UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

FERNANDA BUENO BARBOSA

Perfil de variação no número de
cópias do DNA e regiões de perda
de heterozigose na susceptibilidade
ao lúpus eritematoso sistêmico

Ribeirão Preto
2017

FERNANDA BUENO BARBOSA

Perfil de variação no número de
cópias do DNA e regiões de perda de
heterozigose na susceptibilidade ao
lúpus eritematoso sistêmico

Tese apresentada à Faculdade de Medicina de Ribeirão
Preto da Universidade de São Paulo, como requisito parcial
para a obtenção do título de Doutora em Ciências, área de
concentração: Genética

Versão corrigida

A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja
o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD)

Orientador: Prof. Dr. Aguinaldo Luiz Simões

Coorientadora: Profa. Dra. Milena Simioni

Ribeirão Preto
2017

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

FICHA CATALOGRÁFICA

Barbosa, Fernanda Bueno

Perfil de variação no número de cópias do DNA e regiões de perda de heterozigose na susceptibilidade ao lúpus eritematoso sistêmico. Ribeirão Preto, 2017.

214p. : il. ; 30 cm

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo. Área de concentração: Genética.

Orientador: Simões, Aguinaldo Luiz.

Coorientadora: Simioni, Milena

1. Variação no número de cópias. 2. Perda de heterozigose. 3. Lúpus eritematoso sistêmico. 4. Marcadores informativos de ancestralidade. 5. Base genética.

FOLHA DE APROVAÇÃO

FERNANDA BUENO BARBOSA

Perfil de variação no número de cópias do DNA e regiões de perda de heterozigose na susceptibilidade ao lúpus eritematoso sistêmico

Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, como requisito parcial para a obtenção do título de Doutora em Ciências, área de concentração: Genética

Aprovado em: ____/____/2017

BANCA EXAMINADORA

Prof.Dr. _____

Instituição: _____ Assinatura: _____

Prof.Dr. _____

Instituição: _____ Assinatura: _____

Prof.Dr. _____

Instituição: _____ Assinatura: _____

Prof.Dr. _____

Instituição: _____ Assinatura: _____

Prof.Dr. _____

Instituição: _____ Assinatura: _____

'Glorioso' e 'Rainha', esses são os significados de seus nomes.
E também, o que representam em minha vida. Roberto e
Regina, a quem tenho orgulho de chamá-los de pai e mãe,
obrigada por me ensinarem a lutar pelos meus objetivos.

Ao meu padrinho Marco Aurélio, que por meio da sua música
e de seu exemplo de coragem me inspira a andar com fé,
'Fé na vida, fé no homem, fé no que virá!'.

Dedico

AGRADECIMENTOS

Ao orientador Prof. Dr. Aguinaldo Luiz Simões, a palavra que expressa minha admiração e respeito é *gratidão*. Obrigada pela oportunidade de desenvolver esse trabalho, pela confiança, pelas discussões e todos os ensinamentos ao longo desses anos.

À coorientadora Profa. Dra. Milena Simioni pela contribuição no presente trabalho. Obrigada pelas discussões científicas, pela ajuda, prontidão e disponibilidade.

Ao supervisor durante a BEPE, Prof. Dr. Bernardo Lemos, pela oportunidade de desenvolver parte da pesquisa em seu laboratório, cuja experiência foi muito engrandecedora. Obrigada pela confiança.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por ter concedido a bolsa regular de Doutorado (processo FAPESP 2013/17062-9) e a Bolsa de Estágio em Pesquisa no Exterior (BEPE) (processo FAPESP 2016/10306-8).

Aos pacientes voluntários e a todos os indivíduos saudáveis que doaram suas amostras para a realização de pesquisa científica. Ao Prof. Dr. Eduardo Donadi, da FMRP-USP, por ter gentilmente cedido as amostras de pacientes, possibilitando a realização do presente trabalho. À Profa. Dra. Vera Lúcia Lopes e à Profa. Dra. Íscia Cendes, da FCM-UNICAMP, pela colaboração, fundamental para o desenvolvimento desse trabalho.

Ao Centro de Medicina Genômica da FMRP-USP, onde foi realizada parte dos experimentos de *Cytoscan HD array*. Em especial, agradeço à Kamila e Greice pelo auxílio na padronização dos experimentos.

À Cláudia Wiesel, especialista do laboratório de Genética Bioquímica, pelo auxílio técnico-científico nos experimentos e pela prontidão e disponibilidade em ajudar.

Ao Prof. Dr. Jeremy Squire por ter gentilmente cedido acesso à licença do software *Nexus Copy Number* para análise de parte dos dados do presente trabalho.

Ao Departamento de Genética da FMRP-USP pela oportunidade da realização do Doutorado e a todos os professores, funcionários e colegas da pós-graduação.

À banca examinadora pela disposição em participar na contribuição para o crescimento deste trabalho.

A minha família, minha base, meu porto seguro. Em especial aos meus pais Roberto e Regina, irmãos Roberto, Gabriel e João Eduardo e sobrinhos Eduardo e Thaís, pela compreensão, apoio e respeito as minhas ausências. Obrigada por não medirem esforços para me fazerem feliz, pelo incentivo, pela confiança e amor incondicional.

Ao Julian, agradeço por existir alguém como você em minha vida, exemplo de simplicidade, companheirismo e amor. Obrigada por dividir comigo todos os momentos. Mesmo à distância, sempre esteve presente em meu coração.

A todos os colegas do Bloco B, em especial aos do grupo do Prof. Aguinaldo: Edna, Maria, Ana, Cláudia, Natalia, Sabrina, Thaís, Marcela, Ayling, Aline, Malu, Igor, Vanessa e Alice. A nossa convivência sempre foi baseada no respeito, na união, no companheirismo e na solidariedade. Obrigada por tudo.

A todos os colegas do grupo *Lemos Lab*, por ter sido muito bem recepcionada, pela troca de conhecimento científico. Em especial agradeço ao Dr. Melvin Bonilla, pela ajuda nos experimentos e apoio.

À Vanessa Escolano, pela excelente oportunidade de ser sua coorientadora, pela confiança, pelo interesse e responsabilidade em conduzir pesquisa. Agradeço pela troca de conhecimento, por sermos um *time* e por todos os momentos de descontração.

À Natalia Cagnin, Luiza Araújo, Gustavo França e Allysson Allan por serem pessoas que sempre pude contar, pelas discussões científicas, troca de conhecimento e amizade.

À Maria, obrigada por ser esse anjo na vida de todos nós. Agradeço por ter a oportunidade de aprender tanto com você durante esses anos, pela sensibilidade, cuidado, carinho e atenção.

A minha madrinha Patrícia, por sempre ter feito o papel de *segunda mãe*. Obrigada por ser esse exemplo de ternura em minha vida.

Laís, Helder, Bruno, Hine e Nathalia, a vocês meu eterno agradecimento por poder compartilhar todos os momentos, por serem a minha família de Ribeirão Preto. Obrigada por tornarem tudo mais simples, pelas risadas, pelo carinho e amizade.

Aos amigos que fiz na escola de dança, por todos os momentos de *voar com os pés no chão*. Em especial agradeço a Giselle, Wanessa, Eduardo, Renata, Jean, Gleice e Marco pela amizade, pelo carinho e incentivo.

Às amigas Raíssa, Amanda, Luiza e Carolina. A amizade de você é um presente de Deus. Obrigada por fazerem parte da minha vida.

Aos meus amigos, familiares e a todos que de alguma forma, me auxiliaram, incentivaram ou proporcionaram bons momentos para a continuação e finalização deste trabalho.

A Deus, que me iluminou e me deu forças durante toda a trajetória.

‘L'essentiel est invisible pour les yeux’.

Antoine de Saint-Exupéry

RESUMO

Barbosa FB. Perfil de variação no número de cópias do DNA e regiões de perda de heterozigose na susceptibilidade ao lúpus eritematoso sistêmico [tese]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2017; 214p.

O lúpus eritematoso sistêmico (LES) é uma doença autoimune com forte componente genético, caracterizada por inflamação crônica e produção de autoanticorpos. O objetivo deste trabalho foi determinar o perfil de variação no número de cópias (*CNVs*) e de regiões de perda de heterozigose (*LOH*) na patogênese do LES. A detecção de *CNVs* e *LOH* foi feita pela metodologia *Cytoscan HD array* em pacientes com LES ($n = 23$) e indivíduos saudáveis ($n = 110$). Devido à formação tri-híbrida da população brasileira, foi desenvolvido e validado um painel de 345 marcadores informativos de ancestralidade, a partir dados provenientes do próprio *array*, para estimar as proporções de ancestralidade individual e, em última instância, inseri-las nos modelos de regressão logística como variável de controle nas análises de distribuição de *CNVs* e *LOH*. O perfil de *CNVs* evidenciou que o número e o tamanho de duplicações são maiores nos indivíduos saudáveis do que nos pacientes com LES. Duplicações nos genes *FCGR3B* e *ADAM3A* foram descritas como fator de proteção ao LES, quando tais genes foram avaliados por PCR quantitativa em maior grupo amostral de pacientes ($n = 135$) e controles ($n = 200$). Além disso, mostrou-se o efeito sinérgico da presença da deleção em ambos os *loci* *FCGR3B* e *ADAM3A* no aumento do risco para desenvolver a doença. Deleções em pacientes com LES envolvendo os genes *CFHR4*, *CFHR5* e *HLA-DPB2*, previamente descritos em associação com o LES na literatura, foram identificadas por *array* e confirmadas por PCR digital. O protocolo desenvolvido para identificação de variantes raras, resultou em um conjunto de 21 *CNVs* raras em pacientes com LES. Em relação às regiões de perda de heterozigose, não foram encontradas evidências de que o número médio e a extensão dos segmentos *LOH* seja diferente entre pacientes e indivíduos saudáveis. No entanto, os cromossomos 6 e 12 em pacientes exibem regiões de perda de heterozigose em maior quantidade e tamanho do que os de indivíduos saudáveis, além de apresentarem 17 segmentos *LOH* restritos ao grupo de pacientes com LES. Os resultados aqui descritos evidenciam que novos *loci* de susceptibilidade ao LES podem ser encontrados quando a distribuição de *CNVs* é analisada em todo o genoma, em que a investigação de sua relação com a patogênese pode contribuir para a compreensão da base genética da doença.

Palavras-chave: Variação no número de cópias. Perda de heterozigose. Lúpus eritematoso sistêmico. Marcadores informativos de ancestralidade. Base genética.

ABSTRACT

Barbosa FB. DNA copy number variation and loss of heterozygosity profiles in susceptibility to systemic lupus erythematosus [thesis]. Ribeirão Preto: University of São Paulo, Ribeirão Preto Medical School; 2017; 214p.

Systemic lupus erythematosus (SLE) is an autoimmune disease with a strong genetic background characterized by chronic inflammation and autoantibody production. The purpose of this study was to determine the copy number variation (CNV) and loss of heterozygosity (LOH) profiles in the susceptibility to SLE. The detection of CNVs and LOH was performed by the Cytoscan HD array methodology in SLE patients (n = 23) and healthy subjects (n = 110). Due to the tri-hybrid composition of the Brazilian population, a panel of 345 ancestral informative markers was developed and validated, based on data from the array itself, to estimate the proportions of individual ancestry and, ultimately, to insert them into the logistic regression models as a control variable in the analysis of CNV and LOH distribution. The CNVs profile showed that the burden and the size of duplications are higher in healthy individuals than in SLE patients. Duplications in *FCGR3B* and *ADAM3A* genes were described as a protective factor for SLE, when these genes were evaluated by quantitative PCR in a larger SLE (n = 135) and control (n = 200) groups. In addition, the synergistic effect of the presence of deletion in both *FCGR3B* and *ADAM3A* loci increase the risk of developing the disease. Deletions in SLE patients encompassing the *CFHR4*, *CFHR5* and *HLA-DPB2* genes, previously described in the literature in association to SLE, were identified by the array and confirmed by droplet digital PCR. The pipeline developed here for the identification of rare variants resulted in a set of 21 rare CNVs in SLE patients. Regarding the loss of heterozygosity regions, no evidence was found that the mean number and extent of LOH segments is different between patients and healthy individuals. However, the chromosomes 6 and 12 in SLE patients exhibit greater quantity and size of LOH than those of healthy individuals, besides showing 17 LOH segments restricted to the group of SLE. The results described here show that novel susceptibility loci to SLE can be found once the distribution of variants is analyzed throughout the genome, in which the investigation of its relation to the pathogenesis may contribute to the understanding of the genetic basis of the disease.

Keywords: Copy number variation. Loss of heterozygosity. Systemic lupus erythematosus. Ancestry informative markers. Genetics basis.

LISTA DE ILUSTRAÇÕES

Figura 1. Fluxograma de etapas desenvolvidas no presente trabalho.....	37
Figura 2. Diagrama de <i>Venn</i> mostrando a quantidade de marcadores incluídos no painel de 345 <i>SNP-AIMs</i> capazes de diferenciar cada um dos três pares populacionais: ameríndio-europeu (AME-EUR), europeu-africano (EUR-AFR) e africano-ameríndio (AFR-AME).	55
Figura 3. Estimativa de composição tri-híbrida em grupos de população brasileira. Inferência de ancestralidade grupal (<i>Admix 95</i>) nos pacientes com LES mensurada pelos painéis de 345 e 12 <i>AIMs</i> (A). Proporções de ancestralidade individual (<i>Admixture</i> , $K=3$) nos conjuntos de dados genotípicos de 345 <i>SNP-AIMs</i> (B) e de todos os 749.157 <i>SNPs</i> incluídos no <i>Cytoscan HD array</i> (C) nas populações de referência africana (YRI) e europeia (CEU), em casos (LES) e controles brasileiros (BRA). Cada coluna corresponde a um indivíduo. Vermelho, verde e azul representam as estimativas de contribuição africana, europeia e ameríndia, respectivamente. Gráficos de correlação comparando as estimativas de ancestralidade de ~ 750K <i>SNPs</i> e do subconjunto de 345 <i>AIMs</i> para os componentes europeu (D), africano (E) e ameríndio (F) em 311 indivíduos do projeto <i>HapMap</i> (YRI, $n = 89$; CEU $n = 89$), casos (LES, $n = 23$) e controles (BRA, $n = 110$).....	56
Figura 4. Perfil do número total de <i>CNVs</i> no genoma. Histograma mostrando a frequência de cada classe de quantidade de segmentos <i>CNVs</i> em pacientes com LES (A) e controles (B). Estimativa de densidade <i>kernel gaussiano</i> (R, <i>Kernel Density Plot</i>) mostrando a distribuição não-paramétrica de curva de densidade de probabilidade das classes de frequência de <i>CNVs</i> em pacientes (C) e controles (D).	60
Figura 5. Distribuição do total de <i>CNVs</i> , deleções (del) e duplicações (dup) de acordo com as faixas de tamanho de 1–10 Kb, 10–100 Kb e > 100 Kb nos grupos de pacientes com LES e controles brasileiros (BRA).	61
Figura 6. Perfil de deleções e duplicações nos grupos de pacientes com LES e indivíduos saudáveis. <i>Notched box plot</i> (R, <i>Box Plot</i>) mostrando a distribuição da quantidade (A) e do tamanho (B) de deleções e duplicações nos pacientes com LES e em indivíduos saudáveis. As deleções (del) estão identificadas em vermelho, enquanto as duplicações (dup) estão representadas em azul em pacientes com LES e nos controles brasileiros (BRA). As caixas coloridas são a representação do intervalo interquartil, no qual as linhas pretas indicam a mediana dos dados e os círculos vazados sinalizam valores atípicos (<i>outliers</i>). Os <i>notches</i> exibem intervalo de confiança de 95% em torno da mediana. Se não há sobreposição entre os <i>notches</i> , existe evidência de que as diferenças entre os grupos avaliados são significativas (Chambers, 1983), como pode ser observado no número e também no tamanho das duplicações entre casos e controles.....	61
Figura 7. Distribuição de <i>CNVs</i> por cromossomo em pacientes com LES. <i>Violin plot</i> (pacote R: <i>vioplot</i> v. 0.2) ilustrando a distribuição e a densidade de probabilidade da	

quantidade (A) e do tamanho (B) de CNVs por cromossomo. As linhas pretas mais grossas no centro representam a escala interquartil, enquanto as linhas pretas mais finas e longas correspondem aos intervalos de confiança a 95% em torno da mediana, esta, por sua vez, ilustrada pelo ponto branco. 63

Figura 8. Etapas de filtragem de variantes na identificação de CNVs raras em pacientes com LES. O número total de CNVs, incluindo deleções e duplicações, é mostrado em cada círculo referente a um dos passos subsequentes, utilizados na determinação de CNVs raras de acordo com os filtros populacionais aplicados e o uso de diferentes algoritmos na detecção de variantes. 68

Figura 9. Diagrama de Venn destacando os genes que sobrepõem CNVs pertencentes às listas de CNVs raras, CNVRs e CNVs em genes com relevância funcional. Em negrito estão evidenciados os genes em que há variação no número de cópia, cujo genótipo foi avaliado por metodologia alvo-específica. 69

Figura 10. Distribuição de genótipos de CNVs avaliadas por PCR quantitativa. Frequência do número de cópias (CN), de deleções (del) e duplicações (dup) para os genes *ADAM3A* (A, B) e *FCGR3B* (C, D) nos grupos de pacientes com LES e indivíduos saudáveis. 71

Figura 11. Tendência dose-dependente da presença de variação no número de cópias nos genes *ADAM3A* e *FCGR3B* no desenvolvimento de LES. O risco para desenvolver LES pela quantidade de deleções ($CN < 2$) /duplicações ($CN > 2$) nos genes *ADAM3A*, *FCGR3B* e nos 2 *loci* foi estimado em relação aos indivíduos sem deleções/duplicações nos determinados *loci*. Os valores das estimativas pontuais representam *odds ratio* (OR) com intervalo de confiança a 95% nos modelos de regressão logística ajustados para ancestralidade africana (A, D), europeia (B, E) e ameríndia (C, F). 73

Figura 12. Genótipos de CNVs obtidos por PCR digital. Gráficos mostrando o número de cópias (CN) em cada paciente com LES, confirmando deleções heterozigóticas em *CFHR4* no paciente *LES002* (A), em *CFHR5* no paciente *LES005* (B), em *HLA-DPB2* no paciente *LES019* (C), e duplicação heterozigótica envolvendo os genes *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* e *ST8SIA1*, no paciente *LES018* (D). Em todos os casos, o gene de referência (*FOXP2*) mostra *status* diploide invariável nos indivíduos analisados. A amostra utilizada como referência confirma o número de cópias igual a dois para ambos os genes alvo e de referência. 74

Figura 13. Perfil do número total de segmentos *LOH* no genoma. Histograma mostrando a frequência de cada classe de quantidade de segmentos *LOH* em pacientes com LES (A) e controles (B). Estimativa de densidade *kernel gaussiano* (R, *Kernel Density Plot*) mostrando a distribuição não-paramétrica de curva de densidade de probabilidade das classes de frequência de segmentos *LOH* em pacientes (C) e controles (D). 75

Figura 14. Distribuição de segmentos *LOH* por cromossomo em pacientes com LES. *Violin plot* (pacote R: *vioplot* v. 0.2) ilustrando a distribuição e a densidade de probabilidade da quantidade (A) e do tamanho (B) de *LOH* por cromossomo. As linhas

pretas mais grossas no centro representam a escala interquartil, enquanto as linhas pretas mais finas e longas correspondem aos intervalos de confiança a 95% em torno da mediana, esta, por sua vez, ilustrada pelo ponto branco. 76

Figura 15. Etapas de filtragem de segmentos na identificação de *LOH* restritas ao grupo de pacientes com LES. O número total de *LOH* considerando todos os cromossomos autossômicos e o cromossomo X é mostrado em cada círculo referente a um dos passos subsequentes utilizados na determinação de *LOH* exclusivas de pacientes com LES, de acordo com os filtros populacionais aplicados. 78

Figura 16. *Circo plot* revelando a distribuição de *CNVs* e *LOH* no genoma de pacientes com LES. Os círculos representam: (a) os 22 cromossomos autossômicos e o cromossomo X segundo versão *GRCh37/hg19* do genoma humano; (b) o total de *CNVs* identificadas nos pacientes – deleções representadas em vermelho e duplicações representadas em azul, (c) as *CNVs* raras encontradas em pacientes com LES; (d) o total de segmentos *LOH* identificados nos pacientes; (e) os segmentos *LOH* restritos aos pacientes com LES..... 79

