 (28.9)	27 (14.2)	12 (23.1)	32 (27.8)
	AG	96 (44.0)	108 (56.8)	28 (53.8)	50 (43.5)
	GG	59 (27.1)	55 (28.9)	12 (23.1)	33 (28.7)
p (HWE)		0.079	0.037	0.781	0.191

SNP: single nucleotide polymorphism, FS: *fogo selvagem*, PF: pemphigus foliaceus, CT: control, *p*: probability value, F: relative frequency, HWE: Hardy-Weinberg equilibrium. NC: not calculated due to no genotype data available. In bold *p*-value < 0.05.

TABLE S3: MOTIF FEATURE CONSEQUENCES OF *rs3810265*, *rs8103298*, *rs11083944*, *rs1654431* SNPS ON TRANSCRIPTION FACTOR BINDING SITE WITHIN LRC EXTENDED REGION, AT 19q13.

SNP	Alelle	Transcription factors	Motif position	Motif score change	Binding matrix
RPS9P4					
rs3810265	A	SMAD3	4 (out of 10)	Less like consensus sequence	ENSPFM0489
	A	TEAD4::ONECUT2	10 (out of 16)	More like consensus sequence	ENSPFM0567
	A	MAX, TFAP4::MAX, CLOCK::BHLHA15	6 (out of 17)	More like consensus sequence	ENSPFM0365
rs8103298	C	PAX6	14 (out of 19)	Less like consensus sequence	ENSPFM0435
	C	NKX3-2, NKX2-3, NKX2-8, NKX3-1	4 (out of 10)	Less like consensus sequence	ENSPFM0416
	C	TEAD4::CLOCK	4 (out of 19)	Less like consensus sequence	ENSPFM0524
	C	TEAD4::HOXA2, TEAD4::HOXA3	4 (out of 17)	Less like consensus sequence	ENSPFM0560
	C	ZSCAN4	4 (out of 15)	Less like consensus sequence	ENSPFM0632
	C	GCM1::PITX1	16 (out of 21)	Less like consensus sequence	ENSPFM0244
	C	GCM2::HES7	6 (out of 20)	Less like consensus sequence	ENSPFM0259
	C	TFAP2C::DLX3	9 (out of 22)	More like consensus sequence	ENSPFM0587
	C	TFAP2C::ONECUT2	17 (out of 26)	More like consensus sequence	ENSPFM0601
	C	RXRG, RXRA, RORB	4 (out of 14)	No change	ENSPFM0487
SULT2B1					
rs11083944	G	TEAD4::CLOCK, TEAD4::HES7	5 (out of 20)	Less like consensus sequence	ENSPFM0525
	G	TEAD4::EOMES	10 (out of 22)	Less like consensus sequence	ENSPFM0537
RDH13					
rs1654431	A	CUX1::NHLH1	4 (out of 23)	More like consensus sequence	ENSPFM0049
	A	ELK1::FOXI1	4 (out of 16)	Less like consensus sequence	ENSPFM0096
	A	ETV2::PAX5, ELK1::PAX1, ELK1::PAX5, ELK1::PAX9	4 (out of 20)	Less like consensus sequence	ENSPFM0104
	A	ETV2::SOX15	4 (out of 21)	Less like consensus sequence	ENSPFM0166
	A	ETV2::TEF	17 (out of 23)	Less like consensus sequence	ENSPFM0170

Transcription factor binding site feature promoted by *rs3810265*, *rs8103298*, *rs11083944*, *rs1654431* SNPs (*single nucleotide polymorphism*) associated with FS. Data was obtained on Ensembl genome browser.

TABLE S4: POTENTIAL EFFECT ON GENE EXPRESSION LEVELS BY *rs3810265*, *rs8103298*, *rs11083944*, *rs1654431* SNPS WITHIN LRC EXTENDED REGION, AT 19q13.

SNP	Alelle	Gene	Ensembl_ID	Gene name	Regulation	Tissue
RPS9P4						
rs3810265	A	PTOV1	ENSG00000104960	prostate tumor overexpressed 1	down	Whole blood
		PTOV1-AS2	ENSG00000269352	PTOV1 antisense RNA 2	down	Whole blood
		SNORD32A	ENSG00000201675	small nucleolar RNA, C/D box 32A	down	Esophagus mucosa
		EMC10	ENSG00000161671	ER membrane protein complex subunit 10	up	Skin not sun exposed
		lnc-FTL-1	ENSG00000267898	AC026803.2	up	Whole blood
rs8103298	C	ZNF473	ENSG00000142528	zinc finger protein 473	down	Skin not sun exposed
		SLC6A16	ENSG00000063127	solute carrier family 6 member 16	up	Whole blood
		lnc-FTL-1	ENSG00000267898	AC026803.2	up	Whole blood
SULT2B1						
rs11083944	G	SNORD35B	ENSG00000200530	small nucleolar RNA, C/D box 35B	down	Whole blood
		SNRNP70	ENSG00000104852	small nuclear ribonucleoprotein U1 subunit 70	down	Skin sun exposed
		LHB	ENSG00000104826	luteinizing hormone beta polypeptide	down	Esophagus mucosa
RDH13						
rs1654431	A	RDH13	ENSG00000160439	retinol dehydrogenase 13	up	Whole blood
		lnc-RDH13-12	ENSG00000267149	AC011476.2	up	Whole blood
		lnc-NLRP2-2	ENSG00000267265	AC011476.3	up	Whole blood
		NLRP2	ENSG00000022556	NLR family pyrin domain containing 2	up	Whole blood
		NLRP2	ENSG00000022556	NLR family pyrin domain containing 2	up	Esophagus mucosa
		lnc-MBOAT7-1	ENSG00000223660	AC245052.1	up	Whole blood
		lnc-RDH13-12	ENSG00000267149	AC011476.2	down	Skin sun exposed and not exposed
		lnc-RDH13-12	ENSG00000267149	AC011476.2	down	Esophagus mucosa
		RDH13	ENSG00000160439	retinol dehydrogenase 13	down	Skin sun exposed

Regulatory effect of *rs3810265*, *rs8103298*, *rs11083944*, *rs1654431* SNPs (*single nucleotide polymorphism*) associated with FS. Data was obtained on Ensembl genome browser, considering $p < 0.01$.

TABLE S5: POTENTIAL EFFECT ON GENE EXPRESSION LEVELS BY *rs9304655, rs6509255, rs10412596* AND *rs10439109* SNPS OF *lnc-IGFL2* GENE WITHIN LRC EXTENDED REGION, AT 19q13.

SNP	Alelle	Gene	Ensembl_ID	Gene name	Regulation	Tissue
rs9304655 *rs6509255*	A C					
		lnc-IGFL1-1	ENSG00000268460	AC006262.1	up	Skin sun exposed and not exposed
		lnc-IGFL2-1	ENSG00000267922	AC007785.1	up	Skin sun exposed
		lnc-IGFL3-1	ENSG00000269729	AC006262.2	up	Skin sun exposed
		IGFL2-AS1	ENSG00000268621	IGFL2 antisense RNA	up	Skin sun exposed
		IGFL4	ENSG00000204869	IGF like family member 4	down	Skin sun exposed
		DM1-AS	ENSG00000267395	DM1 locus antisense RNA	down	Skin sun exposed
		ZC3H4	ENSG00000130749	zinc finger CCCH-type containing 4	down	Skin sun exposed
		DMPK	ENSG00000104936	DM1 protein kinase	down	Skin sun exposed
		RPL12P41	ENSG00000269637	ribosomal protein L12 pseudogene 41	down	Skin sun exposed
rs10412596	T	lnc-IGFL1-1	ENSG00000268460	AC006262.1	up	Skin sun exposed and not exposed
		IGFL1	ENSG00000188293	IGF like family member 1	up	Skin not sun exposed
		IGFL2	ENSG00000204866	IGF like family member 2	up	Skin sun exposed
		IGFL3	ENSG00000188624	IGF like family member 3	up	Skin sun exposed
		PPM1N	ENSG00000213889	protein phosphatase, Mg2+/Mn2+ dependent 1N	down	Whole blood
		PPP5C	ENSG00000011485	protein phosphatase 5 catalytic subunit	down	Esophagus mucosa
rs10439109	A	lnc-IGFL1-1	ENSG00000268460	AC006262.1	up	Skin sun exposed and not exposed
		lnc-IGFL1-1	ENSG00000268460	AC006262.1	up	Esophagus mucosa
		lnc-IGFL2-1	ENSG00000267922	AC007785.1	up	Skin not sun exposed
		IGFL2	ENSG00000204866	IGF like family member 2	up	Skin sun exposed
		RPL12P41	ENSG00000269637	ribosomal protein L12 pseudogene 41	down	Skin sun exposed
		PRKD2	ENSG00000105287	protein kinase D2	down	Whole blood

Regulatory effect of *rs9304655, rs6509255, rs10412596 and rs10439109* SNPs (*single nucleotide polymorphism*) associated with FS. Data was obtained on Ensembl genome browser, considering $p < 0.01$.