LISTA DE TABELAS

Tabela 1. CNVs descritas na literatura em associação com LES.....	30
Tabela 2. Sequência de <i>primers</i> utilizados na <i>qPCR</i>	49
Tabela 3. Eficiência dos <i>primers</i> utilizados na <i>qPCR</i>	49
Tabela 4. Sequência de <i>primers</i> utilizados na <i>ddPCR</i>	50
Tabela 5. Quantidade de <i>SNP-AIMs</i> que diferenciaram populações europeias, africanas e/ou ameríndias presentes no <i>Cytoscan HD array</i>	54
Tabela 6. Modelos de regressão logística ajustados para ancestralidade africana, europeia e ameríndia para comparações de <i>SNPs</i> candidatos à susceptibilidade ao LES.....	58
Tabela 7. Descrição de <i>CNVs</i> com frequência significativamente distinta entre pacientes com LES e controles.	65
Tabela 8. Descrição de <i>CNVs</i> presentes em pacientes com LES que sobrepõem genes com relevância funcional para a patogênese do LES.....	66
Tabela 9. Descrição de <i>CNVs</i> raras identificadas em pacientes com LES a partir do protocolo desenvolvido no presente trabalho.	67
Tabela 10. Descrição de <i>CNVs</i> selecionadas para validação por metodologia alvo-específica.	70
Tabela 11. Descrição de regiões de <i>LOH</i> compartilhadas por mais de um indivíduo, exclusivas do grupo de pacientes com LES.....	77
Tabela 12. Descrição de segmentos <i>LOH</i> restritos do grupo de pacientes com LES a partir da comparação com indivíduos saudáveis brasileiros e de outros grupos populacionais.....	78

LISTA DE ABREVIATURAS E SIGLAS

ACMG	<i>American College of Medical Genetics</i>
ACR	<i>American College of Rheumatology</i>
ADAM3A	<i>ADAM Metallopeptidase Domain 3A</i>
ADAM5	<i>ADAM Metallopeptidase Domain 5</i>
AFR	<i>Africano</i>
AIM	<i>Ancestry Informative Marker</i>
AME	<i>Ameríndio</i>
ANA	<i>Antinuclear Antibody</i>
ANKS1B	<i>Ankyrin Repeat and Sterile Alpha Motif Domain Containing 1B</i>
anti-dsDNA	<i>anti-double strand Deoxyribonucleic Acid</i>
anti-Sm	<i>anti-Smith</i>
AOH	<i>Absence of Heterozygosity</i>
ASW	<i>African ancestry in Southwest USA</i>
BANPI	<i>Banco de Amostras do Núcleo de Pesquisa em Imunogenética</i>
BRA	<i>Controles brasileiros</i>
C3	<i>Complement 3</i>
C4	<i>Complement 4</i>
CCL3L1	<i>C-C Motif Chemokine Ligand 3 Like 1</i>
CEU	<i>Utah residents with Northern and Western European ancestry from the CEPH collection</i>
CFH	<i>Complement Factor H</i>
CFHR4	<i>Complement Factor H Related 4</i>
CFHR5	<i>Complement Factor H Related 5</i>
ChAS	<i>Chromosome Analysis Suite</i>
CN	<i>Copy Number</i>
cnLOH	<i>copy-neutral Loss of Heterozygosity</i>
CNP	<i>Copy Number Polymorphism</i>
CNV	<i>Copy Number Variation</i>
CNVR	<i>Copy Number Variation Region</i>
CRP	<i>C-reactive protein</i>
CT	<i>Cycle threshold</i>
ddPCR	<i>droplet digital Polymerase Chain Reaction</i>
DECIPHER	<i>Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources</i>

<i>DEFB4</i>	<i>Defensin Beta 4A</i>
<i>DGV</i>	<i>Database of Genomic Variants</i>
<i>DL</i>	Desequilíbrio de Ligação
<i>DNA</i>	<i>Deoxyribonucleic Acid</i>
<i>EUR</i>	Europeu
<i>EHW</i>	Equilíbrio de Hardy-Weinberg
<i>FCGR3B</i>	<i>Fc Fragment of IgG Receptor IIIb</i>
<i>FCM</i>	Faculdade de Ciências Médicas
<i>FOXP2</i>	<i>Forkhead Box P2</i>
<i>Fst</i>	<i>Estatística F de Wright</i>
<i>GRCh37</i>	<i>Genome Reference Consortium human genome build 37</i>
<i>GWAS</i>	<i>Genome-Wide Association Studies</i>
<i>HapMap</i>	<i>Haplotype Map</i>
<i>HCFMRP</i>	Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto
<i>hg19</i>	<i>Human Genome version 19</i>
<i>HGDP-CEPH</i>	<i>Human Genome Diversity Cell Line Panel</i>
<i>HLA</i>	<i>Human Leucocyte Antigen</i>
<i>HLA-DPB2</i>	<i>Major Histocompatibility Complex, class II, DP Beta 2</i>
<i>HLA-DRB5</i>	<i>Major Histocompatibility Complex, class II, DR Beta 5</i>
<i>HLA-H</i>	<i>Major Histocompatibility Complex, class I, H</i>
<i>HRH4</i>	<i>Histamine Receptor H4</i>
<i>IBGE</i>	Instituto Brasileiro de Geografia e Estatística
<i>IC 95%</i>	Intervalo de confiança a 95%
<i>IgG</i>	Imunoglobulina G
<i>IL17F</i>	<i>Interleukin 17F</i>
<i>IL21</i>	<i>Interleukin 21</i>
<i>IL22</i>	<i>Interleukin 22</i>
<i>InDel</i>	Polimorfismo de Inserção/Deleção
<i>Kb</i>	Kilobase
<i>LDG</i>	<i>Low-Density Granulocyte</i>
<i>LOH</i>	<i>Loss of Heterozygosity</i>
<i>MAF</i>	<i>Minor Allele Frequency</i>
<i>MAPD</i>	<i>Median Absolute Pairwise Difference</i>
<i>Mb</i>	Megabase
<i>MECP2</i>	<i>Methyl-Cpg Binding Protein 2</i>
<i>MHC</i>	<i>Major Histocompatibility Complex</i>

MXL	<i>Mexican ancestry in Los Angeles, California</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
OR	<i>Odds Ratio</i>
PAX6	<i>Paired Box 6</i>
pb	<i>par de base</i>
PCR	<i>Polymerase Chain Reaction</i>
PCR-RFLP	<i>Polymerase Chain Reaction-Restriction Fragment Length Polymorphism</i>
PMN	<i>Polymorphonuclear Neutrophil</i>
qPCR	<i>quantitative Polymerase Chain Reaction</i>
RABGAP1L	<i>RAB GTPase Activating Protein 1 Like</i>
ROH	<i>Runs of Homozygosity</i>
SNP	<i>Single Nucleotide Polymorphism</i>
SNPQC	<i>Single Nucleotide Polymorphism Quality Control</i>
STAT4	<i>Signal Transducer and Activator of Transcription 4</i>
TLR7	<i>Toll-Like Receptor 7</i>
Tm	<i>Temperatura de melting</i>
UNICAMP	<i>Universidade Estadual de Campinas</i>
USP	<i>Universidade de São Paulo</i>
wavinessSD	<i>waviness Standard Deviation</i>
XCI	<i>X chromosome inactivation</i>
YRI	<i>Yoruba in Ibadan, Nigeria</i>

SUMÁRIO

1 INTRODUÇÃO	21
1.1 VARIAÇÃO NO NÚMERO DE CÓPIAS DO DNA	21
1.1.1 Descrição geral.....	21
1.1.2 Efeitos fenotípicos	22
1.2 REGIÕES DE PERDA DE HETEROZIGOSE NO GENOMA	23
1.2.1 Descrição geral.....	23
1.2.2 Efeitos fenotípicos	25
1.3 LÚPUS ERITEMATOSO SISTÊMICO	26
1.3.1 Descrição geral.....	26
1.3.2 Base genética	28
1.4 EFEITO DA ANCESTRALIDADE EM ESTUDOS GENÉTICOS	31
2 JUSTIFICATIVA	33
3 HIPÓTESE.....	34
4 OBJETIVOS.....	35
4.1 OBJETIVO GERAL	35
4.2 OBJETIVOS ESPECÍFICOS.....	35
5 MATERIAL E MÉTODOS	36
5.1 DESENHO EXPERIMENTAL	36
5.2 AMOSTRAS BIOLÓGICAS	37
5.3 ASPECTOS ÉTICOS.....	39
5.4 CYTOSCAN HD ARRAY	39
5.4.1 Análise laboratorial	39
5.4.2 Dados gerados por Cytoscan HD array	40
5.5 ELABORAÇÃO DE PAINEL DE AIMS.....	41
5.5.1 Seleção de SNP-AIMs	41
5.5.2 Aderência ao equilíbrio de Hardy-Weinberg, diferenciação populacional e composição ancestral.....	41
5.5.3 Validação do painel de AIMS.....	42
5.5.4 Correção estatística pela estratificação populacional.....	43
5.6 ANÁLISE DE DADOS DE CNVs	44
5.6.1 Dados e filtros	44
5.6.2 Distribuição de CNVs	44

5.6.3 Regiões de CNV (CNVRs)	44
5.6.4 CNVs em genes com relevância funcional	45
5.6.5 CNVs raras.....	45
5.6.6 Seleção de CNVs para validação.....	46
5.7 VALIDAÇÃO DE CNVs POR PCR QUANTITATIVA EM TEMPO REAL (qPCR) ..	46
5.7.1 Desenho de <i>primers</i>	46
5.7.2 qPCR.....	47
5.7.3 Padronização dos experimentos de qPCR	48
5.7.4 Análise de resultados	49
5.8 VALIDAÇÃO DE CNVs POR PCR DIGITAL (ddPCR)	50
5.8.1 Desenho de <i>primers</i>	50
5.8.2 ddPCR	51
5.8.3 Análise de resultados.....	52
5.9 ANÁLISE DE DADOS DE LOH.....	52
5.9.1 Dados e filtros	52
5.9.2 Distribuição e regiões de LOH.....	53
6 RESULTADOS	54
6.1 INFERÊNCIA DE ANCESTRALIDADE	54
6.1.1 Painel de SNP-AIMs.....	54
6.1.2 Diferenciação populacional e composição ancestral.....	55
6.1.3 Validação do painel de 345 SNP-AIMs	56
6.1.4 Identificação de associações espúrias	57
6.2 VARIAÇÃO NO NÚMERO DE CÓPIAS (CNVs).....	59
6.2.1 Experimentos de Cytoscan HD array.....	59
6.2.2 Distribuição de CNVs	59
6.2.3 CNVRs	64
6.2.4 CNVs em genes com relevância funcional	66
6.2.5 CNVs raras.....	67
6.2.6 CNVs selecionadas para validação	68
6.2.7 CNVs avaliadas por qPCR	70
6.2.8 CNVs avaliadas por ddPCR.....	73
6.3 PERDA DE HETEROZIGOSE	74
6.3.1 Distribuição de segmentos LOH	74
6.3.2 Regiões de LOH restritas ao grupo de pacientes com LES.....	77
6.3.3 Segmentos LOH restritos ao grupo de pacientes com LES	77
6.4 DISTRIBUIÇÃO DE VARIANTES GENÔMICAS NO LES	79
7 DISCUSSÃO	80
7.1 PAINEL DE AIMs	80
7.2 PERFIL DE CNVs	82
7.3 PERFIL DE LOH	93

8 CONCLUSÕES.....	97
9 REFERÊNCIAS BIBLIOGRÁFICAS	99
10 APÊNDICE.....	118
APÊNDICE A – Características clínicas dos pacientes com LES	118
APÊNDICE B – Painel de 345 <i>SNP-AIMs</i>	122
APÊNDICE C – Genotipagem por <i>Cytoscan HD array</i>	135
APÊNDICE D – Listas de genes com relevância funcional para o LES	139
11 ANEXOS.....	147
ANEXO A – Protocolo <i>Cytoscan HD array</i>	147
ANEXO B – Aprovação do comitê de ética	156
B – 1 Aprovação comitê de ética HCFMRP-USP – Parte 1	156
B – 2 Aprovação comitê de ética HCFMRP-USP – Parte 2.....	157
B – 3 Biorrepositório “LÚPUS E VARIANTES ESTRUTURAIS”	158
ANEXO C – Manuscritos	160
C – 1 Comprovante de publicação artigo 01	160
C – 2 Artigo publicado.....	161
C – 3 Comprovante de submissão artigo 02.....	184
C – 4 Manuscrito submetido	185

1 INTRODUÇÃO

1.1 VARIAÇÃO NO NÚMERO DE CÓPIAS DO DNA

1.1.1 Descrição geral

Os segmentos de DNA ao longo dos cromossomos humanos estão em constante modificação de sequência e estrutura, o que configura processo imprescindível para adaptação e evolução ao decorrer das gerações (Sudmant et al., 2015a). A variação estrutural do *status* diploide pode ocorrer como perdas ou ganhos em segmentos genômicos, que resulta em alteração do número de cópias esperado para cada região de acordo com o genoma de referência. As deleções e duplicações do DNA sob tais condições são coletivamente denominadas de variação no número de cópias (*copy number variation*, *CNVs*) (Zarrei et al., 2015).

Embora a maioria dos trabalhos publicados na área tenha adotado o critério de tamanho >1 Kb para definir *CNV*, atualmente, são reconhecidas *CNVs* > 50 pb. Conforme essa definição de magnitude, as variantes < 50 pb são classificadas como polimorfismos de inserção/deleção (InDel) (MacDonald et al., 2014).

As *CNVs* resultam de diferentes mecanismos mutacionais, incluindo recombinação e processos associados à replicação e ao reparo do DNA (Carvalho e Lupski, 2016). Tais rearranjos cromossômicos ocorrem em apenas um dos cromossomos homólogos, conhecidos como alterações hemi ou heterozigóticas, ou em ambos os cromossomos, que por sua vez, são homozigóticas. Ainda, as *CNVs* não apresentam, necessariamente, genótipos bialélicos. A ocorrência de *CNVs* multialélicas tem sido documentada na literatura englobando extensas regiões no genoma (Handsaker et al., 2015).

As *CNVs* identificadas a partir de estudos populacionais têm ampla distribuição no genoma e caráter hereditário. No entanto, em menor frequência, podem ser geradas durante a meiose, assim classificadas como variações estruturais *de novo* (Yim et al., 2015). A análise do conteúdo das *CNVs* revelou que várias regiões deletadas ou duplicadas do genoma incluem genes e que grande parte delas foi observada em mais de um indivíduo, configurando em diversos casos, caráter polimórfico na população (Almal e Padh, 2012; Itsara et al., 2009). Dessa forma, a descoberta das *CNVs* como

fonte significativa de variabilidade ampliou a visão sobre a variação genética humana (Mills et al., 2011; Sudmant et al., 2010).

A dimensão dessa variabilidade se tornou evidente a partir de 2004, quando dois trabalhos independentes demonstraram que indivíduos saudáveis apresentavam diversas regiões genômicas que variam no número de cópias (Iafrate et al., 2004; Sebat et al., 2004). A partir de então, mapas de *CNVs* foram desenvolvidos para avaliar sistematicamente a contribuição da variação estrutural no genoma humano entre indivíduos presumivelmente saudáveis em diversas populações (McCarroll et al., 2008; Redon et al., 2006; Zarrei et al., 2015)

As plataformas para identificação de *CNVs*, que inicialmente foram baseadas em hibridação genômica comparativa por *arrays* construídos em cromossomos artificiais bacterianos e plataformas de oligonucleotídeos de baixa resolução, têm sido gradativamente substituídas por *microarrays* e sequenciamento de nova geração (Duan et al., 2013; Haraksingh et al., 2017). Tais avanços nas técnicas moleculares de identificação e análise de *CNVs* têm possibilitado identificação de variantes com melhor definição e resolução, o que proporciona mapeamento mais preciso da variação estrutural e seus efeitos sobre as doenças humanas, as características complexas e evolução (Ellingford et al., 2017; Liang et al., 2014).

1.1.2 Efeitos fenotípicos

A variação no número de cópias que não leva a alterações fenotípicas evidentes ocorre com alta frequência entre indivíduos saudáveis, o que implica um papel contributivo das *CNVs* na variação humana normal (Itsara et al., 2009; Schrider e Hahn, 2010). Tais *CNVs* têm sido catalogadas no principal banco de dados para variantes genômicas identificadas em indivíduos saudáveis, o *Database of Genomic Variants* (DGV) (MacDonald et al., 2014). A partir da análise de dados catalogados no DGV, estima-se que 4,8–9,5% do genoma é resultante de variantes no número de cópias, e ainda, é possível a ocorrência de deleção homozigótica em determinados genes sem consequências fenotípicas aparentes (Zarrei et al., 2015).

No entanto, além de constituir parte importante da variação genética, as *CNVs* podem apresentar papel funcional (Conrad et al., 2009). O efeito dessas variantes depende da quantidade de cópias e da sua posição relativa nos genes ou nos elementos reguladores, o que pode resultar em alteração do nível de expressão e interferir na

transcrição (Stranger et al., 2007; Sudmant et al., 2015b). Dessa forma, as CNVs podem contribuir para a variabilidade individual do risco de incidência e desenvolvimento de diversas condições patológicas (Craddock et al., 2010).

Os estudos nesse campo têm sido focados na triagem da extensão total da variação estrutural com o intuito de correlacionar seu efeito sobre os processos evolutivos, características complexas e doenças humanas (Mills et al., 2011; Zarrei et al., 2015). Nesse sentido, CNVs em determinados *loci* têm sido associadas com aumento da susceptibilidade a diversas doenças complexas, como a vários tipos de câncer (Bi et al., 2017; Chen et al., 2013; Hieronymus et al., 2014; Shi et al., 2016), obesidade (Carpenter et al., 2015), autismo (Leppa et al., 2016) e esquizofrenia (Marshall et al., 2016).

Adicionalmente, tem sido documentado que CNVs são enriquecidas de regiões que contêm genes codificantes de proteínas envolvidas no sistema imune (Schaschl et al., 2009; Yim et al., 2015). Em particular, os genes sensíveis à dosagem ou os seus elementos reguladores envolvidos nas CNVs podem desempenhar papel contributivo no desenvolvimento de várias doenças autoimunes, tais como artrite reumatoide (Rahbari et al., 2017), doença de Addison (Bronstad et al., 2011), doença de Crohn (Cleynen et al., 2016), diabetes mellitus tipo 1 (Zanda et al., 2014) e lúpus eritematoso sistêmico (Li et al., 2017).

Mediante o exposto, as CNVs patogênicas têm sido descritas em diferentes classes de doenças. O principal desafio é discriminar as variantes com potencial patogênico daquelas benignas, que também estão presentes em pacientes. Além disso, para atribuir às variantes de significado clínico desconhecido, o grau de patogenicidade e seu envolvimento na etiologia de doenças complexas, são necessários estudos de CNVs em larga escala com a inclusão de dados adicionais de frequência em nível populacional (Gamazon e Stranger, 2015; South et al., 2013).

1.2 REGIÕES DE PERDA DE HETEROZIGOSE NO GENOMA

1.2.1 Descrição geral

Perda de heterozigose (*loss of heterozygosity*, *LOH*), também conhecida pelos termos *absence of heterozygosity* (*AOH*) e *runs of homozygosity* (*ROH*), pode ser definida como longos trechos de genótipos homozigotos consecutivos (Szpiech et al., 2013). A

ausência de heterozigose em trechos contínuos e ininterruptos de uma sequência de DNA no estado diploide reflete segmentos compartilhados sob forma idêntica por descendência (Browning e Browning, 2013). Isso resulta de processos demográficos que reduzem o tamanho populacional e aumentam o nível de homozigose ao longo das gerações, como endogamia, efeito de gargalo e seleção natural (Pemberton et al., 2010).

Quando variantes genéticas, como polimorfismos de nucleotídeo único (*single nucleotide polymorphisms*, *SNPs*) e os microssatélites, se localizam dentro dos segmentos homólogos de DNA, apresentam, portanto, dois alelos homólogos (Ku et al., 2011). A leitura de *arrays* usada para detectar *CNVs* do tipo deleção hemizigótica também conduz à identificação de alelos idênticos. Entretanto, o resultado da homozigose nessa situação é devido à ausência do outro alelo, em vez de homozigose verdadeira, na qual dois alelos homólogos estão presentes (Peiffer et al., 2006). Nesse caso, há necessidade de diferenciar as deleções de uma cópia das regiões em que, de fato, ocorreu perda de heterozigose, conhecidas como *copy-neutral LOH* (*cnLOH*) (Mikulasova et al., 2016).

Anomalias citogenéticas, como a dissomia uniparental, em que duas cópias de um único segmento de DNA homólogo são herdadas de apenas um dos progenitores, também podem resultar em homozigose (Carvalho et al., 2015). Como tal, não pode ser distinguida da homozigose resultante de consanguinidade dos pais e de outros fatores demográficos em estudos em que haja disponibilidade de estudos de trios ou de famílias (Ku et al., 2011). No entanto, a probabilidade de que uma fração considerável de *LOH* seja explicada pela dissomia uniparental é baixa, uma vez que esse evento citogenético é raro (Curtis, 2007).

Na população humana, *LOH* varia de dezenas de kilobases a megabases de tamanho e a sua distribuição nos cromossomos não é uniforme (Pemberton et al., 2010). A frequência de *LOH* ao longo do genoma está correlacionada com variáveis locais, tais como taxa de recombinação, bem como com sinais de seleção positiva recente (Di Gaetano et al., 2014). Em geral, a extensão de homozigose é resultado de desequilíbrio de ligação (DL) entre *loci* dentro de uma região cromossômica, que reflete as taxas de recombinação e a história demográfica da população (McVean et al., 2004). Como a dimensão e a estrutura de blocos de DL variam amplamente em todo o genoma e entre as populações, a distribuição de *LOH* também apresenta

padrões distintos de acordo com a localização geográfica (Coop e Przeworski, 2007; Mezzavilla et al., 2015; Pemberton et al., 2010).

Desse modo, a análise de *LOH* e sua variação entre indivíduos fornece um recurso valioso, muitas vezes inexplorado, para estudar a diversidade genética humana, história evolutiva e para avaliar o mecanismo pelo qual a seleção produz padrões de alelos deletérios.

1.2.2 Efeitos fenotípicos

A história e os fatores culturais de uma população podem afetar os níveis de homozigose em genomas individuais. Em algumas populações, mesmo na ausência de endogamia explícita, a taxa de homozigose pode ser alta, decorrente de isolamento geográfico ou efeitos de gargalo populacional, o que conduz ao aumento do grau de parentesco entre os indivíduos (Gross et al., 2011; Humphreys et al., 2011; McQuillan et al., 2008). Em outras, práticas culturais que promovem casamentos consanguíneos podem resultar em altos níveis de homozigose, mesmo em grandes populações (Curik et al., 2014; Mezzavilla et al., 2015).

Como reflexo da história demográfica, *LOH* pode ocorrer sem consequências patológicas evidentes, constituindo parte da variação encontrada em indivíduos saudáveis (Bennett et al., 2014). Entretanto, no contexto de abordagem genômica, *LOH* tem sido amplamente utilizada como modelo para investigar doenças mendelianas por meio de mapeamento de homozigose (Safieh et al., 2010; Stephen et al., 2015; Woods et al., 2006). Essa ideia parte do pressuposto que indivíduos que herdaram uma mutação idêntica por descendência em determinado locus, também compartilham segmentos de DNA adjacentes à mutação. Para um fenótipo recessivo em indivíduos consanguíneos afetados, o locus de risco em homozigose, contudo, provavelmente está localizado em uma região homozigótica invariavelmente longa (Pemberton et al., 2010).

Seguindo tal princípio, o mapeamento de homozigose de longos trechos permitiu a localização de variantes deletérias e de genes associados a doenças recessivas mendelianas em diversos estudos (Bocquet et al., 2013; Harville et al., 2009; Pang et al., 2010).

Adicionalmente, essa abordagem tem se estendido à busca de regiões candidatas em várias doenças multifatoriais. O primeiro estudo que aplicou a abordagem da

associação de homozigose em larga escala para doenças complexas foi feito por Lencz et al. (2007) em pacientes com esquizofrenia. Outras doenças complexas como autismo (Lin PI et al., 2013), deficiência intelectual (Gandin et al., 2015), doença de *Alzheimer* (Ghani et al., 2015), doença de Parkinson de início precoce (Olgiati et al., 2016), doença arterial coronariana (Christofidou et al., 2015), também já foram estudadas utilizando tal abordagem citogenética. Da mesma forma, vários tipos de câncer mostraram associação com a carga total de *LOH*, como glioma (Mizoguchi et al., 2011), carcinoma de células escamosas de cabeça e pescoço (Marescalco et al., 2014), câncer na tireoide (Thomsen et al., 2016), câncer intracranial (Terashima et al., 2014). Já outros tipos como câncer colorretal (Siraj et al., 2012), leucemia linfoblástica aguda (Hosking et al., 2010) e câncer de mama e próstata (Enciso-Mora et al., 2010), não apresentaram resultados significativos em relação ao perfil de heterozigose.

Até mesmo uma característica não clínica, como a altura, foi examinada usando análises de *LOH* e foi demonstrado que um segmento *LOH* localizado na região 12q21.31 está associado com variação da altura em duas populações americanas (Yang TL et al., 2010).

1.3 LÚPUS ERITEMATOSO SISTÊMICO

1.3.1 Descrição geral

O lúpus eritematoso sistêmico (LES; *MIM* 152700) é uma doença complexa, caracterizada por resposta autoimune dirigida a antígenos nucleares/citoplasmáticos e pela deposição de complexos imunes em diferentes tecidos (Cui et al., 2013). Isto resulta em inflamação local e/ou sistêmica, que pode evoluir para disfunção ou mesmo para a falência dos órgãos acometidos (Frieri, 2013).

Assim como na maioria das doenças autoimunes, o LES é prevalente em mulheres, principalmente em idade fértil (Amur et al., 2012), e apresenta uma proporção de 9 mulheres para cada homem afetado (Margery-Muir et al., 2017). Apesar de ter distribuição global, em geral, é mais comum em afro-americanos, hispânicos, afro-caribenhos, índios nativos norte-americanos, indianos e chineses, do que em populações caucasianas (Danchenko et al., 2006; Jiménez et al., 2003; Pons-Estel G.J. et al., 2010). No Brasil, os dados epidemiológicos sobre o LES mostram que a

incidência é de 4,8 e 8,7 casos/100.000 habitantes/ano, em estudos feitos na região sul (Nakashima et al., 2011) e nordeste (Vilar e Sato, 2002) do país, respectivamente.

Sugere-se que o LES se desenvolva quando linfócitos T autorreativos reconhecem antígenos próprios via complexo principal de histocompatibilidade (*major histocompatibility complex, MHC*) de uma célula apresentadora de antígeno, levando à liberação de citocinas e ativação de células B (Cerosaletti et al., 2014). As células B ativadas diferenciam-se em plasmócitos produtores de autoanticorpos, que causam danos teciduais em decorrência de sua formação e deposição na microvasculatura (Dorner et al., 2010).

Células B autorreativas produzem anticorpos contra antígenos expostos na superfície das células apoptóticas (Bouts et al., 2012). Os autoantígenos que ativam as células T e células B em pacientes com LES podem ser atribuídos à eliminação inapropriada de células apoptóticas. Durante o processo de morte celular, fragmentos de material celular se formam na superfície da célula em apoptose, e apresentam-se na superfície celular. Consequentemente, esses resquícios celulares oferecem estímulo necessário para a ativação das células B autorreativas (Mohan e Putterman, 2015).

Logo, a presença de células B autorreativas leva à produção de autoanticorpos, e em conjunto com a eliminação comprometida do material celular apoptótico, resulta na formação de complexos imunes. Na microvasculatura, estes complexos induzem reações autoimunes, causando a inflamação em inúmeros tecidos e os danos subsequentes associados ao LES (Maidhof e Hilar, 2012).

O depósito de complexos imunes em diferentes órgãos que conduz ao comprometimento multissistêmico, resulta em alta heterogeneidade fenotípica do LES, mesmo entre indivíduos aparentados. A evolução da doença provoca manifestações clínicas polimórficas, com períodos de exacerbação e remissão (Mohan e Putterman, 2015). Sinais e sintomas gerais incluem febre, fadiga, perda de peso, erupções cutâneas, artrite, nefrite, anemia e comprometimento do sistema nervoso central (Fortuna e Brennan, 2013). Apesar de algumas manifestações clínicas serem comuns no LES, devido à heterogeneidade da doença, cada paciente apresenta um conjunto único de identificadores (Pons-Estel G.J. et al., 2010).

A heterogeneidade clínica do LES levou ao estabelecimento um conjunto de critérios de diagnóstico propostos pelo *American College of Rheumatology* (ACR) em 1982 (Tan et al., 1982) e revisados em 1997 (Hochberg Marc C, 1997). Para a confirmação do

diagnóstico, é necessária a presença de, no mínimo, quatro das onze seguintes manifestações clínico-laboratoriais: (1) eritema malar, (2) lesão discoide, (3) fotossensibilidade, (4) úlceras orais/nasais, (5) artrite não erosiva, (6) serosite: pleurite/pericardite, (7) comprometimento renal: proteinúria persistente/cilindrúria anormal, (8) distúrbios neurológicos: convulsões/psicoses, (9) alterações hematológicas: anemia hemolítica/leucopenia/linfopenia/plaquetopenia, (10) alterações imunológicas: anticorpos para antígenos nucleares/anti-dsDNA/antifosfolipídeos, (11) anticorpos antinucleares (Yu C et al., 2014).

1.3.2 Base genética

O mecanismo etiológico do LES não está suficientemente esclarecido, pois além dos fatores imunológicos centrados na perda da autotolerância, os fatores genéticos, hormonais e ambientais também desempenham papel importante no desenvolvimento da doença (Kamen, 2014; Rhodes e Vyse, 2008; Yang ML et al., 2012).

Estudos sobre a ligação entre a genética e o LES têm mostrado alta predisposição para desenvolver a doença dentro de famílias (Teruel e Alarcon-Riquelme, 2016). Parentes em primeiro grau de pacientes com LES são significativamente mais susceptíveis à doença em comparação com o restante da população (Maidhof e Hilas, 2012). Além disso, relatos que confirmam a relação do risco elevado de irmãos desenvolverem LES e as maiores taxas de concordância entre gêmeos monozigóticos (24-57%) do que entre gêmeos dizigóticos ou outros irmãos completos (2-5%), sugerem que o LES possui uma base genética complexa (Alarcon-Segovia et al., 2005; Amur et al., 2012).

O aumento do conhecimento da estrutura do genoma humano e o desenvolvimento de tecnologias para identificação e análise do DNA resultaram na disponibilidade de conjuntos de marcadores de alta densidade para a genotipagem de variantes nos indivíduos (LaFramboise, 2009; McCarthy et al., 2008). Dessa forma, tornou-se possível desenvolver estudos de associação genômica em larga escala (*genome-wide association studies*, GWAS), com capacidade para rastrear milhões de SNPs em todo o genoma, sem conhecimento prévio de genes ou de regiões candidatas (Visscher et al., 2012). Com a utilização destes recursos, foram identificados SNPs em mais de 60 *loci* associados com o risco para desenvolver LES em diversos grupos ancestrais, como por exemplo, nos genes: *IRF5*, *STAT4*, *TNFAIP3*, *PTPN22*, *BLK*, *BANK1*, *TNFSF4*,

ITGAM, *KIAA1542*, *PXK* e *MECP2* (Alarcon-Riquelme et al., 2016; Armstrong et al., 2014; Deng e Tsao, 2010; Harley ITW et al., 2009; Sun et al., 2016).

Associação entre CNVs e LES

Além dos *SNPs* associados com LES, as *CNVs*, como fonte de variabilidade genética, podem influenciar nas diferenças interindividuais na susceptibilidade às doenças por meio de diversos mecanismos que afetam a estrutura do DNA e expressão gênica (Craddock et al., 2010; Henrichsen et al., 2009; Ptacek et al., 2008).

Deleções envolvendo os genes *C4* e *FCGR3B* foram associadas com o aumento do risco para o desenvolvimento de LES em diferentes populações. Embora em certas populações estudadas tais achados não tenham sido replicados, resultados de meta-análise confirmaram a diminuição do número de cópias nesses genes como fator de risco para desenvolvimento de LES (Li et al., 2017; Lv et al., 2010; Yuan et al., 2015).

CNVs em outros genes foram associadas ao LES, como *CCL3L1* (Mamtani et al., 2007), *TLR7* (Garcia-Ortiz et al., 2010), *DEFB4* (Zhou XJ et al., 2012) e *HLA-DRB5* (Wu et al., 2014). Todavia, poucos destes achados foram avaliados em diferentes populações, e quando avaliados, a associação da *CNV* com o desenvolvimento de LES não foi replicada (Kelley et al., 2007).

As pesquisas realizadas com *CNVs* e LES citadas anteriormente avaliaram variantes em genes candidatos, previamente relacionados com a doença. Kim JH et al. (2013), utilizaram, pela primeira vez, a metodologia de investigação genômica em larga escala para avaliar a distribuição de *CNVs* de modo global no genoma de mulheres sul-coreanas portadoras de LES e em seus respectivos controles. Três regiões de *CNV* (*CNVRs*) do tipo deleção, duas incluindo genes: *RABGAP1L*, *C4*, e uma região sem gene no cromossomo 10q21.3, foram associadas com aumento da susceptibilidade ao LES (Tabela 1).

Tabela 1. CNVs descritas na literatura em associação com LES.

Localização	Gene	Tipo	População	Referência
1q23	<i>FCGR3B</i>	Del	Várias (meta-análise)	Yuan et al., 2015
1q24	<i>RABGAP1L</i>	Del	Sul-coreanas (mulheres)	Kim JH et al., 2013
4q27	<i>IL-21</i>	Dup	Chineses	Yu B et al., 2011
6p12.2	<i>IL-17F</i>	Dup	Chineses	Yu B et al., 2011
6p21.32	<i>HLA-DRB5</i>	Dup	Chineses	Wu et al., 2014
6p21.3	<i>C4</i>	Del	Várias (meta-análise)	Li et al., 2017
8p23.1	<i>DEFB4</i>	Dup	Chineses	Zhou XJ et al., 2012
10q21.3	-	Del	Sul-coreanas (mulheres)	Kim JH et al., 2013
12q15	<i>IL-22</i>	Dup	Chineses	Yu B et al., 2011
17q21.1	<i>CCL3L1</i>	Del/Dup	Americanos e colombianos	Mamtani et al., 2007
18q11.2	<i>HRH4</i>	Dup	Chineses	Yu B et al., 2010
Xp22.2	<i>TLR7</i>	Del	Mexicanos (crianças)	Garcia-Ortiz et al., 2010

Del=deleção; Dup=duplicação.

Associação entre LOH e LES

Embora o mapeamento de regiões em homozigose tenha sido focado primeiramente em indivíduos consanguíneos, os avanços recentes fornecem base potencial para o mapeamento da homozigose em indivíduos não aparentados portadores de doenças complexas, além do estudo de doenças mendelianas (Christofidou et al., 2015; Ghani et al., 2015; Marescalco et al., 2014).

Especificamente em relação ao LES, Singh et al. (2014) avaliaram o perfil de CNVs e de LOH em pares autólogos de neutrófilos de densidade normal e de uma população anormal de neutrófilos de pacientes com LES, os granulócitos de baixa densidade (*low-density granulocytes; LDGs*). A comparação entre as duas populações de neutrófilos mostrou que há aumento do tanto do número de cópias quanto do número de regiões em homozigose nas amostras de LDGs em relação aos neutrófilos densidade normal. Este trabalho também descreveu uma região de LOH no cromossomo 5q23.3-5q31.1, compartilhada entre 4/13 pacientes com LES analisados, que contém diversos genes de citocinas (*IL-3, IL-4, IL-5, IL-13, CSF2*), e genes envolvidos com regulação da resposta imune (*IRF1, SLC22A4*) e reparo de DNA (*RAD50*).

1.4 EFEITO DA ANCESTRALIDADE EM ESTUDOS GENÉTICOS

Estudos que investigam a base genética tem sido realizados para identificar as variantes que contribuem para a susceptibilidade às doenças complexas em humanos (Lin D e Zeng, 2011). Na concepção clássica de estudos do tipo caso-controle, as frequências alélicas entre indivíduos portadores de determinada doença e indivíduos saudáveis são comparadas (Farhi et al., 2016; Sacks et al., 2014). Entretanto, associações espúrias, não relacionadas aos *loci* causais, podem ser obtidas em decorrência da subestrutura populacional não mensurada (Coelho et al., 2015; Gomes et al., 2017).

A questão dos efeitos de confusão surge quando a população investigada é composta por várias subpopulações ancestrais com diferentes frequências alélicas e risco para desenvolver certas condições patológicas não representado igualmente nos casos e controles. Este viés nas frequências dos alelos pode resultar em associações falso-positivas na análise estatística (Birney et al., 2016; Cardon e Palmer, 2003; Shriner et al., 2011; Tang et al., 2005).

Os marcadores informativos de ancestralidade (*ancestry informative markers, AIMS*), aqueles exibem diferencial de frequência alélica (δ) superior a 30% entre quaisquer duas populações parentais, são comumente utilizados para minimizar os efeitos da estratificação populacional em estudos genéticos (de Andrade Ramos et al., 2016; Maxwell, 2014; Qin et al., 2014).

Nesse contexto, a população brasileira representa um desafio especial devido a sua formação tri-híbrida, decorrente de mais de cinco séculos de mistura populacional em diferentes proporções em todo o território nacional. Mais especificamente, é resultado do processo de mistura entre ameríndios, europeus, africanos e, mais recentemente, indivíduos de outras nacionalidades (Salzano e Bortolini, 2005).

Em 1500, quando a colonização europeia se iniciou com a chegada de aproximadamente 500.000 portugueses, a maioria homens, os ameríndios autóctones representavam uma população de aproximadamente 2,4 milhões de indivíduos (Callegari-Jacques e Salzano, 1999; Ribeiro, 2015). O tráfico de escravos teve início na metade do século XVI e, até meados do século XIX, foram trazidos por volta de 4 milhões de africanos ao Brasil (Rodrigues, 1932; Salzano, 1986). Tais escravos eram provenientes de diferentes regiões da África subsaariana, predominantemente de regiões que pertencem atualmente à Angola, ao Congo e à

Moçambique (Curtin, 1972). Entre 1820 e 1975, estima-se que 6 milhões de imigrantes chegaram ao Brasil, em sua maioria portugueses e italianos, mas também grupos menos numerosos de espanhóis, alemães, japoneses, sírios e libaneses (Ribeiro, 2015).

Em casos de populações que apresentam grau de mistura, como a brasileira, estabelecer um conjunto de *AIMs* aplicável à população em questão é um passo crítico para realizar análises de associação genômica em doenças complexas para inferir ancestralidade e ajustar pela estratificação da população em estudos de caso-controle.

2 JUSTIFICATIVA

Ressalta-se que ainda persistem lacunas significativas na compreensão da base genética do LES, principalmente quando se diz respeito às variantes genômicas, como *CNVs* e *LOH* (Teruel e Alarcon-Riquelme, 2016). A importância de *CNVs* e *LOH* na susceptibilidade e no desenvolvimento de patologias complexas, como o LES, e em particular, a existência de poucos estudos em larga escala relacionando essa patologia a esses tipos de marcadores genéticos, motivam o desenvolvimento de pesquisas nessa área (Kim JH et al., 2013; Singh et al., 2014). Em 2012, foi iniciado o estudo piloto sobre *CNVs* em pacientes com LES no nosso grupo de pesquisa (Barbosa, 2013). A complexidade do tema levou à necessidade de aprimoramento, que se deu na forma de continuação do trabalho.

A abrangência das *CNVs* como fonte de variabilidade genética, seu potencial para alterar significativamente a expressão gênica, os estudos recentes que demonstram a contribuição de *CNVs* na susceptibilidade às doenças e suas implicações diretas para determinados fenótipos clínicos, destacam necessidade de aprimoramento do estudo deste tipo de variação estrutural na compreensão de fenótipos autoimunes (Fanciulli M et al., 2010; Yih Chen et al., 2016; Yim et al., 2015).

Em relação às regiões de *LOH* e sua associação com doenças complexas, diversos estudos têm mostrado o papel contributivo de regiões cujo acúmulo de variantes deletérias em longos trechos de homozigose acarreta em aumento do risco para desenvolver determinadas patologias (Ghani et al., 2015; Lencz et al., 2007; Szpiech et al., 2013). Nesse sentido, a abordagem genômica e a sua utilização na investigação etiológica de doenças complexas é uma estratégia adequada e plausível.