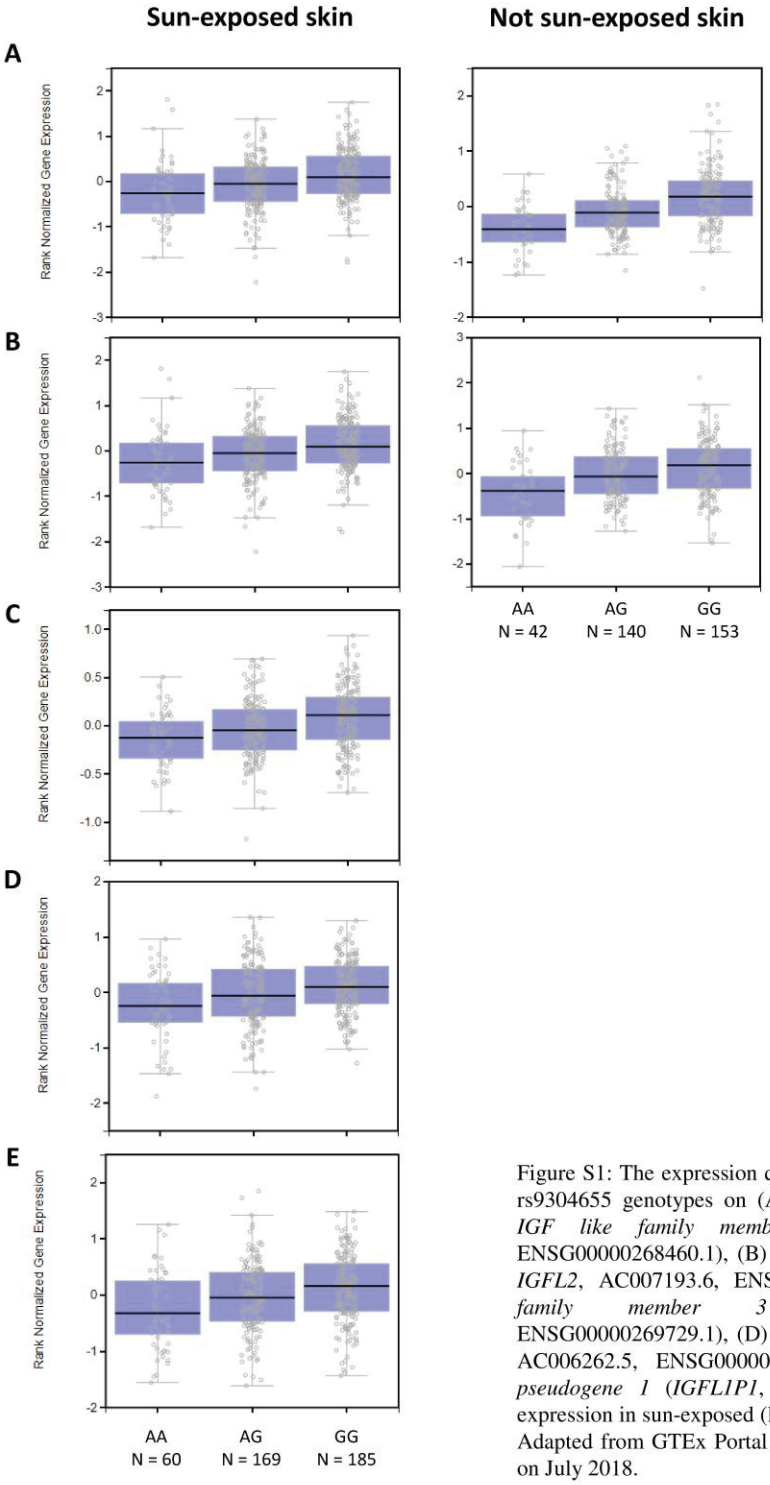


Figure S1: The expression quantitative trait (eQTL) effect of the rs9304655 genotypes on (A) *long non-coding RNA (lncRNA)-IGF like family member 1* (*lnc-IGFL1, AC006262.1,* ENSG00000268460.1), (B) *lnc-IGF like family member 2* (*lnc-IGFL2*, AC007193.6, ENSG00000267922), (C) *lnc-IGF like family member 3* (*lnc-IGFL3*, AC006262.4, ENSG00000269729.1), (D) antisense-RNA *IGFL2* (*IGFL2-AS1*, AC006262.5, ENSG00000268621.1) genes and (E) *IGFL1 pseudogene 1* (*IGFL1P1*, ENSG00000268879.1) pseudogene expression in sun-exposed (left) and not sun-exposed skin (right). Adapted from GTEx Portal (Consortium *et al*., 2015). Accessed on July 2018.