Considerando que a composição ancestral de populações pode alterar resultados de testes de associação, o estudo em populações heterogêneas pode produzir resultados diferentes e até mesmo conflitantes com os da literatura (Cardon e Palmer, 2003; Liu J et al., 2013; Shriner et al., 2011). A insuficiência de estudos que descrevem o perfil de *CNVs* e *LOH* em populações tri-híbridas, tanto quando se trata de indivíduos saudáveis ou de fenótipos patogênicos, dificulta a comparação com os bancos de dados disponíveis atualmente, pois alguns dos resultados obtidos podem ser discrepantes de outras populações, ou até mesmo representarem associações espúrias.

3 HIPÓTESE

A hipótese deste projeto fundamenta-se em que portadores de LES possuem variantes no número de cópias e regiões de perda de heterozigose em segmentos genômicos específicos, mais frequentes nos afetados do que em controles, que contribuem para etiologia da doença. Além disso, sugere-se que o efeito sinérgico de tais *CNVs* possa conduzir ao aumento da susceptibilidade à doença.

4 OBJETIVOS

4.1 OBJETIVO GERAL

O presente trabalho tem como objetivo geral promover análise da distribuição de *CNVs* e de regiões de *LOH* no genoma de portadores de LES em comparação com indivíduos saudáveis.

4.2 OBJETIVOS ESPECÍFICOS

- (1) Genotipar pacientes com LES e indivíduos saudáveis com o intuito de detectar *CNVs* e segmentos *LOH* no genoma;
- (2) Selecionar e validar um conjunto de *AIMs* capaz de diferenciar europeus, africanos e ameríndios, e aplicá-lo para inferência de ancestralidade em pacientes com LES e indivíduos saudáveis;
- (3) Analisar a distribuição e o padrão de *CNVs* no genoma dos pacientes com LES em comparação com indivíduos saudáveis brasileiros e de outros grupos populacionais;
- (4) Desenvolver um protocolo de etapas para a identificação de *CNVs* raras em pacientes com LES;
- (5) Confirmar os resultados de genotipagem de *CNVs* por metodologia alvo-específica;
- (6) Delinear o perfil de *LOH* autossômico e de *LOH* no cromossomo X em pacientes com LES, e compará-lo com o de indivíduos saudáveis brasileiros e de outros grupos populacionais.

5 MATERIAL E MÉTODOS

5.1 DESENHO EXPERIMENTAL

A metodologia *Cytoscan HD array*, que produz dados genômicos de *CNVs*, *SNPs* e *LOH* a partir de amostras de DNA, é adequada para atender aos objetivos iniciais propostos no presente trabalho. Portanto, pacientes com LES e indivíduos saudáveis foram submetidos à genotipagem por esse tipo de ensaio.

Devido à formação tri-híbrida da população brasileira, à heterogeneidade do LES e o risco para desenvolver a doença variar de acordo com a região geográfica e ancestralidade, o primeiro passo foi estabelecer um conjunto de *AIMs*, a partir dos *SNPs* presentes no *Cytoscan HD array*, para inferir ancestralidade e avaliar diferenciação populacional entre os grupos de pacientes com LES e controles. Os modelos de regressão logística foram ajustados pela ancestralidade usando dados obtidos a partir da aplicação do painel de *AIMs* nas análises estatísticas subsequentes.

Em seguida, as *CNVs* presentes nos pacientes com LES e nos controles foram identificadas. A distribuição do número total e extensão de *CNVs* foi estimada nos pacientes com LES e indivíduos saudáveis, bem como a distribuição por cromossomo. Foram identificadas variantes compartilhadas por mais de um indivíduo, e a partir dessa análise, foram selecionadas *CNVs* candidatas à associação com o LES. Com o intuito de confirmar a relação com a doença, *CNVs* candidatas foram genotipadas por metodologia alvo-específica – PCR quantitativa em tempo real (*quantitative real-time PCR*, *qPCR*) – em maior grupo amostral.

Um protocolo de etapas para determinar *CNVs* raras em pacientes com LES foi desenvolvido e, além disso, *CNVs* em genes com relevância funcional para a patogênese foram identificados. A frequência de tais variantes foi avaliada em outras populações e em bancos de dados públicos. As *CNVs* identificadas que apresentaram potencial patogênico foram validadas por metodologia altamente específica – PCR digital (*droplet digital PCR*, *ddPCR*).

Para avaliar o perfil de *copy-neutral LOH* nos pacientes com LES, os eventos de *LOH* foram comparados àqueles presentes nos indivíduos do grupo controle e de outros

grupos externos. A análise decorrente das comparações identificou *LOH* exclusivas do grupo de pacientes com LES.

As etapas desenvolvidas nesse trabalho estão resumidas na Figura 1.

Figura 1. Fluxograma de etapas desenvolvidas no presente trabalho.

5.2 AMOSTRAS BIOLÓGICAS

Para o desenvolvimento desse trabalho, foi utilizado o total de 335 amostras. Nas técnicas iniciais de triagem por *array* foram utilizados 23 pacientes com LES e 110 indivíduos saudáveis, enquanto as etapas de validação foram feitas em 135 casos e 200 controles.

As amostras de LES foram provenientes de pacientes atendidos no ambulatório de Colagenoses da Clínica Médica do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP-USP), e diagnosticados segundo os critérios revisados do *American College of Rheumatology* (Hochberg M.C., 1997). Tais amostras pertencem ao Banco de Amostras do Núcleo de Pesquisa em Imunogenética (BANPI), sob coordenação do colaborador Prof. Dr. Eduardo Antônio Donadi.

Os dados clínicos e laboratoriais dos pacientes com LES foram obtidos a partir de consulta ao prontuários médicos. A média de idade de diagnóstico para o grupo de 135 pacientes foi de 32 ± 12 anos. O número de pacientes do sexo feminino/masculino foi de 125/10. As características clínicas apresentadas pelos portadores foram agrupadas em oito categorias cuja presença da característica em questão mostrou as seguintes frequências: (1) nefrite: 59%, (2) artrite: 60%, (3) rash malar: 36%, (4) úlceras orais e/ou nasais: 12%, (5) fotossensibilidade: 38%, (6) convulsões e psicoses: 22%, (7) alterações hematológicas definidas por presença de anemia ferropriva e/ou leucopenia ($< 4.000/\text{mm}^3$) e/ou linfocitopenia ($< 1.500/\text{mm}^3$) e/ou trombocitopenia ($< 150.000/\text{mm}^3$): 53%, (8) alterações imunológicas definidas por presença de anti-DNA e/ou anti-Sm e/ou anticorpos antifosfolipídeos (lúpus anticoagulante e anticardiolipina): 71%. A tabela completa está descrita no Apêndice A.

As amostras de grupo controle foram obtidas em colaboração com os grupos de pesquisa da Profa. Dra. Vera Lúcia Gil da Silva Lopes e da Profa. Dra. Íscia Lopes Cendes, do Laboratório de Citogenética Humana, Faculdade de Ciências Médicas, Universidade de Campinas (FCM-UNICAMP). Essas amostras foram coletadas de indivíduos saudáveis, voluntários, não aparentados, de ambos os sexos.

O DNA de todas as amostras utilizadas foi obtido em fases já realizadas de outros projetos e encontrava-se estocado. A extração de DNA foi realizada a partir de sangue periférico colhido com anticoagulante ácido etilenodiamino tetra-acético, segundo o método *salting-out* (Miller et al., 1988) em amostras do grupo de pacientes com LES, ou por *kit* comercial *QIAamp® DNA Blood Maxi* (Qiagen, Hilden, Alemanha) em amostras do grupo controle.

5.3 ASPECTOS ÉTICOS

O presente trabalho foi aprovado pelo Comitê de Ética em Pesquisa do HCFMRP-USP (CAAE: 03199712.0.0000.5440) e da instituição coparticipante FCM-UNICAMP (CAAE: 35316314.9.1001.5404).

Para o armazenamento de amostras de DNA para fins de pesquisa e análise científica, foi criado o biorrepositório "Lúpus e Variantes Estruturais" junto ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, sob responsabilidade do Prof. Dr. Aguinaldo Luiz Simões e da pós-graduanda Fernanda Bueno Barbosa. A criação do biorrepositório foi aprovada pelos comitês de ética do HCFMRP-USP (CAAE: 57570716.9.0000.5440) e da instituição coparticipante FCM-UNICAMP (CAAE: 57570716.9.3001.5404).

5.4 CYTOSCAN HD ARRAY

5.4.1 Análise laboratorial

O total de 23 amostras de pacientes com LES e 110 indivíduos saudáveis foi submetido à genotipagem por *Cytoscan HD array* (Affymetrix, CA, EUA). Com o intuito de avaliar a integridade do material genético, foi feita eletroforese em gel de agarose. As amostras de DNA foram diluídas em água livre de nucleases à concentração de 50 ng/μL e, em seguida, quantificadas em espectrofotômetro *NanoDrop™ 1000* (Thermo Fisher Scientific, MA, EUA) para verificar se o DNA atingiu concentração e grau de pureza esperados.

Após a verificação da integridade do DNA e diluição das amostras, os experimentos de *Cytoscan HD array* foram realizados de acordo com as recomendações do fabricante. Reagentes, soluções e procedimentos detalhados das etapas do *Cytoscan HD array* encontram-se no Anexo A. As etapas principais do protocolo são enumeradas a seguir: (1) digestão das amostras de DNA com a enzima *NspI*, (2) reação de ligação aos adaptadores para *NspI*, (3) amplificação dos fragmentos por PCR, (4) purificação do DNA com *beads* magnéticas, (5) fragmentação dos produtos da PCR, (6) marcação do DNA por fluorescência, (7) hibridação das amostras de DNA marcadas aos *chips Cytoscan HD* no *GeneChip® Hybridization Oven 640* (Affymetrix®), (8) lavagem dos chips no *GeneChip® Fluidics Station 450* (Affymetrix®), (9) escaneamento dos chips no *GeneChip® Scanner 3000 7G* (Affymetrix®), (10) análise

das imagens escaneadas de cada *chip* no *GeneChip® Operating software* (GCOS) (Affymetrix®).

5.4.2 Dados gerados por Cytoscan HD array

O *Cytoscan HD array* contém aproximadamente 1,9 milhões de sondas não polimórficas e 750 mil *SNPs*. Para cada *chip*, que corresponde a um indivíduo, a conversão da intensidade do sinal de cada sonda escaneada do *chip* em valores de logaritmo ($\log_2$ ratio) é realizada pelo *software* GCOS, o que origina arquivos com extensão *.cel*. O arquivo *.cel* referente a cada indivíduo é utilizado como *input* para que por meio do *software* *Chromosome Analysis Suite v.3.0* (ChAS) (Affymetrix®) se obtenha: (1) a estimativa da qualidade dos resultados gerados por *Cytoscan HD array*, (2) a determinação dos genótipos das sondas bialélicas e o do número de cópias (*copy number*; *CN*) dos segmentos do genoma utilizando anotação do genoma humano *GRCh37/hg19*, por meio de algoritmo baseado em *hidden markov model*, (3) a definição do sexo do indivíduo, baseado no algoritmo *Y-Gender*, (4) a criação de arquivo com extensão *.cyhd.cychp* para cada indivíduo, utilizado para a análise de dados de *CNVs*, *SNPs* e *LOH* no ChAS.

Para avaliação da qualidade dos experimentos de *Cytoscan HD array* são estimadas três medidas recomendadas pelo fabricante, que também fornece os respectivos valores de limiar: (1) *single nucleotide polymorphism quality control*, *SNPQC* ≥ 12 , (2) *median absolute pairwise difference*, *MAPD* $\leq 0,25$ e (3) *waviness standard deviation*, *wavinessSD* $\leq 0,12$. Essas indicam a qualidade dos dados obtidos na determinação dos genótipos das sondas bialélicas e *CN* dos segmentos de DNA. O *SNPQC* é uma estimativa sobre o nível global da qualidade dos dados de *array* a partir de uma perspectiva de *SNPs*. O *MAPD* e o *wavinessSD* são medidas de variação de curto e longo alcance, respectivamente, do conjunto de sondas do *array* em todo o genoma, que consideram a mediana e o desvio padrão dos valores de $\log_2$ ratio par a par e aqueles referentes a toda extensão da faixa de variação.

Para determinação de genótipos das 750 mil sondas de *SNPs* e do *CN* para cada indivíduo, é necessário que os dados sejam normalizados em relação a um genoma de referência. Para tanto, foi utilizado modelo de referência genômica, disponível no *software* ChAS, composto por amostras de indivíduos saudáveis de ambos os sexos: 270 indivíduos da fase I do projeto *HapMap* e 96 indivíduos europeus pertencentes ao banco *BioServe Biotechnologies* (BioServe Biotechnologies, MD, EUA).

5.5 ELABORAÇÃO DE PAINEL DE AIMs

5.5.1 Seleção de SNP-AIMs

Por meio de busca por periódicos em ferramentas de pesquisa: *Pubmed* (<http://www.ncbi.nlm.nih.gov/pubmed>) e *Google Acadêmico* (<http://scholar.google.com>), foi feito levantamento de *SNPs* autossômicos previamente classificados como *AIMs*, capazes de diferenciar populações europeias (EUR), africanas (AFR) e/ou ameríndias (AME). Após a seleção de painéis e exclusão dos *SNP-AIMs* repetidos entre eles, foi verificado quais dos *AIMs* autossômicos estavam presentes dentre os 750 mil *SNPs* do *Cytoscan HD array*. Os *SNP-AIMs* incluídos no *array* foram selecionados para compor o painel de marcadores de ancestralidade desenvolvido nesse trabalho.

5.5.2 Aderência ao equilíbrio de Hardy-Weinberg, diferenciação populacional e composição ancestral

Genótipos de *SNP-AIMs* correspondentes a cada indivíduo foram exportados do *ChAS*. Frequências alélicas e desvios do equilíbrio de Hardy-Weinberg (EHW) foram calculados usando *Plink* v. 1.9 (Purcell et al., 2007). Diferenciação populacional (*estatística F de Wright, F_{st}*) foi calculada a partir do *software GDA* 1.1 (Weir e Cockerham, 1996).

Para realizar as estimativas de ancestralidade, dois tipos de análises foram realizadas: (1) ancestralidade por grupo: a composição ancestral dos grupos de pacientes com LES e controles foi estimada pelo *software Admix 95* (Chakraborty, 1985), baseada nas frequências alélicas de todos os *SNP-AIMs* incluídos no painel, usando as frequências de europeus, africanos e ameríndios como populações parentais de referência, e (2) ancestralidade por indivíduo: a composição ancestral individual baseada no mesmo conjunto de *SNP-AIMs* utilizado para inferência de ancestralidade grupal foi mensurada pelo *software Admixture* v. 1.23 (Alexander et al., 2009), um modelo baseado em algoritmo bayesiano usado para agrupamento de dados genéticos. Para serem usadas como referência para as análises de inferência de ancestralidade individual, as frequências genotípicas das populações parentais europeia e africana foram obtidas do Projeto *HapMap* (CEU – residentes de Utah com ancestralidade do norte e oeste europeu e YRI – Yoruba em Ibadan, Nigéria). Os ameríndios brasileiros foram baseados nas populações Karitiana e Surui do *HGDP-CEPH Human Genome Diversity Cell Line Panel*.

Com o intuito de avaliar se há diferenças entre as proporções ancestrais dos componentes europeu, africano e ameríndio entre casos e controles, o *Wilcoxon Rank sum test with continuity correction* foi aplicado, utilizando as informações obtidas pelo *software Admixture*.

Dados de ancestralidade individual foram também utilizados para comparar ancestralidade determinada por marcadores genéticos no grupo de pacientes com LES com a ancestralidade autodeclarada. Os pacientes se autodeclararam como pertencentes a um dos três grupos: branco, pardo ou preto – termos empregados de acordo com a classificação do IBGE (Instituto Brasileiro de Geografia e Estatística).

5.5.3 Validação do painel de AIMS

Para validar o painel de *SNP-AIMs* desenvolvido nesse trabalho, resultados de composição ancestral foram comparados com outro painel de *AIMs* previamente avaliado no grupo de pacientes com LES (Cagnin, 2015). Tal painel é composto por 12 *AIMs* anteriormente descritos por Shriver et al. (1997) capazes de diferenciar europeus, africanos e/ou ameríndios. Desses marcadores, cinco foram genotipados por PCR-RFLP como *SNPs* (*CKM* rs4884:C>T, *DRD-2* rs1079598:C>T, *FY-Null* rs2814778:A>G, *LPL* rs285:C>T, *RB2300* rs2252544:C>T), sete foram genotipados por PCR convencional, incluindo inserções *Alu* (*Sb19.3* rs3138524, *APO* rs3138522, *PV92* rs3138523) e *InDel* (*AT3-I/D* rs3138521, *MID-52* rs16344, *MID-93* rs16383, *MID-575* rs140864).

A composição ancestral de pacientes com LES foi estimada por meio do *software Admix 95* e depois comparada com os resultados obtidos para o painel de *SNP-AIMs* usando medida de correlação. As frequências alélicas das populações parentais europeia e africana pra o painel de 12 *AIMs* baseou-se nos dados reportados por Shriver et al., 1997. As frequências parentais de ameríndios foram calculadas a partir da genotipagem de 80 ameríndios brasileiros pertencentes aos seguintes grupos: Baniwa (n = 40), Tikuna (n = 16), Kanamari (n = 12) e Kashinawa (n = 12), usando o mesmo conjunto de 12 *AIMs* (Cagnin, 2015).

Um segundo passo para validação do painel de *SNP-AIMs* foi realizado utilizando os 749.157 *SNPs* disponíveis no *Cytoscan HD array*. O *software Plink v2* (Chang et al., 2015) foi utilizado como controle de qualidade e para compilar os dados. O número total de *SNPs* que se mantiveram após o primeiro controle de qualidade usando ChAS

foi de 749157 marcadores autossômicos para 133 indivíduos (amostras de LES e controles brasileiros). Após compilar os dados com as amostras de referência do HapMap/HGDP e filtrar pela taxa de genotipagem (call rate < 99%), *minor allele frequency* (MAF < 0,01), desvios do EHW ($p < 10^{-8}$), desequilíbrio de ligação ($R^2 > 0,8$) o número final de SNPs autossômicos foi de 69399 (~70K).

As estimativas de ancestralidade individual dos componentes europeu, africano e ameríndio determinadas pelo *Admixture* nos pacientes com LES e controles usando os ~70K SNPs disponíveis no *array* foram comparadas com aquelas obtidas com a utilização do subconjunto de SNP-AIMs selecionados para a composição do painel de ancestralidade.

5.5.4 Correção estatística pela estratificação populacional

Para ilustrar como a associação de LES com determinados *loci* pode ser afetada por níveis de mistura populacional, análise estatística com ajuste para ancestralidade obtida pelo uso do painel de SNP-AIMs foi realizada entre casos e controles. Essa metodologia consistiu na seleção de SNP-AIMs ($\delta > 30\%$) adicionais, não incluídos entre aqueles selecionados para constituir o painel de ancestralidade, e de 371 SNPs dentre os 749.157 *Cytoscan HD array*, localizados em dez genes previamente associados com LES: *ETS1*, *FAM167A*, *IRF5*, *JAZF1*, *MECP2*, *PTPN22*, *PXK*, *SCUBE1*, *STAT4*, *XKR6* (Dang et al., 2014; Gateva et al., 2009; Harley ITW et al., 2009; Harley JB et al., 2008).

Os genótipos individuais foram exportados do *ChAS*. As frequências alélicas nos grupos de pacientes e controles foram calculadas e, em seguida, comparadas por Teste Exato de Fisher usando o *Plink*. Aplicando um valor de corte de $p < 0,05$, dez SNPs foram selecionados aleatoriamente para análise de regressão logística com correção para ancestralidade. Os dados de proporções individuais de ancestralidade obtidos pelo *Admixture* foram adicionados como uma variável independente nos modelos de regressão executados no R (<http://www.R-project.org/>). Três modelos independentes com correção para ancestralidade europeia, africana e ameríndia foram ajustados. Com o intuito de identificar associação falso-positivas, os valores de p ajustados para cada componente ancestral obtido pelos modelos de regressão foram usados para comparação com os resultados obtidos a partir dos modelos não ajustados pela ancestralidade.