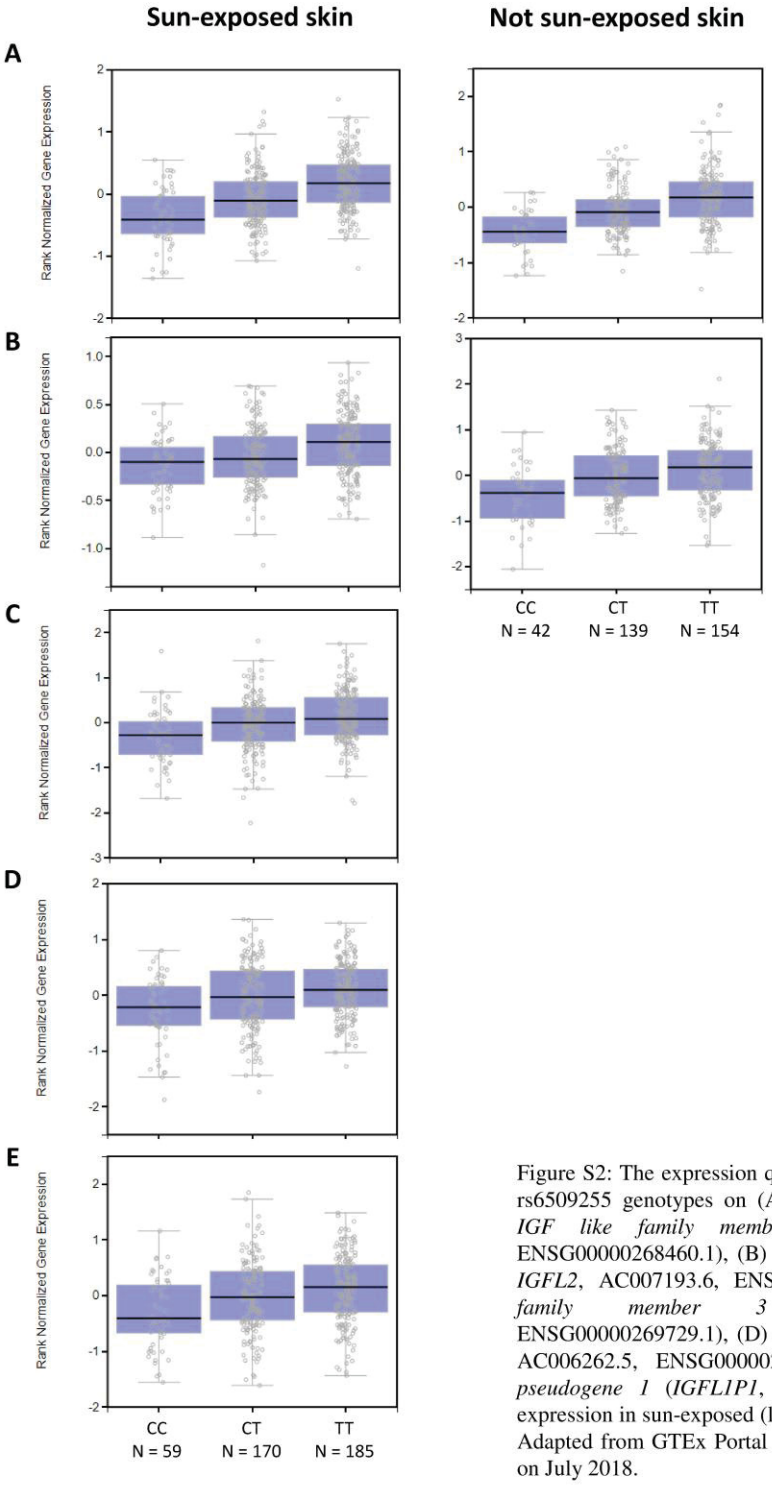


Figure S2: The expression quantitative trait (eQTL) effect of the rs6509255 genotypes on (A) *long non-coding RNA (lncRNA)-IGF like family member 1* (*lnc-IGFL1, AC006262.1,* ENSG00000268460.1), (B) *lnc-IGF like family member 2* (*lnc-IGFL2*, AC007193.6, ENSG00000267922), (C) *lnc-IGF like family member 3* (*lnc-IGFL3*, AC006262.4, ENSG00000269729.1), (D) antisense-RNA *IGFL2* (*IGFL2-AS1*, AC006262.5, ENSG00000268621.1) genes and (E) *IGFL1 pseudogene 1* (*IGFL1P1*, ENSG00000268879.1) pseudogene expression in sun-exposed (left) and not sun-exposed skin (right). Adapted from GTEx Portal (Consortium *et al*., 2015). Accessed on July 2018.

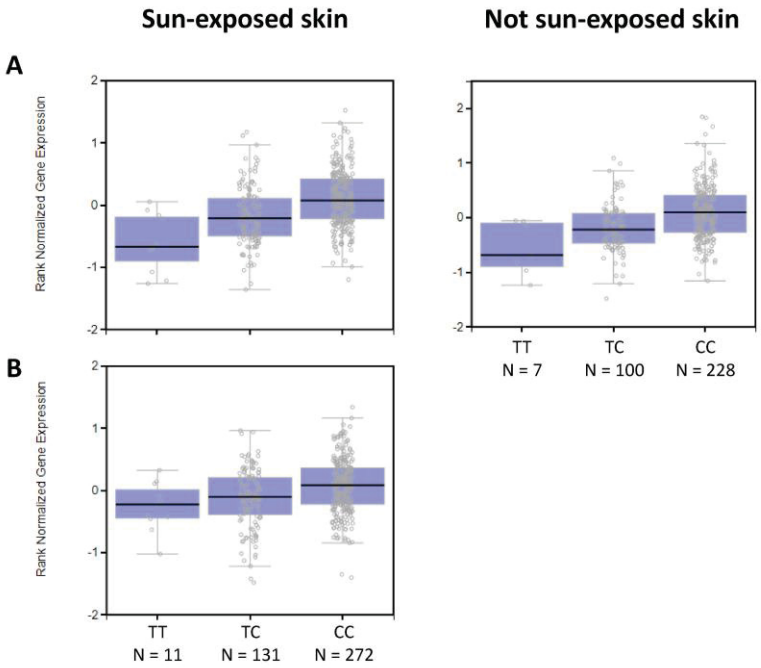


Figure S3: The expression quantitative trait (eQTL) effect of the *rs10412596* genotypes on (A) *long non-coding RNA (lncRNA)-IGF like family member 1* (*lnc-IGFL1, AC006262.1,* ENSG00000268460.1) and (B) *IGF like family member 3* (*IGFL3*, ENSG00000188624.2) genes expression in sun-exposed (left) and not sun-exposed skin (right). Adapted from GTEx Portal (Consortium *et al*., 2015). Accessed on July 2018.

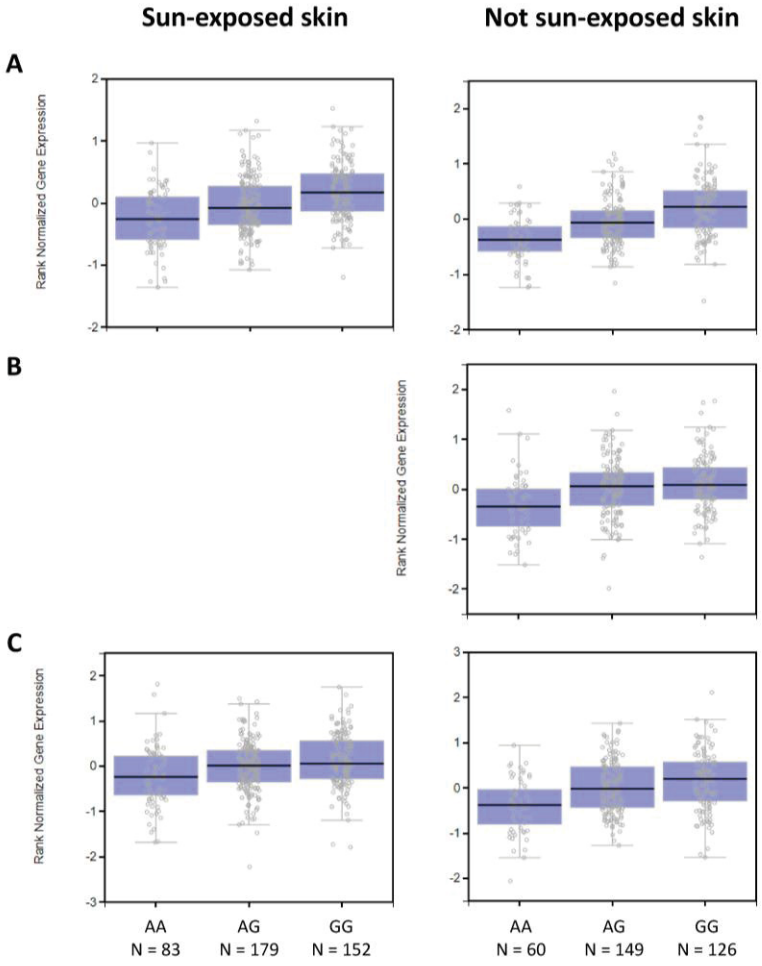


Figure S4: The expression quantitative trait (eQTL) effect of the *rs10439109* genotypes on (A) *long non-coding RNA (lncRNA)-IGF like family member 1* (*lnc-IGFL1, AC006262.1,* ENSG00000268460.1), (B) *IGFL1 pseudogene 1* (*IGFL1P1*, ENSG00000268879.1), and (C) *lncRNA-IGF like family member 2* (*lnc-IGFL2-1*, AC007193.6, ENSG00000267922) genes expression in sun-exposed (left) and not sun-exposed skin (right). Adapted from GTEx Portal (Consortium *et al*., 2015). Accessed on July 2018.

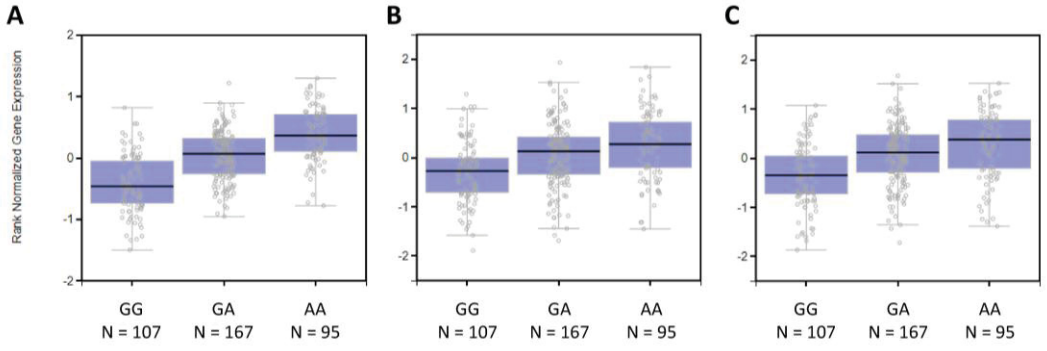


Figure S5: The expression quantitative trait (eQTL) effect of the *rs1654431* genotypes on (A) *retinol dehydrogenase 1* (*RDH13,* ENSG00000160439.11), (B) *NLR pyrin domain containing 2* (*NLRP2*, ENSG00000022556.11), and (C) *long non-coding RNA (lncRNA)- NLRP2* (*lnc-NLRP2*, ENSG00000267265.1) genes expression in whole blood. Adapted from GTEx Portal (Consortium *et al*., 2015). Accessed on July 2018.

APÊNDICE B – DETALHES DE MATERIAIS E MÉTODOS

Amostra de pênfigo foliáceo endêmico e controles

O banco de amostras de DNA do Laboratório de Genética Molecular Humana (LGMH) é composto por aproximadamente 300 amostras de pacientes com pênfigo foliáceo endêmico (PFE) não relacionados geneticamente entre si, diagnosticados com PF através de dados clínicos e imunohistoquímicos. As coletas de material biológico e informações pessoais/clínicas vêm sendo realizadas desde a década de 1980 pela equipe do Laboratório de Genética Molecular Humana (LGMH) da UFPR no Hospital Adventista do Pênfigo em Campo Grande (MS), no Hospital de Dermatologia Sanitária de Piraquara (PR), Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (SP), Hospital das Clínicas da UFPR (PR) e, mais recentemente, também no Hospital de Caridade em Uberaba (MG).

A amostra controle (CTR) é composta por aproximadamente 300 indivíduos sem diagnóstico de PF e sem histórico pessoal e familial de outra doença autoimune. Os indivíduos foram contatados em regiões endêmicas da doença, não são geneticamente relacionados entre si, nem com os pacientes. A distribuição das frequências de etnia, sexo e idade dos controles será semelhante à dos pacientes.

Os pacientes e controles foram esclarecidos quanto aos objetivos da pesquisa, assinando um termo de compromisso livre esclarecido (TCLE, apêndice B) e foram submetidos a averiguação de dados familiares e clínicos (Apêndice C e D). Este estudo foi aprovado pela Comissão Nacional de Ética em Pesquisa (CONEP), CAAE 02727412.4.0000.0096, parecer 505.988 de 2 de janeiro de 2014.

Além de responderem ao questionário, os participantes da pesquisa concederam amostras biológicas – sangue – para posterior processamento e armazenagem de células, extração de DNA e separação de soro/plasma realizados no LGMH/UFPR.

O banco de amostras de pacientes e controles do LGMH continua em crescimento ao realizar contínuas coletas nos hospitais supracitados.

Amostras de DNA e qualidade

O DNA tem sido extraído pelo método de fenol, clorofórmio, álcool isoamílico (SAMBROOK *et al.,* 1989). As amostras de DNA passaram por um controle de qualidade antes de serem submetidas a técnica de microarranjo. Este controle incluiu análise pelos seguintes métodos:

- Medição da concentração de ácidos nucléicos (DNA e RNA) em espectrofotômetro NanoDrop™ (ThermoFisher Scientific), através da absorbância no comprimento de onda de 260 nm e verificação da razão das absorbâncias a 260 nm e a 280 nm para avaliação da contaminação por proteínas que absorvem a 280 nm, e da razão 260 nm / 230 nm para avaliação da contaminação por fenol e outras substâncias;
    - A razão de absorbância 260/280 deve ser maior que 1,75.
    - A razão de absorbância 260/230 deve ser maior que 1,5.
- Análise de integridade ou possível degradação do material genético da amostra de DNA estoque pela técnica de eletroforese em gel de agarose (1)), durante 35 minutos.

Algumas amostras que não foram consideradas com boa qualidade pela presença de contaminantes foram submetidas a um protocolo de purificação por kits comerciais da Qiagen (QIAamp DNA Micro Kit (50)) e Zymo (DNA Clean & Concentrator™-5), conforme indicação do fabricante. As amostras de DNA com as razões de absorbâncias ideais e consideradas de boa qualidade e integridade foram diluídas a concentração de 80ng/ul em um volume total de 30ul em placas com 96 poços.

APÊNDICE C – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO PARA PACIENTES COM PÊNFIGO E GRUPO CONTROLE

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Este é um convite para você participar voluntariamente do projeto de pesquisa **GENES E SUAS VARIAÇÕES: ANÁLISE MOLECULAR, EVOLUTIVA E FUNCIONAL, EM POPULAÇÕES E EM DOENÇAS**. Por favor, leia com atenção as informações abaixo antes de dar ou não seu consentimento para participar deste estudo. Se houver qualquer dúvida sobre o estudo ou sobre este documento, pergunte ao pesquisador com quem você está conversando neste momento.

- OBJETIVO DO ESTUDO

O objetivo desse estudo é conhecer porque certas pessoas manifestam o pênfigo enquanto outras aparentemente são resistentes ao desenvolvimento dessa doença. A predisposição ao pênfigo depende de vários fatores. Alguns desses fatores são do ambiente no qual a pessoa vive e de seus hábitos alimentares ou medicamentos. Outros são genéticos, hereditários. Entretanto, os fatores de predisposição ainda são pouco conhecidos. Esse trabalho de pesquisa científica contribuirá para esclarecer quais os fatores genéticos (hereditários) que tornam a pessoa mais, ou menos, predisposta à doença e como isso acontece. Os resultados poderão auxiliar para a prevenção, tratamento e cura da doença, no futuro.

- PROCEDIMENTOS

Se você participar deste estudo, será colhida uma amostra de seu sangue (cerca de 50 ml, ou seja, correspondente ao volume de um copo de cafezinho) e/ou uma amostra da pele, do mesmo tipo que é colhida para os exames de laboratório solicitados pelo seu médico. Essa amostra de pele (biópsia) será obtida pelos médicos responsáveis. A amostra de sangue será dividida em três partes que servirão para a obtenção de DNA (10 ml), de RNA (15 ml) e de células mononucleares (PBMC) (25 ml). Seu médico ou o prontuário médico será consultado, para confirmação do diagnóstico e tratamento. Além disso, faremos algumas perguntas sobre a sua origem e a origem de seus ancestrais, assim como sobre os seus hábitos e condições de vida. Você pode recusar-se a responder a qualquer das perguntas, se assim desejar.

- RISCOS

Os riscos envolvidos na execução do estudo são raros e transitórios. A coleta da amostra de sangue e, se houver, a coleta da amostra de pele, poderão causar desconforto e, raramente, poderão causar alguma outra complicação como sangramento, hematoma, infecção ou desmaio. Embora não seja esperado, caso você tenha algum problema e não possa ir ao trabalho, você receberá um atestado médico para justificar a sua falta. Se houverem danos decorrentes da pesquisa, você será devidamente indenizado.

- BENEFÍCIOS

Não há nenhum benefício direto desta pesquisa para você. Porém, o conhecimento adquirido através deste estudo poderá auxiliar na compreensão das causas da doença. Isso poderá contribuir para o desenvolvimento de métodos mais eficientes de prevenção e tratamento, com menos efeitos indesejados.

- PARTICIPAÇÃO VOLUNTÁRIA

A sua participação neste estudo é voluntária. Mesmo que você decida participar, terá plena e total liberdade para desistir a qualquer momento, sem que isso acarrete qualquer prejuízo para você.