5.6 ANÁLISE DE DADOS DE CNVs

5.6.1 Dados e filtros

Os dados gerados na determinação de CNVs pelo *ChAS* foram visualizados na forma de gráficos e tabelas. Nos gráficos, são observados $\log_2 \text{ratio}$ e o CN. A cópia diploide esperada é representada pelo $CN = 2$, enquanto que $CN < 2$ corresponde à deleção e $CN > 2$ corresponde à duplicação de segmentos genômicos.

Para cada indivíduo, foi exportada do *ChAS* uma tabela na forma de arquivo com extensão *.txt*, contendo informações relacionadas a cada CNV detectada, tais como: localização genômica exata (cromossomo, citobanda, posição inicial e final), CN, número de marcadores utilizados para determinar cada CNV, o tamanho do fragmento deletado/duplicado, os genes, total ou parcialmente, sobrepostos às CNVs.

Dois tipos de filtros foram aplicados ao dados de CNVs: (1) tamanho: CNVs $\geq 1\text{Kb}$, (2) marcadores: mínimo de 25 e 50 marcadores consecutivos para detectar deleções e duplicações, respectivamente.

5.6.2 Distribuição de CNVs

O *Plink* foi utilizado para sumarizar os dados de CNVs por indivíduo e por grupo. Para avaliar o perfil de variação estrutural no genoma, as CNVs foram divididas em três classes de tamanho: (1) entre 1–10 Kb, (2) entre 10–100 Kb, (3) $> 100\text{Kb}$.

Tanto o número de segmentos, de deleções e de duplicações, quanto o tamanho de CNVs por classe, foi estimado e comparado entre controles e pacientes por meio de regressão logística ajustada para as variáveis sexo e ancestralidade.

Além disso, a distribuição de CNVs por cromossomo foi avaliada e as comparações entre a quantidade e extensão de variantes por cromossomo foram feitas por *teste T de Student*.

5.6.3 Regiões de CNV (CNVRs)

A localização genômica das CNVs recorrentes entre os indivíduos não apresentam, necessariamente, exatos pontos de quebra. Segundo o conceito determinado por

Redon et al (2006), as CNVs presentes em dois ou mais indivíduos sobrepostas entre si são agrupadas em regiões de CNV (CNVRs).

As CNVRs foram definidas por meio de ferramentas de agrupamento de regiões disponíveis no *Plink*. Deleções e duplicações foram analisadas separadamente. Posteriormente, a frequência de cada CNVR identificada foi comparada entre pacientes com LES e controles utilizando Teste Exato de Fisher. A frequência na população saudável de CNVRs provenientes de comparações entre caso-controle que resultaram em $p < 0,05$ foi avaliada usando dados do DGV e do projeto *HapMap*.

5.6.4 CNVs em genes com relevância funcional

Para avaliar CNVs presentes em genes já descritos em associação com LES ou fenótipos autoimunes, foi feito um levantamento de genes previamente associados com LES e com autoimunidade, por meio de pesquisa em periódicos na literatura. Os genes foram categorizados em três listas: (1) genes que sobrepõem CNVs do tipo deleção ou duplicação descritas em associação com o LES, (2) genes previamente associados ao LES em estudos de ligação gênica e/ou GWAS, (3) genes relacionados com autoimunidade de forma geral.

Com a utilização da ferramenta *cytoregion* do ChAS, que permite a busca de regiões ou genes de interesse que sobrepõem deleções ou duplicações, para cada lista de genes foi criada uma *cytoregion* correspondente. Para essa análise, o número de marcadores consecutivos para detecção de deleções e duplicações foi reduzido de 25/50 para 15/15.

Após a obtenção das CNVs que sobrepõem genes candidatos à associação com LES, a frequência da ocorrência de variação englobando o gene em questão foi estimada em indivíduos saudáveis: (1) no grupo controle, (2) em dois grupos externos do projeto *HapMap* (YRI e CEU), (3) no DGV.

5.6.5 CNVs raras

O protocolo para identificação de CNVs raras no grupo de pacientes com LES seguiu quatro etapas. Inicialmente, foi feita a detecção de CNVs com frequência menor do que 1% presentes em pacientes em relação aos controles brasileiros usando o *Plink*.

As CNVs resultantes dessa análise foram comparadas com aquelas identificadas nos indivíduos das populações africana (YRI) e europeia (CEU) do *HapMap*, e então, as CNVs presentes nos indivíduos do *HapMap* foram excluídas. Em seguida, a frequência de tais CNVs foi verificada no DGV, principal banco de dados de variantes estruturais em indivíduos controles; Apenas CNVs catalogadas no DGV com frequência < 1% foram mantidas na análise.

Por fim, com o intuito de aumentar a confiabilidade das CNVs raras detectadas, foi feita uma etapa de validação das CNVs utilizando o *software Nexus Copy Number 8.0* (BioDiscovery, CA, USA). Os arquivos .cel de pacientes com LES foram submetidos à análise por esse *software* alternativo, que utiliza sequência de algoritmos para detecção de CNVs diferente do *ChAS*. Além disso, inclui etapa de correção sistemática dos dados de acordo com o tipo de *array* utilizado. Dessa forma, a lista de CNVs raras resultante após os filtros de frequência populacional foi contrastada com a lista de CNVs detectadas com a utilização do *Nexus Copy Number*. Apenas as CNVs identificadas por ambos *software* permaneceram na análise de CNVs raras.

5.6.6 Seleção de CNVs para validação

Após obtenção das listas de CNVRs, CNVs em genes com relevância funcional e CNVs raras, sete CNVs foram selecionadas para validação por metodologia alvo-específica. Os critérios de seleção aplicados incluíram análise dos genes sobrepostos pelas regiões deletadas e duplicadas, e sua descrição na literatura para avaliar potencial envolvimento na etiologia do LES.

As CNVRs selecionadas foram validadas em maior grupo amostral de pacientes e controles por PCR quantitativa (*qPCR*), enquanto as CNVs em genes com relevância funcional e CNVs raras foram avaliadas por PCR digital (*ddPCR*).

5.7 VALIDAÇÃO DE CNVs POR PCR QUANTITATIVA EM TEMPO REAL (*qPCR*)

5.7.1 Desenho de *primers*

Inicialmente, foi desenhado um par de *primers* para uma região cujo CN fosse, invariavelmente, igual a dois, para ser usado como referência. O critério de seleção para escolha do gene foi baseado em alto grau de conservação entre as espécies,

expressão constitutiva no genoma, ausência de descrição de deleções e duplicações em banco de dados de variantes genômicas, como o *DGV* e *DECIPHER*.

Para o desenho de *primers*, as sequências de interesse foram obtidas em formato *FASTA* a partir do *NCBI* (<http://www.ncbi.nlm.nih.gov/>), versão do genoma *GRCh37/hg19*. O *software* online *Primer3Plus* (Untergasser et al., 2007) foi utilizado para o desenho das sequências de *primers forward* e *reverse*, seguindo os critérios: (1) sequência do fragmento (*amplicon*) entre 70–150 pb de tamanho, (2) conteúdo de bases GC entre 40–60%, (3) tamanho da sequência do *primer* entre 18–23 pb; (4) temperatura de pareamento entre 58–65°C.

Após obtidas as sequências para cada par de *primers*, o *software Gene Runner* (Hastings Software, NY, USA; <http://www.generunner.net/>) foi utilizado para verificar a formação de algum artefato molecular (*hairpin* ou dímero) que pudesse prejudicar a especificidade do mesmo. Na simulação, os *primers* em que houve formação de artefatos em temperatura de melting (T_m) próxima em 10°C à T_m do *primer*, as sequências foram rejeitadas e então, refeitas.

Para verificar a especificidade do *primer*, ou seja, se o *primer* parecia apenas com a região de interesse, ou se pode parear com outras regiões do genoma, foi utilizada a ferramenta *PCR in silico* do *UCSC Genome Browser* (Tyner et al., 2016).

Dispondo da sequência para cada um dos pares de *primers* obtida, os oligonucleotídeos foram sintetizados (Sigma-Aldrich, MO, EUA).

5.7.2 qPCR

Os experimentos de *qPCR* foram realizados no equipamento *StepOnePlus™ Real-Time PCR System* (Applied Biosystems, CA, EUA) para genotipagem de *CNVs* no total de 135 pacientes com LES e 200 indivíduos saudáveis.

O *mix* para *qPCR* foi preparado para o total de 15 µL por reação, contendo as seguintes quantidades: 3,5 µL de água livre de nucleases; 7,5 µL de SybrGreen; 1,0 µL de *primer forward*; 1,0 µL de *primer reverse*; 2,0 µL de DNA genômico na concentração de 25 ng/µL.

A ciclagem padrão de *qPCR* foi: etapa de desnaturação da dupla fita: 95°C por 10 minutos; ciclagem (40X): 95°C por 15 segundos; 60°C por 1 minuto; curva de dissociação: 95°C por 15 segundos; 60°C por 1 minuto; 95°C por 15 segundos.

Na etapa de padronização dos *primers*, foram testadas quatro concentrações para cada par de *primers*: 2,5 µM; 1,25 µM; 0,625 µM; 0,31 µM. A escolha das concentrações ideais foi baseada nos seguintes critérios: (1) ausência de amplificação dos controles negativos na concentração em questão, (2) *cycle threshold* (CT) entre 20–25, (3) ausência de hibridação inespecífica, verificada pela curva de dissociação após a etapa de ciclagem da *qPCR*.

Em seguida, com intuito de elaborar a curva padrão foi feito um gradiente de concentração de DNA (100 ng/µL; 50 ng/µL; 25 ng/µL; 12,5 ng/µL e 6,25 ng/µL), usando um *pool* de DNA de quatro amostras, dois pacientes e dois controles. A eficiência dos *primers* foi determinada a partir da curva padrão de concentração. Valores de eficiência entre 90 e 110% foram considerados adequados.

As reações de *qPCR* para determinar os genótipos da *CNVs* para os alvos de interesse foram realizadas em triplicata de acordo com a concentração ideal de cada *primer*, seguindo a ciclagem e o *mix* especificados anteriormente. Para fins de cálculo do número de cópias, o gene de referência foi genotipado para todas as amostras analisadas. A amostra de referência, que assim como o gene de referência apresenta cópia diploide conhecida ($CN = 2$) para o alvo de interesse, foi repetida em cada placa de 96 poços utilizada nos experimentos de genotipagem. Com o intuito de diminuir o viés de variabilidade técnica entre diferentes reações de PCR, os experimentos foram delineados de modo que a genotipagem dos genes de referência e de interesse para cada amostra foi realizada na mesma placa, além de todas as placas conterem amostras alternadas de pacientes e controles.

5.7.3 Padronização dos experimentos de *qPCR*

O gene *PAX6* foi utilizado como referência para a genotipagem do *ADAM3A*. Os *primers* usados para a genotipagem do gene *FCGR3B* foram descritos por Fanciulli Manuela et al. (2007) (Tabela 2). A concentração ideal utilizada nos ensaios de *qPCR* foi de 2,5 µM para os *primers* *PAX6/FOXP2/FCGR3B* e de 3,5 µM para o *ADAM3A*.

Tabela 2. Sequência de *primers* utilizados na *qPCR*.

Alvo	Sequência <i>forward</i> 5'–3'	Sequência <i>reverse</i> 5'–3'
<i>PAX6</i>	TTGCTTTCTGTGCGGTTGTG	AAGCTGCCAATGACTGAAGG
<i>ADAM3A</i>	AACACTTCAGGACAGCTTAGCC	GTTCCAGAGCTTTGTGAATGG
<i>FOXP2</i>	TGACATGCCAGCTTATCTGTTT	GAGAAAAGCAATTTTCACAGTCC
<i>FCGR3B</i>	CACCTTGAATCTCATCCCCAGGGTCTTG	CCATCTCTGTACCTGCCAG

A curva padrão de concentração do DNA para os quatro pares de *primers* foi obtida conforme esperado, e não houve amplificação dos controles negativos ou desvio padrão significativo entre as replicatas. A eficiência dos *primers* mostrou-se dentro da faixa do esperado, entre 90–110%, e foi considerada suficiente para a realização de ensaios de *qPCR* (Tabela 3).

A amplificação dos produtos de *qPCR* e curva de dissociação dos *primers* mostraram eficiência da reação de *qPCR*, sem amplificação inespecífica de DNA ou formação de dímeros de *primers*.

Tabela 3. Eficiência dos *primers* utilizados na *qPCR*.

Alvo	<i>Slope</i>	<i>Y-Inter</i>	R ²	Eficiência
<i>PAX6</i>	-3,49	26,38	0,67	93,3%
<i>ADAM3A</i>	-3,32	27,48	0,99	99,9%
<i>FOXP2</i>	-3,43	27,76	0,97	95,5%
<i>FCGR3B</i>	-3,51	27,30	0,99	92,4%

5.7.4 Análise de resultados

O número de cópias para cada indivíduo foi calculado a partir do método de quantificação relativa $\Delta\Delta C_t$ (Livak e Schmittgen, 2001). A razão do número de cópias (R) é representada por $R = 2^{-\Delta\Delta C_t}$, de forma que se $R > 1$, o número de cópias do alvo é maior no paciente do que no controle, enquanto que $R < 1$, o número de cópias do alvo é menor no paciente do que no controle.

Após determinação dos genótipos de cada indivíduo do grupo de pacientes e controles, foi feita análise multivariada no R v. 3.1.1 (<http://www.r-project.org/>)

usando modelo de regressão logística ajustado para duas variáveis independentes: sexo e ancestralidade. O sexo foi inserido no modelo como variável qualitativa binária. A ancestralidade, por sua vez, como variável quantitativa contínua, a partir de dados resultantes da aplicação do painel de *AIMs* desenvolvido no presente trabalho. Para os pacientes com LES usados na etapa de validação que não foram genotipados por *Cytoscan HD array*, os dados de ancestralidade individual foram obtidos por meio da aplicação do painel de 12 *AIMs* descrito no tópico 5.5.3.

Para as análises, os indivíduos que apresentaram $CN < 2$ foram agrupados na classe *deleção*, enquanto aqueles que apresentaram $CN > 2$ foram agrupados na classe *duplicação*. Para cada classe foram ajustados três modelos independentes, europeu, africano e ameríndio. Genótipos que apresentaram diferenças entre caso e controles com p ajustado $< 0,05$ e *odds ratio* (OR) $> 1,0$ em um intervalo de confiança a 95% foram considerados significativos.

5.8 VALIDAÇÃO DE CNVs POR PCR DIGITAL (*ddPCR*)

5.8.1 Desenho de *primers*

O desenho dos *primers* usados na *ddPCR* seguiu os mesmos passos descritos no tópico 5.7.1. Como esse experimento inclui etapa de digestão com enzima de restrição, os *amplicons* foram avaliados para a presença de sítios de corte da enzima *HindIII* (AAGCTT).

A lista de *primers* desenhados para *ddPCR* para confirmar a presença da CNV previamente identificada por *Cytoscan HD array* está descrita na Tabela 4.

Tabela 4. Sequência de *primers* utilizados na *ddPCR*.

Alvo	Sequência <i>forward</i> 5'–3'	Sequência <i>reverse</i> 5'–3'
<i>CFHR4</i>	CGATCCAAGTCATCCCTAGAAG	AGGAGGAAACACCTGGAATC
<i>CFHR5</i>	CAAATTCATGTGCACTTTGTGG	CGCCCCATATCTTTGTATTTG
<i>ABCC9</i>	GGAAAACTGGCAGGTGAAAC	TCTCCTCATTCCCTTCACG
<i>HLA-DPB2</i>	TGAAGTTTGATTGCTTCTAGGC	TTACAATCCATTGGTGTAACAGG
<i>STAT4</i>	GCAGTCTGAGTACTCCGTTGAG	TGCCTTAGTTTCGAGGTCAG

5.8.2 ddPCR

Os experimentos de *ddPCR* (Bio-Rad Laboratories, CA, EUA) foram realizados nas instalações da *Bauer Core Facility/FAS Center for Systems Biology* (<http://bauercore.fas.harvard.edu/>), durante o período da bolsa de estágio em pesquisa no exterior no *Department of Environmental Health, Harvard T.H. Chan School of Public Health*, Boston, MA, EUA.

Os passos que precedem as reações de *ddPCR* incluem a digestão do DNA genômico com enzima de restrição. O *mix* para digestão foi preparado para o total de 50 µL por reação, contendo as seguintes quantidades: 41,5 µL de água livre de nucleases; 5 µL de tampão 10X NEBuffer 2.1; 1,0 µL de enzima de restrição *HindIII*; 3,0 µL de DNA genômico na concentração de 70 ng/µL. As amostras foram incubadas a 37°C por uma hora e depois mantidas a -20°C.

Após 16-24 horas da etapa de digestão, o *mix* para as reações de *ddPCR* foi preparado total de 22 µL por reação, contendo as seguintes quantidades: 8,25 µL de água livre de nucleases; 11 µL de *QX200 ddPCR EvaGreen Supermix* (Bio-Rad®); 0,55 µL de *primer mix* (*forward* e *reverse*) a 10 µM; 2,2 µL de DNA genômico digerido com *HindIII* na concentração de 2,5 ng/µL.

O tubo contendo o *mix* é agitado com auxílio de agitador magnético, centrifugado e incubado à temperatura ambiente por 3 minutos. O total de 20 µL do *mix* de reação é transferido para um poço do cartucho *DG8™ Cartridge for QX200 Droplet Generator*, seguido por 70 µL do reagente *QX200 Droplet Generation Oil* para EvaGreen nos poços designados para o óleo.

Após serem gerados os *droplets* no instrumento *QX200 Droplet Generator* (Bio-Rad®), transfere-se lentamente todo o material dos cartuchos para a placa de 96 poços com auxílio de uma pipeta multicanal *P200*. Em seguida, a placa é selada usando *PX1 PCR Plate Sealer* e levada ao termociclador.

A ciclagem padrão de *ddPCR* foi: etapa de ativação da enzima: 95°C por 1 minuto; ciclagem (40X): 95°C por 30 segundos; 60°C por 1 minuto; estabilização do sinal: 4°C por 5 minutos; extensão da dupla fita: 90°C por 5 minutos.

Após a termociclagem, a placa selada é levada ao instrumento *QX200 Droplet Reader* (Bio-Rad®) para leitura dos dados. O *software QuantaSoft™* (Bio-Rad®), acoplado ao

leitor, é utilizado para configurar o *layout* da placa de acordo com o desenho experimental e em seguida, a leitura dos *droplets* é iniciada.

5.8.3 Análise de resultados

Após da aquisição de dados no *software QuantaSoft™*, o *thresholding* aplicado automaticamente aos dados de amplitude 1-D gera os valores de concentração em cópias/ μ L, baseada na distribuição de *Poisson* com intervalo de confiança a 95%, de acordo com a quantidade de *droplets* positivos e negativos em cada reação. Em seguida, calcula-se a razão entre a concentração do alvo e a concentração do gene de referência ($CN = 2$). O resultado é multiplicado por 2 para que se obtenha o valor final do número de cópias referente a organismos diploides.

5.9 ANÁLISE DE DADOS DE LOH

5.9.1 Dados e filtros

Os dados de 750 mil *SNPs* presentes no *Cytoscan HD array* são utilizados para determinação de regiões em variantes consecutivas que se encontram em homozigose no genoma.

Dois tipos de filtros foram aplicados aos dados de *LOH*: (1) filtro de tamanho: segmentos *LOH* > 2 Mb, e (2) mínimo de 50 *SNPs* consecutivos com genótipos homozigotos definir cada segmento *LOH*.

Com o intuito de selecionar apenas *copy-neutral LOH*, os segmentos *LOH* identificados para cada indivíduo foram comparados com as *CNVs* do tipo deleção anteriormente detectadas. Dessa forma, as regiões determinadas como perda de heterozigose que ocorreram concomitantemente a deleções na mesma região genômica foram excluídas do total de segmentos *LOH*.

O total de segmentos em que ocorreu *LOH* foi avaliado individualmente por meio de tabelas individuais geradas pelo *ChAS* em extensão *.txt*, que continham as seguintes informações: (1) localização genômica de cada região de *LOH*, (2) extensão de *LOH* em Mb, (3) número total de sondas bialélicas em homozigose de cada região detectada, (4) genes afetados. A partir desses dados foi possível analisar o perfil de *LOH* nos pacientes e controles para avaliar a distribuição dos segmentos no genoma.

5.9.2 Distribuição e regiões de LOH

Ferramentas do *Plink* foram utilizadas para sumarizar os dados em tabelas contendo o número total/extensão de segmentos *LOH* por indivíduo e por cromossomo.