- ARMAZENAMENTO DAS AMOSTRAS

Caso você autorize, seu material biológico e informações associadas permanecerão armazenados em biorrepositório para utilização no projeto citado acima. Caso haja necessidade de utilizá-lo no futuro, você será contatado para consentir ou não sobre o novo uso de sua amostra. Nesse caso, seu consentimento será formalizado através de um novo TCLE específico. Quando isso não for possível, o fato será justificado perante o Comitê de Ética em Pesquisa (CEP). Toda nova pesquisa a ser feita com o seu material será submetida à apreciação do CEP e sempre que couber também da CONEP, em conformidade com as normas do Conselho Nacional de Saúde (Resolução CNS nº 441/2011). Seu material biológico será armazenado por 05 anos, sob a responsabilidade da Profa. Dra. Maria Luiza Petzl Erler, sendo que, se necessário, serão solicitados mais 05 anos para o Comitê de Ética. Contudo, a qualquer momento você poderá solicitar que seu material biológico não seja mais utilizado, sem que isso acarrete qualquer prejuízo ou constrangimento a você. Entretanto, o material biológico poderá ser descartado devido à inadequação por critérios de qualidade, seguindo normas definidas na legislação brasileira.

- ESCLARECIMENTO DE DÚVIDAS

Você pode e deve fazer todas as perguntas que julgar necessárias antes de concordar em participar do estudo.

- GARANTIA DE SIGILO E PRIVACIDADE

A sua identificação e todos seus dados pessoais, clínicos, de exames laboratoriais e de tratamentos serão mantidos em sigilo. Os resultados do estudo serão publicados sem revelar a sua identidade.

- ACESSO AOS RESULTADOS

Você deseja ser informada(o) sobre os resultados obtidos com esse material biológico?

( ) Não quero ser informada(o) sobre os resultados.

( ) Quero ser informada(o) sobre os resultados.

Estaremos à disposição para informar e explicar sobre todos os resultados obtidos com sua amostra biológica e sobre os estudos nos quais ela foi utilizada. Você poderá nos contatar no Departamento de Genética da UFPR e nos telefones que se encontram neste documento.

- EQUIPE DE PESQUISADORES
- Pesquisadora responsável: Profa. Dra. Maria Luiza Petzl-Erler - Departamento de Genética, UFPR.
- Colaboradores: Dra. Angélica B. Winter Boldt (UFPR); Dra. Danielle Malheiros Ferreira (UFPR); Dr. Luis Eduardo Agner Machado Martins (UFPR); Dra. Ana Maria Roselino; Dr. Alfredo Marquardt Filho; M.Sc. Rodolfo Pessato Timóteo; Dra. Noemi F. Pereira; Dra. Luiza T. Tsuneto (UEM).

Número de telefones: (+41) 3361-1746 e 3361-1724

Eu,_____ abaixo assinado, declaro que li esse termo de consentimento e compreendi a natureza e objetivo do estudo do qual concordei em participar. A explicação que recebi menciona os riscos e benefícios. Eu entendi que sou livre para interromper minha participação a qualquer momento sem justificar minha decisão e sem que esta decisão afete meu tratamento. Declaro também não possuir nenhum grau de dependência profissional ou educacional com os pesquisadores envolvidos nesse projeto (ou seja, os pesquisadores desse projeto não podem me prejudicar de modo algum no trabalho ou nos estudos), não me sentindo pressionado de nenhum modo a participar dessa pesquisa.

Eu concordo voluntariamente em participar deste estudo, e autorizo a coleta, o armazenamento e guarda do meu material biológico sabendo que, se houver uma nova pesquisa, essa será avaliada pelo Comitê de Ética em Pesquisa e sempre que couber também pela CONEP e que serei contatado novamente, para consentir ou não o uso de meu material, através de um TCLE específico.

_____, ____ de _____ de _____
local data

_____	_____
Assinatura do Voluntário	**Assinatura** do Pesquisador
_____	_____
Nome do Voluntário RG	**Nome** do Pesquisador RG

Este Termo de Consentimento Livre e Esclarecido é assinado em duas vias, uma para você e outra que será arquivada pela pesquisadora responsável.

Este Termo de Consentimento foi revisado pelo Comitê de Ética em Pesquisas em Seres Humanos do Hospital de Clínicas da UFPR. Esse Comitê é responsável pela análise das pesquisas e defesa dos interesses dos seus sujeitos, em sua integridade e dignidade. Dessa forma, contribui para o desenvolvimento da pesquisa respeitando os princípios éticos preconizados pela Resolução CNS 466/12 e suas complementares.

Endereço: Bloco Central, Térreo, Hall da Direção, HC-UFPR, Curitiba, PR.

Horário de atendimento: das 08:00 h às 17:00. Fone/Fax: (+41) 3360-1041

APÊNDICE D – FICHA DE AVERIGUAÇÃO PARA PACIENTES DE PÊNFIGO

CÓDIGO DO INDIVÍDUO	Nº DO PRONTUÁRIO	LOCAL	DATA DA AVERIGUAÇÃO	AVERIGUADOR

Material coletado

Biópsia de pele. (1) Sim (2) Não.
Sangue periférico: (1) Sim (2) Não.
Volume total _____ml.
Volume em heparina _____ml.
Volume em EDTA para DNA _____ml, para células __ml.
Volume em ativador de coagulação (para soro) _____ml.

1) IDENTIFICAÇÃO

Nome: _____

Sexo: (1) Masculino (2) Feminino

Estado civil: (1) solteiro(a) (4) divorciado(a)
(2) casado(a) (5) Outros: _____
(3) viúvo(a)

Data de nascimento: ____ / ____ / ____ Município: _____ UF: _____

Endereço: CEP: _____

Telefone: (__) _____

Município: _____ UF: ____ Próximo a: _____

Nome, telefone e endereço de pessoa para contato:

_____

_____

Municípios onde residiu:

Município	Tempo de residência	Exposição a (usar os códigos abaixo e especificar para cada exposição):	Obs.

Código: (1) Animais, (2) Insetos, (3) Lavouras, (4) Rios

Grupo étnico:

(1) Branco (5) Negro
(2) Mulato-claro (6) Indígena
(3) Mulato-médio (7) Oriental
(4) Mulato-escuro (8) Branco brasileiro

Obs.:_____

_____

_____

Escolaridade: _____ anos completos de estudo

Grau de instrução: (1) Analfabeto (5) Médio completo
(2) Fundamental incompleto (6) Superior incompleto
(3) Fundamental completo (7) Superior completo
(4) Médio incompleto (8) Pós-graduação

História ocupacional:

Emprego/Ocupação	Período	Obs.

2) EXPOSIÇÃO A AGENTES QUÍMICOS (considerar: agrotóxicos, tintas, solventes...)

Agente	Período	Frequência (vezes por semana)	Obs.

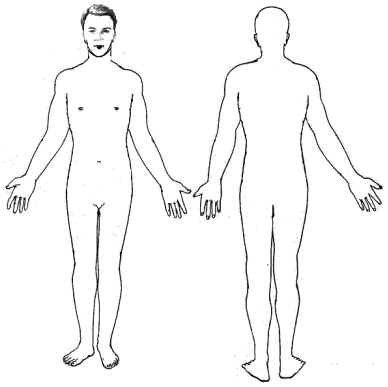
Tipo de habitação: (1) madeira (2) alvenaria (3) mista (4) Outros: _____
Nº de cômodos: _____
Nº banheiros: _____
Saneamento básico: (1) Sim (2) Não
Energia elétrica: (1) Sim (2) Não
Nº de pessoas que vivem na habitação: _____
3) HISTÓRIA FAMILIAR
Nome do pai:
Município de nascimento: _____ UF: _____
Ascendência do pai: (1) européia (2) africana (3) indígena (4) oriental (5) Outros: ___
Cor de pele semelhante à sua? (1) Sim (2) Não (se não, descrever)
Nome da mãe:
Município de nascimento: _____ UF: _____
Ascendência da mãe: (1) européia (2) africana (3) indígena (4) oriental (5) Outros: __
Cor de pele semelhante à sua? (1) Sim (2) Não (se não, descrever)
Pais consanguíneos (considerar mesmo se primos distantes)? (1) Sim (2) Não (se sim, descrever com heredograma, no verso)
Número de irmãos - total: _____
Número de irmãos vivos: _____
Todos filhos dos mesmos pais? (1) Sim (2) Não
Ordem de nascimento dos irmãos:
_____
_____
Quantas vezes engravidou / sua esposa engravidou? _____
Quantos filhos nasceram vivos? _____
Quantos filhos nasceram mortos? _____
Algum aborto (perca)? _____

Alguém na família tem a mesma doença? (1) Sim (2) Não
Nome e grau de parentesco: _____
Caso haja algum caso, fazer o heredograma dos familiares (no verso)
A doença foi semelhante à sua? (1) Sim (2) Não

Caso não seja semelhante, especificar abaixo a diferença:

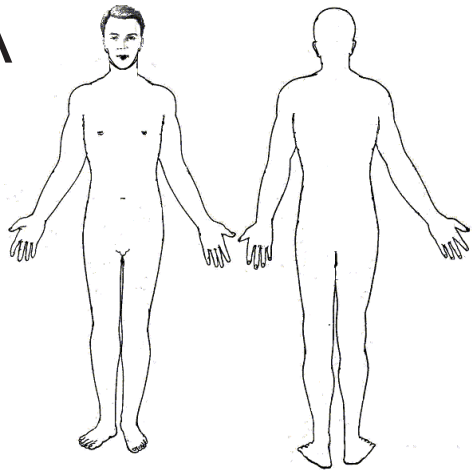
_____
_____
_____

4) ASPECTOS CLÍNICOS
LESÕES
Apresenta lesões no momento da coleta? (1) Sim (2) Não
Região do corpo: marcar os lugares em que há lesões ativas.

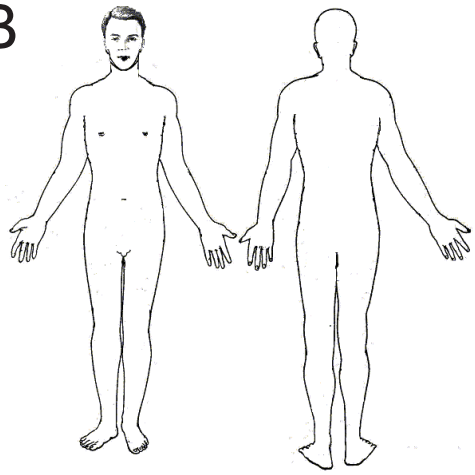


EVOLUÇÃO DAS LESÕES
Município em que apareceu a lesão primária: _____
Idade: _____ Ano: _____ Mês do ano em que apareceu: _____
A lesão se disseminou? (1) Sim (2) Não
Utilize os esquemas abaixo para mostrar: **A**) lesão primária; **B**) evolução em ordem crescente (ex: 1 marca o lugar onde surgiram novas lesões, 2 o segundo e assim sucessivamente). Utilizar a mesma chave numérica na tabela abaixo.

A



B



Local (usar número colocado no esquema acima)	Tempo após lesão primária
1	
2	
3	
4	

DIAGNÓSTICO DEFINITIVO

Base do diagnóstico: (1) Informações clínicas (2) Histopatológico (3) Imunohistoquímica

Pênfigo Vulgar ( ) Pênfigo Foliáceo ( ).

Local do diagnóstico - Instiuição: _____ Município: _____ UF: ____

Informações adicionais sobre diagnóstico:
(1) Informações clínicas:
(2) Histopatológico:
(3) Imunohistoquímica:

TITULAÇÃO DE ANTICORPOS

Anticorpo	Data	Título

TRATAMENTO ATUAL DE PÊNFIGO

1- Prednisona ( ) Dose _____ Posologia _____ Início _____
2- Azatioprina ( ) Dose _____ Posologia _____ Início _____
3- _____ Dose _____ Posologia _____ Início _____

TRATAMENTOS ANTERIORES DE PÊNFIGO

Medicamento	Dose	Via de administração (usar código abaixo)	Posologia	Início (mês/ano)	Fim (mês/ano)	Melhora das lesões? (1) Sim (2) Não

Código: (1) tópico; (2) via oral; (3) via intravenosa

Houve interrupção do tratamento? (1) Sim (2) Não

Interrupção no tratamento:
1- Prednisona ( ) Sim ( ) Não Tempo: _____ Motivo: _____
2- Azatioprina ( ) Sim ( ) Não Tempo: _____ Motivo: _____
3- _____ ( ) Sim ( ) Não Tempo: _____ Motivo: _____

OUTROS MEDICAMENTOS

MEDICAMENTO	DOSE	VIA DE ADMINISTRAÇÃO (usar código abaixo)	POSOLOGIA	INÍCIO (mês/ano)	FIM (mês/ano)

Código: (1) tópico; (2) via oral; (3) via intravenosa

OUTRAS DOENÇAS:

DOENÇA	DATA DE DIAGNÓSTICO	DATA DE CURA	TRATAMENTO

HÁBITOS ALIMENTARES

Costuma ingerir (marcar todos que se aplica):

Alimento	Frequência	Quantidade	OBS
(1) manga			
(2) alho			
(3) alho-poró			
(4) cebola			
(5) água de poço/rio			

APÊNDICE E – FICHA DE AVERIGUAÇÃO PARA GRUPO CONTROLE

CÓDIGO DO INDIVÍDUO	LOCAL	DATA DA AVERIGUAÇÃO	AVERIGUADOR

Material coletado

Biópsia de pele. (1) Sim (2) Não.
Sangue periférico: (1) Sim (2) Não.
Volume total _____ml.
Volume em heparina _____ml.
Volume em EDTA para DNA ____ml, para células ___ml.
Volume em ativador de coagulação (para soro) _____ml.

1) IDENTIFICAÇÃO
Nome: _____
Sexo: (1) Masculino (2) Feminino
Estado civil: (1) solteiro(a) (4) divorciado(a)
(2) casado(a) (5) Outros: _____
(3) viúvo(a)
Data de nascimento: ____ / ____ / ____ Município: _____ UF: _____
Endereço: CEP: _____
Telefone: (__) _____
Município: _____ UF: ____ Próximo a: _____
Nome, telefone e endereço de pessoa para contato:
_____
_____

Municípios onde residiu:

Município	Tempo de residência	Exposição a (usar os códigos abaixo e especificar para cada exposição):	Obs.

Código: (1) Animais, (2) Insetos, (3) Lavouras, (4) Rios

Grupo étnico:

(1) Branco (5) Negro
(2) Mulato-claro (6) Indígena
(3) Mulato-médio (7) Oriental
(4) Mulato-escuro (8) Branco brasileiro

Obs.:_____
_____
_____

Escolaridade:___ anos completos de estudo
Grau de instrução: (1) Analfabeto (5) Médio completo
(2) Fundamental incompleto (6) Superior incompleto
(3) Fundamental completo (7) Superior completo
(4) Médio incompleto (8) Pós-graduação

História ocupacional:

Emprego/Ocupação	Período	OBS

Exposição a agentes químicos (considerar: agrotóxicos, tintas, solventes...)

Agente	Período	Frequência (vezes por semana)	Obs.