O número total/extensão de segmentos *LOH* foi comparado entre controles e pacientes por meio de regressão logística ajustada para as variáveis sexo e ancestralidade, enquanto comparações entre a quantidade e extensão de segmentos *LOH* por cromossomo foram feitas por *teste T de Student*.

Considerando pacientes com LES e controles brasileiros, as regiões de *LOH* compartilhadas entre dois ou mais indivíduos foram definidas com a utilização de ferramentas de agrupamento. Dois tipos de localização genômica foram considerados: (1) a totalidade do segmento *LOH*, (2) a fração compartilhada entre todos segmentos *LOH*.

Posteriormente, a frequência dos segmentos *LOH* compartilhados exclusivamente entre pacientes com LES foi verificada em comparação com populações africanas (YRI) e europeias (CEU) do *HapMap*. Com o intuito de adicionar comparação com outros grupos populacionais além do *HapMap*, foi utilizada uma lista de segmentos *LOH* com frequência > 5% na população americana, criada pela Dra. Hutton Kearney, do *Fullerton Genetics Lab* (Asheville, NC, EUA). Tal lista compreende a informação compilada sobre segmentos *LOH* identificados a partir de genotipagem por *Cytoscan HD array*. Os segmentos *LOH* únicos e restritos aos pacientes de LES por sua vez, foram analisados seguindo os mesmos critérios de comparação acima citados.

6 RESULTADOS

6.1 INFERÊNCIA DE ANCESTRALIDADE

6.1.1 Paineis de *SNP-AIMs*

O levantamento de painéis de *SNP-AIMs* capazes de diferenciar os três grupos parentais que compõem a população brasileira (europeus, africanos e ameríndios) resultou na identificação de cinco painéis compostos por 806 *SNP-AIMs* (Galanter et al. 2012; Halder et al. 2008; Kosoy et al. 2009; Lins et al. 2010; Shriver et al. 2003) (Tabela 5). Dentre esses, 345 *SNP-AIMs* estão incluídos dentre os 750 mil *SNPs* pertencentes ao *Cytoscan HD array* e foram selecionados para constituir o painel de marcadores utilizado no presente trabalho. Esse conjunto é capaz de diferenciar os três pares populacionais: europeu-africano (EUR-AFR), africano-ameríndio (AFR-AME), ameríndio-europeu (AME-EUR) (Figura 2).

A identificação e localização genômica dos 345 *SNP-AIMs*, bem como suas frequências nas populações parentais, o diferencial de frequência de cada marcador para cada par populacional e as sondas correspondentes no *Cytoscan HD array*, estão listadas no Apêndice B.

Tabela 5. Quantidade de *SNP-AIMs* que diferenciaram populações europeias, africanas e/ou ameríndias presentes no *Cytoscan HD array*.

Painel	<i>SNP-AIMs</i> no painel	<i>SNP-AIMs</i> no <i>Cytoscan HD array</i>	Referência
1	446	287	Galanter et al., 2012
2	128	32	Kosoy et al., 2008
3	176	19	Halder et al., 2008
4	28	5	Shriver et al., 2003
5	28	2	Lins et al., 2010
Total	806	345

Figura 2. Diagrama de *Venn* mostrando a quantidade de marcadores incluídos no painel de 345 *SNP-AIMs* capazes de diferenciar cada um dos três pares populacionais: ameríndio-europeu (AME-EUR), europeu-africano (EUR-AFR) e africano-ameríndio (AFR-AME).

6.1.2 Diferenciação populacional e composição ancestral

Dentre os 345 *SNP-AIMs*, desvios do EHW ($p < 0,05$) foram encontrados em seis *loci* no grupo de pacientes e em 14 *loci* nos controles (Apêndice B).

Os resultados de diferenciação populacional revelaram que pacientes e controles apresentam diferenciação genética moderada ($F_{ST} = 0,0113$; [IC 95% 0,0081–0,0150]).

A análise de composição ancestral revelou que o principal componente nos grupos de pacientes e controles é europeu ($66,2\% \pm 0,11\%$ e $77,8\% \pm 0,07\%$; $R^2 > 0,99$), seguido de africano ($22,0\% \pm 0,09\%$ e $12,6\% \pm 0,06\%$; $R^2 > 0,99$) e ameríndio ($11,8\% \pm 0,10\%$ e $9,5\% \pm 0,06\%$; $R^2 > 0,99$). Comparações usando as proporções individuais de cada componente ancestral mostraram diferenças significativas entre as ancestralidades europeia ($p = 6,47 \times 10^{-5}$; [IC 95% -0,20– -0,08]) e africana ($p = 1,14 \times 10^{-3}$; [IC 95% 0,03–0,16]) entre casos e controles. Nenhuma diferença significativa foi encontrada em relação à ancestralidade ameríndia ($p = 0,09$; [IC 95% -0,005–0,06]).

Em relação à ancestralidade autodeclarada no grupo de LES, pacientes que se autodeclararam como brancos mostraram, em média, maior contribuição europeia ($64,5\% \pm 14,0\%$). Entretanto, contribuição significativa das ancestralidades africana ($20,3\% \pm 13,2\%$) e ameríndia ($15,5\% \pm 9,5\%$) foram identificadas nesses indivíduos.

Os pacientes autodeclarados como pardos mostraram alta contribuição dos componentes africano ($36,1\% \pm 17,8\%$) e europeu ($60,3\% \pm 14,8\%$). O único paciente

autodeclarado como preto exibiu proporção elevada de ancestralidade africana (52,2%), mas também contribuição europeia (38,9%) significativa.

6.1.3 Validação do painel de 345 *SNP-AIMs*

Valores similares foram observados entre as estimativas de ancestralidade obtidas com a aplicação dos painéis de 345 *SNP-AIMs* e de 12 *AIMs* no grupo de LES. Da mesma forma, a inferência de ancestralidade individual dos componentes europeu, africano e ameríndio utilizando os 70K *SNPs* do *Cytoscan HD array* mostrou alta correlação com as contribuições ancestrais obtidas pelo subconjunto de 345 *SNP-AIMs* (Figura 3).

Figura 3. Estimativa de composição tri-híbrida em grupos de população brasileira. Inferência de ancestralidade grupal (*Admix 95*) nos pacientes com LES mensurada pelos painéis de 345 e 12 *AIMs* (A). Proporções de ancestralidade individual (*Admixture*, $K=3$) nos conjuntos de dados genotípicos de 345 *SNP-AIMs* (B) e de ~70K *SNPs* incluídos no *Cytoscan HD array* (C) nas populações de referência africana (YRI), europeia (CEU) e ameríndia (AME), em casos (LES) e controles brasileiros (BRA). Cada coluna corresponde a um indivíduo. Vermelho, verde e azul representam as estimativas de contribuição africana, europeia e ameríndia, respectivamente. Gráficos de correlação comparando as estimativas de ancestralidade de ~70K *SNPs* e do subconjunto de 345 *AIMs* para os componentes europeu (D), africano (E) e ameríndio (F) em indivíduos do projeto *HapMap* (YRI, $n = 89$; CEU, $n = 89$), HGDP (Karitiana e Surui, $n = 45$), casos (LES, $n = 23$) e controles (BRA, $n = 110$).

6.1.4 Identificação de associações espúrias

Os efeitos do nível de mistura populacional em estudos de associação genética foi observado usando três *SNP-AIMs* que diferenciam europeus e africanos e dez *SNPs* selecionados randomicamente a partir de 371 *SNPs*, cuja associação caso-controle foi significativa ($p < 0,05$ no Teste Exato de Fisher).

Quatro modelos de regressão logística foram aplicados para avaliar o efeito da estratificação populacional devido a mistura: modelo não ajustado e modelos 1, 2 e 3, ajustados respectivamente pelas ancestralidade africana, europeia e ameríndia (Tabela 6). Após ajustar os modelos para avaliar a influência de ancestralidade, foi demonstrado que os *SNP-AIMs* adicionais testados apresentaram associação significativa com o LES no modelo não ajustado, que se tornou não significativa após a correção pela ancestralidade nos modelos europeu e africano, demonstrando a eficiência da correção estatística por esse sistema utilizado. Em relação aos *SNPs* localizados em genes candidatos ao LES, associações significativas se mantiveram para 8/10 *SNPs* (modelo 1: africano), 7/10 *SNPs* (modelo 2: europeu) e 10/10 *SNPs* (modelo 3: ameríndio). Portanto, usando esta abordagem, identificamos 3/10 associações falsas positivas, destacando o efeito da estratificação populacional na análise genética.

Tabela 6. Modelos de regressão logística ajustados para ancestralidade africana, europeia e ameríndia para comparações de *SNP-AIMs* e *SNPs* candidatos à susceptibilidade ao LES.

ID SNP	Tipo	rs dbSNP	Alelo (A1A2)	Modelo não-ajustado		Model 1 (Africano)		Model 2 (Europeu)		Model 3 (Ameríndio)	
				<i>p</i>	OR (IC 95%)	<i>p</i>	OR (IC 95%)	<i>p</i>	OR (IC 95%)	<i>p</i>	OR (IC 95%)
SNP1a	SNP-AIM	rs4926851	CT	0,004	4,4 (1,6 – 12,1)	0,089	2,7 (0,8 – 8,6)	0,109	2,5 (0,8 – 7,7)	0,005	4,3 (1,5 – 11,9)
SNP2a	SNP-AIM	rs11264110	GA	0,019	3,2 (1,3 – 8,9)	0,110	2,3 (0,8 – 6,7)	0,194	2,0 (0,7 – 6,0)	0,030	3,0 (1,1 – 8,4)
SNP3a	SNP-AIM	rs6663966	CT	0,038	2,9 (1,02 – 7,8)	0,409	1,6 (0,5 – 5,1)	0,570	1,4 (0,4 – 4,5)	0,056	2,7 (0,9 – 7,4)
SNP1b	Candidato / ETS1	rs3935795	CT	0,003	5,0 (1,7 – 14,9)	0,007	4,5 (1,5 – 13,7)	0,004	5,3 (1,7 – 16,8)	0,002	6,1 (2,0 – 19,2)
SNP2b	Candidato/ ETS1	rs12277199	CA	0,027	8,1 (1,3 – 64,5)	0,121	4,8 (0,6 – 41,5)	0,094	5,3 (0,7 – 46,0)	0,021	9,0 (1,4 – 73,1)
SNP3b	Candidato/ FAM167A	rs10503422	CT	0,029	2,8 (1,1 – 7,1)	0,017	3,2 (1,2 – 8,9)	0,010	3,7 (1,4 – 10,7)	0,020	3,1 (1,2 – 8,1)
SNP4b	Candidato/ FAM167A	rs7007712	TC	0,033	2,9 (1,1 – 8,0)	0,021	3,3 (1,2 – 9,9)	0,024	3,3 (1,2 – 9,7)	0,037	2,8 (1,1 – 7,9)
SNP5b	Candidato/ PXX	rs59956777	AG	0,001	5,2 (1,9 – 14,6)	0,002	4,9 (1,8 – 14,4)	0,002	5,3 (1,9 – 16,1)	0,001	5,8 (2,1 – 17,2)
SNP6b	Candidato/ JAZF1	rs202	CT	0,019	3,2 (1,3 – 8,9)	0,016	3,5 (1,3 – 10,4)	0,014	3,6 (1,3 – 10,8)	0,018	3,2 (1,2 – 9,0)
SNP7b	Candidato/ JAZF1	rs78276498	AG	0,005	7,4 (1,8 – 32,3)	0,008	7,1 (1,7 – 36,2)	0,013	6,6 (1,5 – 31,2)	0,007	7,0 (1,7 – 30,8)
SNP8b	Candidato/ JAZF1	rs10951192	GC	0,035	2,9 (1,1 – 8,7)	0,031	3,2 (1,2 – 9,8)	0,051	2,8 (0,9 – 8,6)	0,047	2,8 (1,1 – 8,3)
SNP9b	Candidato/ XKR6	rs17725480	TC	0,025	2,9 (1,2 – 7,8)	0,019	3,3 (1,3 – 9,4)	0,011	3,8 (1,4 – 11,3)	0,017	3,3 (1,3 – 9,1)
SNP10b	Candidato/ XKR6	rs17152782	TC	0,013	7,4 (1,5 – 40,4)	0,192	3,3 (0,5 – 21,3)	0,169	3,3 (0,6 – 20,1)	0,010	8,0 (1,6 – 43,9)

ID= identificação; OR = Odds ratio; IC 95% = interval de confiança a 95%; A1: alelo 1; A2: alelo 2.

(a) SNP-AIM não incluído no painel de 345 SNP-AIMs; (b) SNP localizado em gene candidato para susceptibilidade ao LES.

Variável dependente na regressão logística: LES; variável independente: A1; covariável: ancestralidade mensurada usando os 345 SNP-AIMs.

p < 0,05 em itálico.

6.2 VARIAÇÃO NO NÚMERO DE CÓPIAS (CNVs)

6.2.1 Experimentos de Cytoscan HD array

Obtiveram-se valores acima do limiar em todos os controles de qualidade, *SNPQC*, *MAPD* e *wavinessSD*, nas 133 amostras processadas por *Cytoscan HD array*, 23 pacientes e 110 controles. Os valores para cada uma das medidas estão disponíveis no Apêndice C.

6.2.2 Distribuição de CNVs

O total de 447 CNVs nos 23 pacientes com LES e 2652 CNVs nos 110 controles, correspondendo a uma média de 19 ± 10 e 24 ± 11 CNVs por indivíduo, foi identificado nos respectivos grupos. A quantidade de CNVs por indivíduo mostrou-se heterogênea em ambos os grupos, visto que pacientes com LES apresentaram de 5–55 CNVs, enquanto que no grupo controle a amplitude foi ainda mais pronunciada, de 7–74 CNVs (Figura 4, Apêndice C).

Figura 4. Perfil do número total de CNVs no genoma. Histograma mostrando a frequência de cada classe de quantidade de segmentos CNVs em pacientes com LES (A) e controles (B). Estimativa de densidade *kernel gaussiano* (R, *Kernel Density Plot*) mostrando a distribuição não-paramétrica de curva de densidade de probabilidade das classes de frequência de CNVs em pacientes (C) e controles (D).

Considerando o total de CNVs, a maioria concentra-se entre 10–100 Kb tanto em pacientes, quanto em controles, totalizando 61% e 58% (Figura 5). Em geral, tanto pacientes quanto controles apresentam maior quantidade de CNVs do tipo deleção, correspondendo a, respectivamente, 76% e 62% do total de CNVs. Entretanto, o tamanho médio das deleções foi de 3–4 vezes menor, correspondendo a 54 Kb nos pacientes e 67 Kb controles, comparado ao das duplicações, que apresentaram 219 Kb e 216 Kb nos respectivos grupos (Figura 6).

Figura 5. Distribuição do total de CNVs, deleções (del) e duplicações (dup) de acordo com as faixas de tamanho de 1-10 Kb, 10-100 Kb e > 100 Kb nos grupos de pacientes com LES e controles brasileiros (BRA).

Figura 6. Perfil de deleções e duplicações nos grupos de pacientes com LES e indivíduos saudáveis. *Notched box plot* (R, *Box Plot*) mostrando a distribuição da quantidade (A) e do tamanho (B) de deleções e duplicações nos pacientes com LES e em indivíduos saudáveis. As deleções (del) estão identificadas em vermelho, enquanto as duplicações (dup) estão representadas em azul em pacientes com LES e nos controles brasileiros (BRA). As caixas coloridas são a representação do intervalo interquartil, no qual as linhas pretas indicam a mediana dos dados e os círculos vazados sinalizam valores atípicos (*outliers*). Os *notches* exibem intervalo de confiança de 95% em torno da mediana. Se não há sobreposição entre os *notches*, existe evidência de que as diferenças entre os grupos avaliados são significativas (Chambers, 1983), como pode ser observado no número e também no tamanho das duplicações entre casos e controles.

Diferenças significativas mostraram que as duplicações têm maiores frequência ($p = 0,001$; OR = 0,7 [IC 95%, 0,6–0,9]) e tamanho ($p = 0,002$; OR = 1,0 [IC 95%, 1,0–1,0]) em indivíduos saudáveis do que em pacientes com LES. Nenhuma evidência de diferença entre a quantidade ($p = 0,552$; OR = 1,0 [IC 95%, 0,9–1,0]) e tamanho ($p = 0,341$; OR = 1,0 [IC 95%, 1,0–1,0]) de deleções foi observada entre casos e controles.

Em relação à distribuição nos cromossomos, foram encontradas CNVs em todos os cromossomos, exceto no cromossomo Y de pacientes. Em ambos os grupos, o cromossomo X apresentou a maior quantidade de CNVs, representando uma média de 10 ± 4 e 6 ± 5 CNVs por pacientes com LES e controles, respectivamente. Todavia, as CNVs de maior tamanho concentraram-se no cromossomo 14 (Figura 7).

Diferenças significativas tanto na quantidade quanto no tamanho de CNVs entre pacientes e controles foram encontradas nos cromossomos 14 ($p = 0,020$; $p = 0,004$) e 22 ($p = 0,0002$; $p = 0,0002$). Enquanto no cromossomo 14 foi identificado maior número de CNVs de maior tamanho em pacientes, no cromossomo 22, os valores foram mais expressivos no grupo controle.

Diferenças apenas na quantidade de CNVs entre pacientes e controles foram encontradas nos cromossomos 1 ($p = 0,041$), 10 ($p = 0,032$) e 16 ($p = 0,010$). O cromossomo 1 apresentou maior número médio de CNVs por paciente, enquanto a quantidade média de CNVs nos cromossomos 10 e 16 foi maior em indivíduos saudáveis. Em relação ao tamanho médio de CNVs, o cromossomo 11 apresentou valores mais altos em pacientes do que em controles ($p = 0,033$).

Figura 7. Distribuição de CNVs por cromossomo em pacientes com LES. *Violin plot* (pacote R: vioplot v. 0.2) ilustrando a distribuição e a densidade de probabilidade da quantidade (A) e do tamanho (B) de CNVs por cromossomo. As linhas pretas mais grossas no centro representam a escala interquartil, enquanto as linhas pretas mais finas e longas correspondem aos intervalos de confiança a 95% em torno da mediana, esta, por sua vez, ilustrada pelo ponto branco.

6.2.3 CNVRs

Considerando os pacientes e controles, a análise de regiões de *CNV* compartilhadas por dois ou mais indivíduos revelou a presença de 254 *CNVRs* do tipo deleção e 137 *CNVRs* do tipo duplicação.

Na comparação de frequência de cada *CNVR* no grupo de pacientes e controles, obteve-se $p < 0,05$ em 15 *CNVRs*, quatro duplicações e 11 deleções. A maioria englobou regiões gênicas. Do total de *CNVRs*, 11 foram mais frequentes nos pacientes. Quatro delas pertenceram ao cromossomo X, enquanto as demais foram localizadas nos cromossomos 1, 5, 7, 8, 11 e 19. Das três *CNVRs* mais frequentes no grupo controle, foram encontradas duas deleções no cromossomo 1 e uma duplicação no cromossomo 22 (Tabela 7).

Tabela 7. Descrição de CNVRs com frequência significativamente distinta entre pacientes com LES e controles.

Localização genômica	CNVR		Tam (Kb)	Frequência		Teste Exato de Fisher		Freq pop	
	Genes	Tipo		LES	BRA	p	OR [IC 95%]	YRI	CEU
chr1:72762662-72812440	NA	Del	50	0,00	0,16	4,17x10 ⁻²	NA	0,09	0,09
chr1:152761909-152776589	<i>LCE1D</i>	Del	15	0,09	0,30	3,82x10 ⁻²	0,22 [0,0-1,0]	0,31	0,32
chr1:210604732-210612839	<i>HHAT</i>	Del	8	0,09	0,00	2,88x10 ⁻²	NA	0,00	0,00
chr5:180378753-180442428	<i>BTNL3</i>	Del	64	0,39	0,16	2,14x10 ⁻²	3,2 [1,1-9,6]	0,23	0,24
chr7:154720-158615	NA	Del	4	0,22	0,06	3,44x10 ⁻²	4,0 [0,9-16,7]	0,03	0,04
chr8:39213948-39389003	<i>ADAM5, ADAM3A</i>	Del	175	0,83	0,59	3,52x10 ⁻²	3,3 [1,0-14,1]	0,56	0,58
chr11:48658169-48892181	NA	Del	234	0,09	0,00	2,88x10 ⁻²	NA	0,00	0,00
chr11:54701631-55039246	<i>TRIM48</i>	Del	338	0,13	0,01	1,64x10 ⁻²	15,8 [1,2-86,8]	0,01	0,01
chr14:22397988-22970492	NA	Del	573	0,13	0,45	4,65x10 ⁻³	0,2 [0,0-0,7]	0,00	0,00
chr19:41345129-41392187	<i>CYP2A6, CYP2A7</i>	Del	47	0,22	0,05	1,41x10 ⁻²	5,7 [1,2-27,7]	0,07	0,07
chrX:115576235-115589080	<i>SLC6A14</i>	Del	13	0,09	0,00	2,88x10 ⁻²	NA	0,00	0,00
chr22:23151743-23183112	<i>MIR650</i>	Dup	31	0,04	0,55	3,53x10 ⁻⁶	0,03 [0,0-0,2]	0,14	0,14
chrX:356232-387749	NA	Dup	32	0,17	0,02	8,30x10 ⁻³	11 [1,5-13,0]	0,01	0,01
chrX:395646-401509	NA	Dup	6	0,17	0,04	3,01x10 ⁻²	5,5 [0,9-32,1]	0,00	0,00
chrX:1793059-1818776	NA	Dup	26	0,13	0,01	1,64x10 ⁻²	15,8 [1,2-86,8]	0,01	0,01

chr = cromossomo; NA = não apresenta; Del= deleção; Dup = duplicação; Tam = tamanho; OR = odds ratio; IC = intervalo de confiança; BRA = controles brasileiros; YRI = africanos do projeto HapMap; CEU = europeus do projeto HapMap; Freq pop = frequência populacional.