Tipo de habitação: (1) madeira (2) alvenaria (3) mista (4) Outros: _____
Nº de cômodos: _____
Nº banheiros: _____
Saneamento básico: (1) Sim (2) Não
Energia elétrica: (1) Sim (2) Não
Nº de pessoas que vivem na habitação: _____

2) HISTÓRIA FAMILIAR
Nome do pai:
Município de nascimento: _____ UF: _____
Ascendência do pai: (1) européia (2) africana (3) indígena (4) oriental (5) Outros: _____
Cor de pele semelhante à sua? (1) Sim (2) Não (se não, descrever)
Nome da mãe:
Município de nascimento: _____ UF: _____
Ascendência da mãe: (1) européia (2) africana (3) indígena (4) oriental (5) Outros: _____
Cor de pele semelhante à sua? (1) Sim (2) Não (se não, descrever)

Pais consanguíneos (considerar mesmo se primos distantes)? (1) Sim (2) Não (se sim, descrever com heredograma, no verso)
Número de irmãos - total: _____
Número de irmãos vivos: _____
Todos filhos dos mesmos pais? (1) Sim (2) Não
Ordem de nascimento dos irmãos:_____
Quantas vezes engravidou / sua esposa engravidou? _____
Quantos filhos nasceram vivos? _____
Quantos filhos nasceram mortos? _____
Algum aborto (perca)? _____
Alguém na família tem pênfigo? (1) Sim (2) Não
Nome e grau de parentesco: _____
Caso haja algum caso, fazer o heredograma dos familiares (no verso)

3) MEDICAMENTOS

Toma algum medicamento? (1) Sim (2) Não

MEDICAMENTO	DOSE	VIA DE ADMINISTRAÇÃO (usar código abaixo)	POSOLOGIA	INÍCIO (mês/ano)	FIM (mês/ano)

Código: (1) tópico; (2) via oral; (3) via intravenosa

4) DOENÇAS

Tem ou teve alguma doença? (1) Sim (2) Não

DOENÇA	DATA DE DIAGNÓSTICO	DATA DE CURA	TRATAMENTO

5) HÁBITOS ALIMENTARES

Costuma ingerir (marcar todos que se aplica):

Alimento	Frequência	Quantidade	OBS
(1) manga			
(2) alho			
(3) alho-poró			
(4) cebola			
(5) água de poço/rio			

APÊNDICE F – DESCRIÇÃO AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO E ESPORÁDICO E CONTROLES

TABELA S1: DESCRIÇÃO DAS AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (BRASIL) E ESPORÁDICO (ALEMANHA) E CONTROLES QUANTO AO SEXO E IDADE.

	Brasil		**Alemanha**	
	PFE	CTR	PF	CTR
Sexo				
	F (%)	F (%)	F (%)	F (%)
Masculino	91 (43,8)	70 (50,0)	40 (53,3)	74 (49,7)
Feminino	112 (53,8)	70 (50,0)	35 (46,7)	75 (50,3)
Idade				
Média	40,5	44,0	65,0	65,0
Mediana	40,0	45,1	61,1	60,3

FIGURA S1: DISTRIBUIÇÃO DA IDADE DE AMOSTRAS DE PÊNFIGO FOLIÁCEO ENDÊMICO (PFE) E CONTROLES (CTR) DO BRASIL.

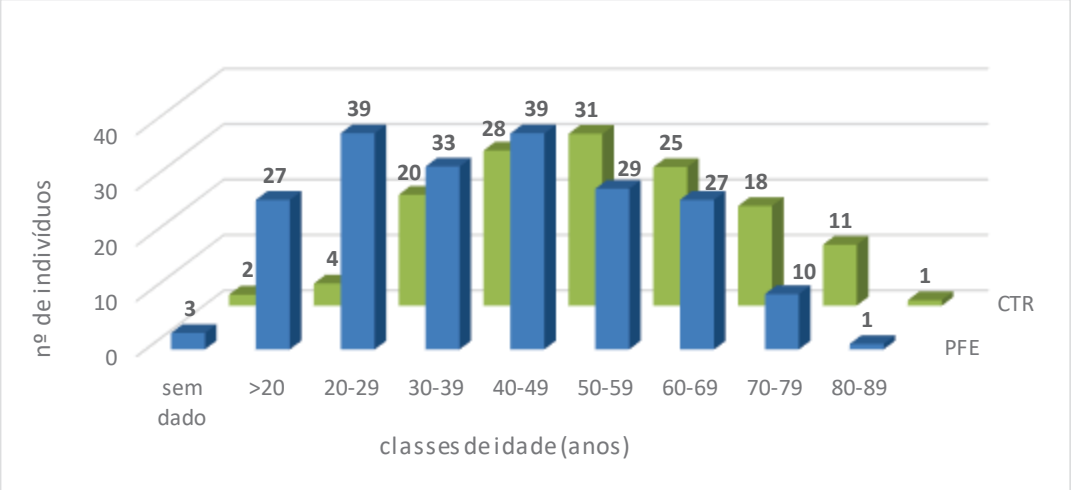
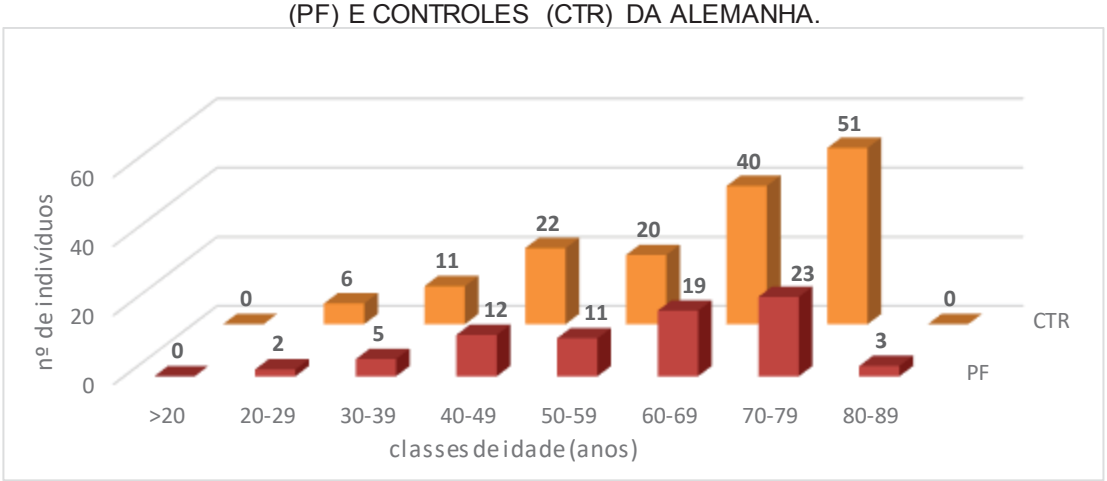


FIGURA S2: DISTRIBUIÇÃO DA IDADE DE AMOSTRAS DE PÊNFIGO FOLIÁCEO ESPORÁDICO (PF) E CONTROLES (CTR) DA ALEMANHA.



APÊNDICE G - EFEITOS FUNCIONAIS DOS SNPS LOCALIZADOS NA REGIÃO 19Q13.42

TABELA S1: POSSÍVEIS EFEITOS FUNCIONAIS DOS SNPS *rs4806488*, *rs465169*, *rs7249071*, *rs681198*, *rs2619223* LOCALIZADOS NA REGIÃO INTERGÊNICA (RI) DO LRC, EM 19Q13.42.

SNP	Posição chr19 (GRCh37)	Alelos	MAF	Distância (pb)	DL (r2)	SNiPA		Regulome DB		Haploreg						
						Efeito de regulação	Genes potenc. regulados	Score	Obs	Marca de histonas no promotor	Marca de histonas no intensificador	DNAse	Ligação proteína	Motivos modificados	eQTL	GENCODE
rs4806488	54526420	T, C	0.37	-552	0.70	Efeito putativo na regulação		5	Evidência fraca		Células sanguíneas, tecido epitelial, muscular, digestivo e outros	Músculo		TEF-1, Zbtb3		10kb 3' do CACNG6
rs465169	54526972	G, A	0.37	0		ENCODE promotor distal associado (intensificador)	LENG8, LILRA3, TMBOAT7, ZNF331, LAIR1, LILRA5, LILRA6, LILRB2, MBOAT7, MYADM, NDUFA3, NLRP12, OSCAR, TSEN34, VSTM1	4	Proteína, cromatina, histona	Células sanguíneas, pulmão		Muitas células e tecidos		Maf	CACNG6	11kb 3' do CACNG6
rs7249071	54527521	T, A	0.27	549	< 0.2	Sítio de ligação de CTCF		2b	Possivelmente afeta ligação	Células sanguíneas, células tronco, pulmão		Muitas células e tecidos	ZNF263	EWSR1, GATA, HDAC2, PAX-4, Pbx3		12kb 3' do CACNG6
rs681198	54527948	A, G	0.40	976	0.77	Intensificador, sítio de ligação de CTCF		4	Proteína, cromatina, histona		Células sanguíneas e cérvix	Célula sanguínea cérvix		Ascl2, HEN1		12kb 3' do CACNG6
rs2619223	54528930	C, G	0.39	1958	0.71	no		5	Evidência fraca			Pele		CHD2, Ik-1, PU.1, Pax-4, RREB-1, SP1, TFII, UF1H3beta, VDR, ZBRB7A, Zfp281, Zfp740		13kb 3' do CACNG6

Efeitos reguladores preditos dos SNPs (*single nucleotide polymorphism*) *rs4806488*, *rs465169*, *rs7249071*, *rs681198*, *rs2619223* presentes na região intergênica (RI) associada com pênfigo foliáceo. Nesta tabela os é possível ver o desequilíbrio de ligação (DL) entre os SNPs com SNP *rs465169* de acordo com os resultados do sequenciamento de nova geração (NGS) realizado neste trabalho. Além disso, é mostrada a posição genômica, distância deles com o SNP *rs461569* em pares de bases (pb) e predição de seus efeitos em elementos reguladores. Os dados foram obtidos pelas ferramentas *online*: SNiPA, RegulomeDB e HaploReg.

Posição: localização em pares de base de cada SNP no cromossomo (chr) 19 (versão do genoma GRCh37/hg19)

eQTL: *expression quantitative trait loci*; GENCODE: Enciclopédia dos genes e suas variantes.

TABELA S2: POSSÍVEL EFEITO NOS NÍVEIS DE EXPRESSÃO GÊNICA PELOS SNPS *rs2619223*, *rs465169* E *rs681198* LOCALIZADOS NA REGIÃO INTERGÊNICA (RI) DO LRC, EM 19q13.42.

SNP	Alelo	Gene	Número no Ensembl	Nome do gene	Regulação	Tecido
rs2619223	G	NLRP2	ENSG00000022556	NLR family pyrin domain containing 2	Alto	Pele exposta ao sol
	G	KIR3DX1	ENSG00000104970	killer cell immunoglobulin like receptor, three Ig domains X1	Alto	Sangue total
	G	TSEN34	ENSG00000170892	tRNA splicing endonuclease subunit 34	Alto	Sangue total
	G	lnc-LENG9; lnc-LILRA4	ENSG00000227407	AC245884.1	Baixo	Pele exposta ao sol
	G	LILRA1	ENSG00000104974	leukocyte immunoglobulin like receptor A1	Baixo	Sangue total
	G	LILRA5	ENSG00000187116	leukocyte immunoglobulin like receptor A5	Baixo	Mucosa do esôfago
	G	LILRB3	ENSG00000204577	leukocyte immunoglobulin like receptor B3	Baixo	Mucosa do esôfago
	G	FCAR	ENSG00000186431	Fc fragment of IgA receptor	Baixo	Mucosa do esôfago
	G	NCR1	ENSG00000189430	natural cytotoxicity triggering receptor 1	baixo	Mucosa do esôfago
rs465169	A	LENG8	ENSG00000167615	leukocyte receptor cluster member 8	Alto	Sangue total
	A	KIR3DX1	ENSG00000104970	killer cell immunoglobulin like receptor, three Ig domains X1	Alto	Sangue total
	A	ZNF813	ENSG00000198346	zinc finger protein 813	Alto	Mucosa do esôfago
	A	lnc-LENG9; lnc-LILRA4	ENSG00000227407	AC245884.1	Baixo	Sangue total Pele exposta ao sol
	A	LILRB3	ENSG00000204577	leukocyte immunoglobulin like receptor B3	Baixo	Mucosa do esôfago
	A	ZNF665	ENSG00000197497	zinc finger protein 665	Baixo	Mucosa do esôfago
	A	FCAR	ENSG00000186431	Fc fragment of IgA receptor	Baixo	Mucosa do esôfago
	A	NCR1	ENSG00000189430	natural cytotoxicity triggering receptor 1	Baixo	Mucosa do esôfago
rs681198	G	NLRP2	ENSG00000022556	NLR family pyrin domain containing 2	Alto	Pele exposta ao sol
	G	KIR3DX1	ENSG00000104970	killer cell immunoglobulin like receptor, three Ig domains X1	Alto	Sangue total
	G	TSEN34	ENSG00000170892	tRNA splicing endonuclease subunit 34	Alto	Sangue total
	G	lnc-LENG9; lnc-LILRA4	ENSG00000227407	AC245884.1	Baixo	Pele exposta ao sol Mucosa do esôfago
	G	LILRA5	ENSG00000187116	leukocyte immunoglobulin like receptor A5	Baixo	Mucosa do esôfago
	G	LILRB3	ENSG00000204577	leukocyte immunoglobulin like receptor B3	Baixo	Mucosa do esôfago
	G	FCAR	ENSG00000186431	Fc fragment of IgA receptor	Baixo	Mucosa do esôfago
	G	NCR1	ENSG00000189430	natural cytotoxicity triggering receptor 1	Baixo	Mucosa do esôfago

Efeitos reguladores preditos dos SNPs (*single nucleotide polymorphism*) *rs465169*, *rs681198* e *rs2619223* presentes na região intergênica (RI) associada com pênfigo foliáceo. Os dados foram obtidos pelo navegador genômico Ensembl considerando $p < 0.01$.

TABELA S3: POSSÍVEL LIGAÇÃO DE FATORES DE TRANSCRIÇÃO NOS SÍTIOS DOS SNPS *rs465169* E rs7249071 LOCALIZADOS NA REGIÃO INTERGÊNICA (RI) DO LRC, EM 19q13.42.

SNP	Alelo	Fatores de transcrição	Posição no motivo	Alta informação da posição	Impacto da mudança de motivo	Matrix de ligação
rs465169	A	ETV2::ONECUT2	16 (de 19)	Sim	Sequência consenso menos semelhante	ENSPFM0155
	A	ETV2::RFX5	20 (de 20)	Não		ENSPFM0161
	A	TFAP2C::ONECUT2	5 (de 25)	Não		ENSPFM0600
	A	MAFF, MAFG, MAFK	14 (de 15)	Não	Sequência consenso mais semelhante	ENSPFM0363
rs7249071	A	HOXB2::ESRRB	15 (de 20)	Sim	Sequência consenso menos semelhante	ENSPFM0309
	A	CUX1, CUX2	2 (de 18)	Sim	Sequência consenso mais semelhante	ENSPFM0044
	A	POU2F1::FOXO6	14 (de 31)	Não		ENSPFM0457
	A	POU2F1::DLX2	3 (de 21)	Não		ENSPFM0448
	A	PITX1::HES7	2 (de 16)	Não		ENSPFM0443
	A	GCM2::DLX3	1 (de 19)	Não		ENSPFM0254
	A	ETV5::TCF3	15 (de 20)	Não		ENSPFM0176
	A	ETV2::SOX15	3 (de 20)	Não		ENSPFM0165
	A	ETV2::ONECUT2, ERF::ONECUT2, FLI1::ONECUT2	11 (de 20)	Não		ENSPFM0156
	A	ETV2::DLX3	3 (de 20)	Não	Sequência consenso menos semelhante	ENSPFM0138
	A	ELK1::SPDEF, ETV2::ETV7, ETV2::SPDEF, ERF::ETV7, ELK1::ETV7, FLI1::ETV7, ETV5::ETV7	7 (de 18)	Não		ENSPFM0132
	A	ETV2::GSC2, ERF::PITX1	3 (de 17)	Não		ENSPFM0122
	A	ETV2::HOXA2, ERF::HOXA3	3 (de 20)	Não		ENSPFM0118
	A	ETV2::FOXI1, ELK1::FOXI1, ERF::FOXI1, FOXO1::ELF1, FOXO1::ELK1, FLI1::FOXI1, ETV5::FOXI1	11 (de 13)	Não		ENSPFM0114
	A	CUX1::HOXA13	9 (de 17)	Não		ENSPFM0047

Efeitos reguladores preditos dos SNPs (*single nucleotide polymorphism*) *rs4806488*, *rs465169*, *rs7249071*, *rs681198*, *rs2619223* presentes na região intergênica (RI) associada com pênfigo foliáceo. Os dados foram obtidos pelo navegador genômico Ensembl.