6.2.4 CNVs em genes com relevância funcional

Do levantamento de genes previamente relacionados ao LES descritos na literatura, foram encontrados 12 genes afetados por CNVs descritas em associação com o LES (lista 1), outros 61 genes relacionados especificamente ao LES (lista 2), além de 80 genes relacionados com autoimunidade de forma geral (lista 3). O total de 153 genes com relevância funcional para a patogênese do LES está listado no Apêndice D, juntamente com suas respectivas referências.

A partir da análise da presença de tais genes dentro do intervalo das variantes no número de cópias, foram encontrados 8/153 genes sobrepostos por CNVs nos pacientes com LES. Desses, quatro pertenciam à lista 2: genes relacionados ao LES (*CFHR4*, *CFHR5*, *STAT4*, *MECP2*) e os outros quatro pertenciam à lista 3: genes relacionados com autoimunidade de forma geral (*HLA-DPB2*, *HLA-H*, *ANKS1B*, *ADAM3A*) (Tabela 8).

Tabela 8. Descrição de CNVs presentes em pacientes com LES que sobrepoem genes com relevância funcional para a patogênese do LES.

Localização genômica	CNV			Presença de CNVs (Del/Dup)				
	Genes	Tipo	Tam (Kb)	LES (n=23)	BRA (n=110)	YRI (n=89)	CEU (n=85)	Freq pop
chr1:196827841-196877037	<i>CFHR4</i>	Del	49	1/0	2/1	0/4	0/1	> 1% / > 1%
chr1:196964969-196987806	<i>CFHR5</i>	Del	23	1/0	0/0	1/0	0/0	< 1% / < 1%
chr2:191904170-191921930	<i>STAT4</i>	Del	18	1/0	0/0	0/0	0/0	< 1% / < 1%
chr6:33083147-33095011	<i>HLA-DPB2</i>	Del	12	1/0	0/0	0/0	0/0	< 1% / < 1%
chr6:29837187-29874629	<i>HLA-H</i>	Del	37	1/0	8/4	5/0	2/0	> 1% / > 1%
chr12:99791778-99798721	<i>ANKS1B</i>	Del	7	6/0	25/0	9/0	8/1	> 1% / < 1%
chr8:39247097-39386952	<i>ADAM3A</i>	Del	140	19/0	14/51	9/0	34/10	~50% / ~20%
chrX:153334931-153335831	<i>MECP2</i>	Dup	1	0/7	0/28	0	0	> 1% / > 5%

chr = cromossomo; *Del* = deleção; *Dup* = duplicação; *Tam* = tamanho; *BRA* = controles brasileiros; *YRI* = africanos do projeto HapMap; *CEU* = europeus do projeto HapMap; *Freq pop* = frequência populacional.

6.2.5 CNVs raras

A partir das 447 CNVs detectadas no grupo de pacientes com LES, foram identificadas de 88 CNVs com frequência menor do que 1% comparada àquelas encontradas em controles brasileiros. Após filtragem utilizando dados populacionais disponíveis no projeto *HapMap* e em banco de dados públicos, foram identificadas 49 CNVs raras. Dessas, 21 CNVs foram detectadas por algoritmos de dois *software* diferentes, *ChAS* e *Nexus Copy Number*.

As 21 CNVs raras resultantes após aplicação do protocolo de identificação desenvolvido, consistem em 13 deleções e oito duplicações distribuídas em 13/23 pacientes, que apresentaram de 1–4 CNVs raras por portador de LES (Tabela 9, Figura 8).

Tabela 9. Descrição de CNVs raras identificadas em pacientes com LES a partir do protocolo desenvolvido no presente trabalho.

Localização genômica	Genes	Tipo	Tam (Kb)	ID LES	Freq pop
chr1:196964969-196987806	CFHR5	Del	23	LES005	NA
chr2:34219274-34313413	NA	Del	94	LES009	< 1%
chr2:36932842-36964864	<i>VIT</i>	Del	32	LES010	NA
chr3:65517399-65595008	MAGI1	Dup	78	LES015	< 1%
chr5:96625630-96865984	NA	Del	240	LES014	NA
chr6:152303218-152311895	ESR1	Del	9	LES003	< 1%
chr7:11983316-12078163	NA	Dup	95	LES005	< 1%
chr7:85522989-85709099	NA	Del	186	LES003	< 1%
chr7:114209440-114215739	FOXP2	Del	6	LES023	NA
chr7:146859536-146868357	CNTNAP2	Del	9	LES003	NA
chr8:30079990-30202711	<i>MIR54802</i>	Dup	123	LES018	NA
chr9:137354612-137529641	NA	Dup	175	LES023	< 1%
chr12:21796951-22445614	LDHB, KCNJ8, ABCC9, CMAS, ST8SIA1	Dup	649	LES018	< 1%
chr13:93378273-93386333	GPC5	Del	8	LES023	NA
chr13:94506273-94515199	GPC6, GPC6-AS2	Del	9	LES023	< 1%
chr13:104521177-104647267	NA	Del	126	LES013	NA
chr14:67111441-67148873	GPHN	Dup	37	LES011	NA
chr14:67193022-67217571	GPHN	Del	25	LES016	< 1%
chr22:25066486-25112325	NA	Del	46	LES006	< 1%
chrX:15723836-15831401	CA5B, INE2, ZRSR2	Dup	108	LES018	NA
chrX:52833688-52904643	XAGE5, XAGE3	Dup	71	LES017	NA

chr = cromossomo; NA = não apresenta; Del= deleção; Dup = duplicação; Tam = tamanho; ID = identificação; Freq pop = frequência populacional.

Genes descritos no OMIM destacados em negrito.

Figura 8. Etapas de filtragem de variantes na identificação de CNVs raras em pacientes com LES. O número total de CNVs, incluindo deleções e duplicações, é mostrado em cada círculo referente a um dos passos subsequentes, utilizados na determinação de CNVs raras de acordo com os filtros populacionais aplicados e o uso de diferentes algoritmos na detecção de variantes.

6.2.6 CNVs selecionadas para validação

A partir das três listas resultantes das análises de CNVRs, CNVs em genes com relevância funcional e CNVs raras, foram selecionadas CNVs para validação por metodologia alvo-específica.

De acordo com a frequência do alvo em questão, dois tipos de técnicas de PCR, quantitativa e digital, foram empregadas nessa etapa. A PCR quantitativa foi utilizada para validação de CNVRs, em que a frequência é $> 5\%$ na população, demonstrando a necessidade de genotipagem em maior grupo amostral de casos e controles para avaliar a associação com o LES. Já a PCR digital foi utilizada para confirmação dos genótipos de CNVs raras e CNVs em genes com relevância funcional encontrados em determinados pacientes com LES, presentes em baixa ocorrência, ou até mesmo ausentes na população geral.

A análise que levou à seleção de CNVs para validação foi baseada nos genes sobrepostos pelas CNVs das três listas. O total de 38 genes encontrou-se parcial ou totalmente sobrepostos por CNVs pertencentes às listas de CNVRs, CNVs em genes com potencial relevância funcional e CNVs raras. Os genes *CFHR5* e *ADAM3A*, incluídos na lista de genes com potencial relevância funcional, foram também identificados pela análise de CNVs raras e de CNVRs, respectivamente (Figura 9).

A seleção de CNVs teve como base a literatura a respeito dos 38 genes para avaliar o possível envolvimento na etiologia do LES ou de outras doenças autoimunes, a descrição do gene no banco de dados *Online Mendelian Inheritance in Man (OMIM)* e a frequência na população saudável de variantes no número de cópias envolvendo tal gene. Dessa forma, a análise *in silico* de tais regiões resultou na seleção de sete CNVs para validação por metodologia alvo-específica que sobrepõem os genes: (1) *FCGR3B* e (2) *ADAM3A*, avaliados por *qPCR* e (3) *CFHR4*, (4) *CFHR5*, (5) *STAT4*, (6) *HLA-DPB2* (7) *ABCC9*, genotipados por *ddPCR* (Tabela 10).

Figura 9. Diagrama de Venn destacando os genes que sobrepõem CNVs pertencentes às listas de CNVs raras, CNVRs e CNVs em genes com relevância funcional. Em negrito estão evidenciados os genes em que há variação no número de cópia, cujo genótipo foi avaliado por metodologia alvo-específica.

Tabela 10. Descrição de CNVs selecionadas para validação por metodologia alvo-específica.

Localização genômica	Genes	Tipo	Lista	Método
chr8:39213948-39389003	ADAM5, ADAM3A	Del	CNVR, CNV funcional*	qPCR
chr1:161623196-161631963	FCGR3B	Del	Literatura	qPCR
chr1:196964969-196987806	CFHR5	Del	CNV funcional, CNV rara	ddPCR
chr1:196827841-196877037	CFHR4	Del	CNV funcional	ddPCR
chr2:191904170-191921930	STAT4	Del	CNV funcional	ddPCR
chr6:33083147-33095011	HLA-DPB2	Del	CNV funcional	ddPCR
chr12:21796951-22445614	LDHB, KCNJ8, ABCC9, CMAS, ST8SIA1	Dup	CNV rara	ddPCR

chr = cromossomo; Del= deleção; Dup = duplicação; Tam = tamanho; qPCR = PCR quantitativa; ddPCR = PCR digital.

*CNV em genes com relevância funcional.

6.2.7 CNVs avaliadas por qPCR

Deleção no gene ADAM3A

Na análise de CNVRs realizada a partir de dados de *Cytoscan HD array*, deleção (CN < 2) na região que inclui parcialmente o gene *ADAM5* e inteiramente o gene *ADAM3A* foi identificada como potencial candidata ao aumento de susceptibilidade ao LES ($P = 0,0352$; OR = 3,3; [IC 95% = 1,0–14,1]). Com o intuito de confirmar os genótipos obtidos por *array* e aumentar o número de amostras genotipadas, a CNV nesse gene foi selecionada para validação por qPCR em 134 pacientes com LES e 191 controles.

Deleção no gene FCGR3B

A metodologia *Cytoscan HD array* não apresenta cobertura suficiente para detectar CNVs no gene *FCGR3B*, em que deleções foram previamente associadas com LES em outras populações. Portanto, a determinação do número de cópias nesse alvo foi realizada diretamente por metodologia alvo-específica em 130 pacientes e 139 controles.

A quantificação relativa do número de cópias em cada amostra pelo método $\Delta\Delta C_t$ depende de amostra de referência, a qual apresenta conhecidamente duas cópias da região de interesse. Como não existiram experimentos prévios com o gene *FCGR3B*, a detecção de CNVs nesse alvo incluiu um passo precedente à genotipagem dos indivíduos: a obtenção de amostra de referência por ddPCR. Uma vez encontrada a razão 1:1 entre o gene de referência e o alvo *FCGR3B* por ensaios de ddPCR, a amostra foi utilizada como referência em todos os experimentos de genotipagem por qPCR.

Determinação do número de cópias

No alvo *ADAM3A*, os genótipos dos indivíduos determinados por ambas metodologias, *Cytoscan HD array* e *qPCR*, foram 100% e 90% concordantes nos grupos de casos e controles, respectivamente. Indivíduos com genótipos discordantes foram excluídos das análises estatísticas finais.

O número de cópias do gene *ADAM3A* variou de 0–3 cópias nos pacientes, enquanto no grupo controle variou de 0–8 cópias. Para o gene *FCGR3B*, a variação foi de 0–4 cópias nos pacientes e 0–5 cópias nos controles (Figura 10).

Figura 10. Distribuição de genótipos de CNVs avaliadas por PCR quantitativa. Frequência do número de cópias (CN), de deleções (del) e duplicações (dup) para os genes *ADAM3A* (A, B) e *FCGR3B* (C, D) nos grupos de pacientes com LES e indivíduos saudáveis.

Modelos de regressão logística

Os modelos de regressão logística ajustados para as variáveis sexo e ancestralidade mostraram que deleção no gene *FCGR3B* está associada ao aumento da susceptibilidade ao LES (coeficiente $\beta = 1,3$; $p = 1,66 \times 10^{-3}$, OR = 3,6 [IC 95% 1,7–8,3]), bem como duplicação nesse gene foi descrita como fator de proteção ao desenvolvimento do LES (coeficiente $\beta = -1,6$; $p = 1,55 \times 10^{-4}$, OR = 0,2 [IC 95% 0,1–0,4]).

A associação da deleção no gene *ADAM3A* com LES não foi replicada usando amostra maior (coeficiente $\beta = 0,0$; $p = 0,99$, OR = 1,0 [IC 95% 0,6–1,8]). Por outro lado, duplicação no *ADAM3A* foi descrita como fator de proteção ao desenvolvimento de LES (coeficiente $\beta = -1,4$; $p = 1,23 \times 10^{-2}$, OR = 0,2 [IC 95% 0,1–0,7]).

Quando a presença simultânea da deleção em ambos os *loci FCGR3B e ADAM3A* foi avaliada, observou-se que há um efeito sinérgico entre os *loci* que contribui para o aumento da susceptibilidade ao LES (coeficiente $\beta = 1,8$; $p = 1,4 \times 10^{-4}$, OR = 5,9 [IC 95% 2,5–15,9]).

Os valores apresentados anteriormente correspondem aos modelos de regressão logística ajustados para os valores de ancestralidade individual do componente africano. Os modelos ajustados para os componentes europeu e ameríndio, por sua vez, mostraram resultados similares (Figura 11).

Figura 11. Tendência dose-dependente da presença de variação no número de cópias nos genes *ADAM3A* e *FCGR3B* no desenvolvimento de LES. O risco para desenvolver LES pela quantidade de deleções ($CN < 2$) / duplicações ($CN > 2$) nos genes *ADAM3A*, *FCGR3B* e nos 2 loci foi estimado em relação aos indivíduos sem deleções/duplicações nos determinados loci. Os valores das estimativas pontuais representam *odds ratio* (OR) com intervalo de confiança a 95% nos modelos de regressão logística ajustados para ancestralidade africana (A, D), europeia (B, E) e ameríndia (C, F).

6.2.8 CNVs avaliadas por ddPCR

Os genótipos anteriormente obtidos por *Cytoscan HD array* foram confirmados por *ddPCR* para as 4/5 CNVs selecionadas para validação. Dessa forma, foram descritas três deleções heterozigóticas ($CN = 1$) no genes *CFHR4*, *CFHR5*, *HLA-DPB2*, e uma duplicação heterozigótica ($CN = 3$) que engloba cinco genes: *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* e *ST8SIA11*. A deleção heterozigótica no gene *STAT4* não foi confirmada por *ddPCR* (Figura 12).

Figura 12. Genótipos de CNVs obtidos por PCR digital. Gráficos mostrando o número de cópias (CM) em cada paciente com LES, confirmando deleções heterozigóticas em *CFHR4* no paciente *LES002* (A), em *CFHR5* no paciente *LES005* (B), em *HLA-DPB2* no paciente *LES019* (C), e duplicação heterozigótica envolvendo os genes *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* e *ST8SIA1*, no paciente *LES018* (D). Em todos os casos, o gene de referência (*FOXP2*) mostra *status* diploide invariável nos indivíduos analisados. A amostra utilizada como referência confirma o número de cópias igual a dois para ambos os genes alvo e de referência.

6.3 PERDA DE HETEROZIGOSE

6.3.1 Distribuição de segmentos LOH

A partir da análise de *LOH* nos cromossomos autossômicos, foram identificados 190 segmentos *LOH* em 23 pacientes com LES e 786 em 110 controles. Após verificação de sobreposição dos segmentos *LOH* com regiões deletadas do genoma, foram excluídos 6 e 23 segmentos no grupo de pacientes e no grupo controle, respectivamente, totalizando 184 segmentos *LOH* nos pacientes e 763 nos controles. A distribuição do número de segmentos por indivíduo em ambos os grupos foi assimétrica. Nos controles, o número de segmentos variou de 0–18, enquanto que nos pacientes a amplitude foi maior, variando de 2–30 segmentos por indivíduo (Figura 13, Apêndice C).

Em geral, pacientes apresentaram valores maiores quanto ao número de segmentos *LOH* por indivíduo e extensão de *LOH* no genoma. A média de segmentos por indivíduo correspondeu a 8 ± 6 nos pacientes e 7 ± 3 nos controles. A média de tamanho das regiões de *LOH* foi de 29 Mb nos pacientes e de 20 Mb nos controles. Entretanto, na análise comparativa dos resultados obtidos para pacientes e controles, as diferenças encontradas no número ($p = 0,109$; OR = 1,1 [IC 95%, 1,0–1,3]) e extensão ($p = 0,121$; OR = 1,0 [IC 95%, 1,0–1,0]) de *LOH* não foram significativas.

Figura 13. Perfil do número total de segmentos *LOH* no genoma. Histograma mostrando a frequência de cada classe de quantidade de segmentos *LOH* em pacientes com LES (A) e controles (B). Estimativa de densidade *kernel gaussiano* (R, *Kernel Density Plot*) mostrando a distribuição não-paramétrica de curva de densidade de probabilidade das classes de frequência de segmentos *LOH* em pacientes (C) e controles (D).

Quanto à distribuição de *LOH* por cromossomo, foram encontrados segmentos *LOH* em todos os cromossomos, exceto no 18 e no 22. Dentre todos os cromossomos autossômicos, o cromossomo 16 foi que apresentou a maior quantidade e extensão de segmentos *LOH* no genoma (Figura 14). A quantidade e extensão de segmentos *LOH* nos cromossomos 6 ($p = 0,045$; $p = 0,027$), 12 ($p = 0,008$; $p = 0,013$) foi maior nos pacientes com LES do que em controles. O cromossomo 7, por sua vez, foi significativo apenas para o número de *LOH* por indivíduo ($p = 0,043$), enquanto o cromossomo 5 apresentou maior extensão de segmentos *LOH* ($p = 0,001$). Nos controles, o cromossomo 21 apresentou maior quantidade e extensão de *LOH* do que os pacientes ($p = 0,002$; $p = 0,007$).

A análise de *LOH* no cromossomo X de mulheres do grupo de pacientes e controles identificou 75 segmentos *LOH* em 18 pacientes com LES e 324 em 70 controles, apresentando uma média de 4 ± 3 e 5 ± 2 segmentos por indivíduo. Não houve diferença relevante entre o número de segmentos ($p = 0,897$; OR = 1,0 [IC 95%, 0,8–1,3]) e extensão de *LOH* ($p = 0,979$; OR = 1,0 [IC 95%, 1,0–1,0]) no cromossomo X de pacientes e controles.

Figura 14. Distribuição de segmentos *LOH* por cromossomo em pacientes com LES. Violin plot (pacote R: vioplot v. 0.2) ilustrando a distribuição e a densidade de probabilidade da quantidade (A) e do tamanho (B) de *LOH* por cromossomo. As linhas pretas mais grossas no centro representam a escala interquartil, enquanto as linhas pretas mais finas e longas correspondem aos intervalos de confiança a 95% em torno da mediana, esta, por sua vez, ilustrada pelo ponto branco.

6.3.2 Regiões de *LOH* restritas ao grupo de pacientes com LES

Em relação aos segmentos *LOH* compartilhados por dois ou mais indivíduos, foram encontrados 163 segmentos autossômicos e 86 no cromossomo X (Figura 14). Destes últimos, seis segmentos foram compartilhados exclusivamente por pacientes com LES (Tabela 11).

Tabela 11. Descrição de regiões de *LOH* compartilhadas por mais de um indivíduo, exclusivas do grupo de pacientes com LES.

ID <i>LOH</i>	Localização genômica	Genes	Genes <i>OMIM</i>	Tam (Mb)	ID LES	YRI	CEU	Freq pop
LOH1	chr6:166820898-169656248	35	19	4,1	LES016, LES017	NA	NA	NA
	chr6:167830771-169656248	9	5	1,8				
LOH2	chr9:23539654-40087758	138	72	16,5	LES016, LES017	NA	NA	NA
	chr9:35966698-40087758	35	15	4,1				
LOH3	chr10:45124203-48424759	41	11	3,3	LES005, LES014, LES022	1	8	NA
	chr10:45966421-47597451	24	6	1,6				
LOH4	chr11:57984776-60107747	41	18	2,1	LES001, LES021	NA	NA	NA
	chr11:58001955-60107105	40	18	2,1				
LOH5	chr12:23988752-34835837	58	29	10,8	LES003, LES016	NA	2	NA
	chr12:32674644-34835837	6	5	2,2				
LOH6	chrX: 93531629-96648212	3	2	3,1	LES013, LES015	NA	NA	NA
	chrX: 93754040-95652065	1	0	1,9				

chr = cromossomo; *NA* = não apresenta; *OMIM* = Online Mendelian Inheritance in Man; *Tam* = tamanho; *ID* = identificação; *YRI* = africanos do projeto HapMap; *CEU* = europeus do projeto HapMap; *Freq pop* = frequência populacional.

6.3.3 Segmentos *LOH* restritos ao grupo de pacientes com LES

A partir da identificação de 184 segmentos *LOH* nos cromossomos autossômicos e 75 segmentos *LOH* no cromossomo X de pacientes, foram aplicados filtros subsequentes para determinar quais desses segmentos pertenciam exclusivamente ao grupo de pacientes com LES. Dentre a totalidade de segmentos *LOH* únicos, 17 foram encontrados apenas nos portadores de LES (Figura 15). Tais segmentos estão ausentes em indivíduos pertencentes a grupos externos como o *HapMap* e da lista de *LOH* comuns na população americana. Os segmentos únicos foram encontrados distribuídos em 8/23 pacientes, com a presença variando de 1–6 por indivíduo (Tabela 12).

Tabela 12. Descrição de segmentos *LOH* restritos do grupo de pacientes com LES a partir da comparação com indivíduos saudáveis brasileiros e de outros grupos populacionais.

Localização genômica	Genes	Genes OMIM	Tam (Mb)	ID LES	Freq pop
chr19:48866213-50927358	127	78	2,1	LES002	NA
chr4:92342885-94701185	2	1	2,4	LES006	NA
chr10:85433274-87951821	10	3	2,5	LES011	NA
chr6:103924199-106577507	8	6	2,7	LES014	NA
chr4:183503610-190921709	50	20	7,4	LES016	NA
chr7:27844434-30165896	17	9	2,3	LES016	NA
chr10:7112075-14412084	33	17	7,3	LES016	NA
chr12:130805662-133778166	34	21	3,0	LES016	NA
chr20:3698495-9587591	42	28	5,9	LES016	NA
chr21:14386012-16673128	13	6	2,3	LES016	NA
chr2:221534426-235094377	112	70	13,6	LES017	NA
chr9:7789576-14075095	7	3	6,3	LES017	NA
chr15:22752398-26129249	110	18	3,4	LES017	NA
chr21:26092452-34399401	85	25	8,3	LES017	NA
chr6:144131236-147039527	15	11	2,9	LES019	NA
chr8:141498752-146292734	113	61	4,8	LES022	NA
chr12:64797404-66857938	18	12	2,1	LES022	NA

chr = cromossomo; NA = não apresenta; OMIM = Online Mendelian Inheritance in Man; Tam = tamanho; ID = identificação; Freq pop = frequência populacional.

Figura 15. Etapas de filtragem de segmentos na identificação de *LOH* restritas ao grupo de pacientes com LES. O número total de *LOH* considerando todos os cromossomos autossômicos e o cromossomo X é mostrado em cada círculo referente a um dos passos subsequentes utilizados na determinação de *LOH* exclusivas de pacientes com LES, de acordo com os filtros populacionais aplicados.

6.4 DISTRIBUIÇÃO DE VARIANTES GENÔMICAS NO LES

Com intuito de sumarizar e relacionar os dados de variantes no número de cópias e regiões de perda de heterozigose, a representação circular de tais variantes a partir do *software* *Circos* v. 0.69-4 (Krzywinski et al., 2009) é mostrada na Figura 16. Tal representação evidencia as regiões cromossômicas de alta/baixa densidade de variantes no genoma, como o cromossomo X e o 18, respectivamente. Além disso, mostra a localização genômica de *CNVs* raras e de segmentos *LOH* encontrados apenas em pacientes.

Figura 16. *Circos plot* revelando a distribuição de *CNVs* e *LOH* no genoma de pacientes com LES. Os círculos representam: (a) os 22 cromossomos autossômicos e o cromossomo X segundo versão *GRCh37/hg19* do genoma humano; (b) o total de *CNVs* identificadas nos pacientes – deleções representadas em vermelho e duplicações representadas em azul, (c) as *CNVs* raras encontradas em pacientes com LES; (d) o total de segmentos *LOH* identificados nos pacientes; (e) os segmentos *LOH* restritos aos pacientes com LES.

7 DISCUSSÃO

7.1 PAINEL DE AIMS

A inclusão de *SNPs* com altos valores de alelo de menor frequência (*minor allele frequency; MAF*) dentre o conjunto de marcadores do *Cytoscan HD array* permitiu a identificação de 345 *SNP-AIMs*, previamente relatados como adequados para distinguir as populações europeias, africanas e ameríndias (Galanter et al., 2012; Halder et al., 2008; Kosoy et al., 2009; Lins et al., 2010; Shriver et al., 2003).

Estudos caso-controle de doenças complexas em populações com alto grau de mistura requerem o uso de *AIMs* para evitar associações espúrias (Martin e Eskin, 2016; Shriner et al., 2011). A estimativa de ancestralidade em pacientes com LES e indivíduos saudáveis com base no conjunto de 345 *SNP-AIMs* mostrou que ambos os grupos tiveram maior contribuição ancestral europeia, seguida por africana e ameríndia. Esse achado é consistente com a composição ancestral da população do Sudeste brasileiro que apresenta 60–80% de contribuição europeia, seguida de menores contribuições africana e ameríndia (Egan et al., 2017; Muniz et al., 2008; Pena et al., 2009; Rolim et al., 2016)

Embora em ambos pacientes e controles a contribuição europeia tenha sido a principal, foram detectadas diferenças significativas nas ancestrais europeias e africanas entre os dois grupos. Os pacientes com LES apresentaram maior contribuição africana (22%) do que os indivíduos saudáveis (13%), enquanto os controles mostraram uma contribuição europeia 12% maior do que os pacientes com LES. Essa diferença explica a subestrutura populacional moderada detectada entre os grupos caso-controle. A maior proporção de componente africano no grupo LES parece estar em desacordo com a baixa incidência da doença nas populações africanas (Petri, 2002; Pons-Estel Guillermo J et al., 2010). No entanto, é consistente com relatos anteriores de afro-americanos como mais susceptíveis ao LES do que os euro-americanos (Borchers et al., 2010; Danchenko et al., 2006). Esses resultados enfatizam a diversidade da doença, considerada geneticamente complexa, que apresenta diversas taxas de incidência de acordo com a ancestralidade e apresenta características clínicas e laboratoriais heterogêneas.

As comparações entre a ancestralidades genética e autodeclarada em pacientes com LES apresentaram pelo menos 30% de contribuição ancestral não declarada, incluindo africano/ameríndio em brancos e europeu/ameríndio em indivíduos autodeclarados como pretos, segundo a nomenclatura proposta pelo IBGE. Essa divergência entre a ancestralidade autodeclarada e a ancestralidade genética individual está de acordo com relatos anteriores que descrevem a estrutura genética da população brasileira (Cardena et al., 2013; Lins et al., 2011). Este achado enfatiza a necessidade de avaliar a ancestralidade baseada em marcadores genéticos em vez de usar a percepção dos indivíduos sobre suas características fenotípicas.

Ressalta-se que associações espúrias têm sido documentadas na literatura como efeito da estratificação populacional em estudos de caso-controle em populações que apresentam grau de mistura, como a brasileira (Cardon e Palmer, 2003; Gomes et al., 2017; Shriner et al., 2011; Tang et al., 2005). Nesse contexto, o uso do painel de 345 *AIMs* não se restringe apenas ao presente trabalho. O painel tem potencial para ser amplamente utilizado em pesquisas de genética molecular e citogenéticas para o estudo de doenças cuja incidência é afetada pela ancestralidade, aproveitando os dados do próprio *array*, sem requerimento de experimentos adicionais.

Para avaliar a possível influência da ancestralidade nas associações genéticas, foi demonstrado que a partir do conjunto de *SNPs* selecionados para análise caso-controle, 30% de associações falsas positivas foram encontradas após os ajustes para ancestralidade, o que ressalta a importância de controlar tal fator de confusão nas análises estatísticas.

Nesse sentido, o painel de 345 *SNP-AIMs* foi aplicado nas análises de regressão logística para comparar a distribuição de *CNVs* e *LOH* entre pacientes com LES e indivíduos saudáveis, bem como para avaliar a associação das *CNVs* selecionadas para validação por *qPCR* com o risco para desenvolver LES. Assim, a informação de proporções individuais de cada componente parental foi inserida como variável independente em três modelos de regressão logística independentes ajustados para os componentes europeu, africano e ameríndio. Como os três modelos apresentaram resultados similares, apenas os resultados da regressão ajustada para o modelo europeu foi apresentada na descrição das comparações da distribuição da quantidade e extensão de *CNVs* e de segmentos

LOH, como uma estratégia para resumir e simplificar os achados. Já na apresentação de resultados da associação de *CNVs* com o aumento da susceptibilidade ao LES nos genes *ADAM3A*, *FCGR3B* e para presença da deleção/duplicação em ambos os *loci*, os três modelos foram apresentados (Figura 11).

As associações espúrias resultantes da estratificação populacional podem ser contornadas usando o painel 345 *SNP-AIMs* aqui descrito como um método simples e eficiente para inferir ancestralidade nesse e em outros estudos caso-controle em população brasileira.

7.2 PERFIL DE *CNVs*

No presente trabalho, para que na determinação de *CNVs* a detecção de falsos-positivos fosse minimizada, o número mínimo de marcadores consecutivos para identificar cada *CNV* seguiu as recomendações do fabricante, e portanto, foi maior para duplicações do que para deleções, totalizando 50 e 25, respectivamente. A exigência de um número maior de marcadores para a detecção de duplicações apresenta um viés que é neutralizado quando se considera o fato de que, de forma geral, duplicações envolvem segmentos genômicos maiores do que as deleções e, logo, cobrem regiões que possuem usualmente mais marcadores do que o limite aqui determinado (Choy et al., 2010; Sudmant et al., 2015a). Além disso, mesmo igualando o número de marcadores consecutivos para detectar ganhos e perdas, a proporção encontrada anteriormente foi mantida para ambos os grupos (dados não mostrados).

O perfil de *CNVs*, tanto de pacientes quanto no grupo controle, mostrou que as variantes têm ampla distribuição no genoma. Em geral, as deleções foram três vezes mais frequentes do que as duplicações, entretanto, o tamanho médio dos segmentos deletados foi quatro vezes menor do que o das duplicações. No estudo em larga escala desenvolvido por Kim JH et al. (2013), em que foi analisado o perfil de *CNVs* em mulheres coreanas portadoras de LES, o padrão de deleções/duplicações foi semelhante ao encontrado no presente trabalho. No entanto, as diferenças na razão foram menos expressivas, visto que as deleções foram 1,4 vezes mais frequentes e 1,6 vezes menores do que as duplicações.

Os resultados encontrados em relação ao número e extensão dos eventos de perda/ganho também estão de acordo com o que foi descrito por Zarrei et al. (2015) no mapa de CNVs a partir de meta-análise de dados de indivíduos saudáveis catalogados no DGV.

Em relação à distribuição de CNVs, foi observado excesso de variantes no cromossomo X. Tais resultados foram concordantes com o de trabalhos metodologicamente semelhantes, cuja predominância de CNVs no cromossomo X em relação aos outros cromossomos foi relatada (Kidd et al., 2008; Singh et al., 2014). Embora a carga de CNVs encontrada no cromossomo X não tenha tido diferença significativa entre pacientes e controles, foram encontradas quatro variantes genômicas candidatas ao LES no cromossomo X: deleção em Xq23 (*SLC6A14*) e três duplicações no Xp22.33. A presença de CNVs no cromossomo X sugerem que tais regiões podem estar envolvidas com o desenvolvimento de LES e devem ser melhor caracterizadas para a compreensão do papel de tais genes no desencadeamento e na manutenção da resposta autoimune.

A maioria das imunodeficiências primárias ligadas ao X originam manifestações autoimunes significativas, ilustrando o papel crítico desempenhado por produtos de um único gene situado no cromossomo X, na função e na homeostasia do sistema imune (Bianchi et al., 2012). Adicionalmente, a grande quantidade de genes relacionados ao sistema imune localizados no cromossomo X, que somado à prevalência de doenças autoimunes no sexo feminino, sugerem que alterações no cromossomo X constituem um traço comum na autoimunidade e podem contribuir para a compreensão de sua etiologia (Guerra et al., 2012; Liu KC, 2016). Embora o conhecimento nesse campo tenha se tornado mais refinado, as razões para o predomínio no sexo feminino destas condições ainda não estão claras (Amur et al., 2012; Kukurba et al., 2016).

Mulheres portadoras de diversos tipos de doenças autoimunes apresentam taxas mais elevadas de células circulantes com monossomia do X (Invernizzi et al. 2009). No entanto, tal observação não foi demonstrada em mulheres portadoras de LES (Invernizzi et al., 2007) e relatos da coexistência de LES e síndrome de Turner (45, X0) são extremamente raros (Cooney et al., 2009). Por outro lado, pacientes com síndrome de Klinefelter (47, XXY) apresentam risco semelhante para desenvolver LES em comparação com as mulheres (Liu K et

al., 2016; Scofield et al., 2008). Esses dados sugerem efeito de dose do cromossomo X na susceptibilidade ao LES em que a presença de dois cromossomos X confere maior risco ao desenvolvimento de LES do que apenas uma cópia (Weckerle e Niewold, 2011). Nesse sentido, duplicações de regiões no X podem contribuir para o aumento da susceptibilidade à doença.

Alguns genes de regiões pseudoautosômicas do cromossomo X estão presentes ou têm genes homólogos no Y (Hinch et al., 2014). A dosagem de tais genes é compensada entre os sexos devido a um mecanismo de escape da inativação do cromossomo X (*X chromosome inactivation*, *XCI*), que permite a transcrição em ambos os cromossomos X (Galupa e Heard, 2015). Enquanto desvios no escape de *XCI* foram observados em doenças autoimunes (Özbalkan et al., 2005; Ozcelik et al., 2006), estudos envolvendo ambos modelos murino e humano não encontraram desvios de inativação do cromossomo X no LES (Chitnis et al., 2000). Nossos resultados indicam presença de variantes genômicas que podem estar envolvidas na patogênese do LES e podem ajudar a explicar as razões da prevalência do LES em mulheres, que não é atualmente explicada por mecanismos comuns a outras doenças autoimunes como monossomia do X e desvios do escape de inativação do X.

Há uma série de diferentes mecanismos pelos quais *CNVs* podem ser variantes causais na susceptibilidade e severidade de doenças complexas (Beckmann et al. 2007). Existem, pelo menos, dois modelos distintos, mas não exclusivos, de associação *CNV*-fenótipo. Um modelo envolve *CNVs* de ocorrência comum, também conhecidos como polimorfismos no número de cópias (*copy number polymorphisms*; *CNPs*) (Itsara et al., 2009). Comumente, os *CNPs* relatados em associação com doenças constituem fenótipos multialélicos e são herdadas (Choy et al., 2010; Handsaker et al., 2015; Zuccherato et al., 2017). Genes que sobrepõem *CNPs* são enriquecidos para as funções biológicas associadas com a resposta a drogas, resposta imune, percepção sensorial, entre outras (Almal e Padh, 2012; Cooper et al., 2007; Schaschl et al., 2009). Em princípio, doenças complexas podem ser mais susceptíveis a formas de variação com consequências fenotípicas mais brandas, como *CNVs* em sequências não codificantes, ou em regiões gênicas em que há alteração da dosagem do gene sem abolir completamente a sua função (McCarroll e Altshuler, 2007). Sob esse modelo, variantes comuns que alteram a dosagem de genes ou de outros elementos funcionais, foram relatadas em associação a diversos fenótipos de

doenças complexas. Especificamente em relação ao LES, exemplos de CNVs multialélicas que ocorrem com frequência > 5% nos genes *C4* (Li et al., 2017) e *FCGR3B* (Yuan et al., 2015) foram descritas em associação com o LES em diversas populações.

No presente trabalho, a deleção no gene *FCGR3B* foi associada com aumento do risco para desenvolver LES na população brasileira. Esse achado corrobora com os resultados da meta-análise feita por Yuan et al. (2015) considerando estudos realizados em populações africanas, chinesas e caucasianas. Lee et al. (2015) mostrou que esse locus também está associado com outros fenótipos autoimunes, principalmente granulomatose de Wegener, artrite reumatoide e síndrome de Sjögren.

A ligação de receptores à porção Fc da imunoglobulina G (*Fc gamma receptors*, *FCGRs*) é um mecanismo de resposta imune crucial. A ligação à FCGR pode desencadear respostas biológicas tais como fagocitose e processamento de complexos imunes (Ravetch e Bolland, 2001). Os genes que codificam *FCGRs* são mapeados na região cromossômica 1q21-1q23, e em humanos, há cinco genes que codificam *FCGRs*: *FCGR2A*, *FCGR2B*, *FCGR2C*, *FCGR3A* e *FCGR3B* (Brown et al., 2007).

Foram descritas variantes no número de cópias em três dos cinco genes que codificam *FCGRs*: *FCGR3A*, *FCGR2C* e *FCGR3B* (Breunis et al., 2009). O *FCGR3B* é expresso quase exclusivamente em neutrófilos e atua como sítio de ligação para a imunoglobulina G e complexos imunes. Além disso, *FCGR3B* está envolvido no recrutamento de neutrófilos polimorfonucleares (*polymorphonuclear neutrophils*, *PMNs*) para os locais da inflamação e na depuração de complexos imunes (Mayadas et al., 2009; Nimmerjahn e Ravetch, 2008). Deste modo, deleção no *FCGR3B* pode resultar em redução do tráfego de neutrófilos para lesões inflamatórias e uma diminuição na capacidade de conter complexos imunes (Yuan et al., 2015). Assim, o *FCGR3B* é um gene envolvido com a susceptibilidade a doenças autoimunes com base na sua localização cromossômica e o seu papel funcional.

Adicionalmente, foi observado que na presença simultânea da deleção nos genes *FCGR3B* e *ADAM3A* há um efeito aditivo que contribui para o aumento de 5,9 vezes no risco de desenvolver LES. O efeito sinérgico da deleção em outros três loci: *RABGAP1L*, *C4*, e uma região sem gene no cromossomo 10q21, resultando

no aumento de 5,5 vezes no risco de desenvolver LES foi observado por Kim et al. (2013). Tais resultados evidenciam o efeito aditivo de CNVs na etiologia do LES.

Além desses achados, foi visto que duplicações no gene *ADAM3A*, no cromossomo 8p11.22, cuja função está ligada à atividade pró-inflamatória e sobrevivência dos macrófagos (Yang ZF et al., 2005), conferem proteção ao desenvolvimento de LES. As variantes nesse gene têm frequência > 50% na população saudável e apresentam múltiplos alelos. No presente estudo, foram encontradas tanto duplicações quanto deleções nesse gene, com número de cópias variando de 0–8.

Em relação a outras doenças, Younkin et al. (2015) identificaram uma região de 206 Kb abrangendo os genes *ADAM3A* e *ADAM5*, que mostrou significativa das deleções são transmitidas para os filhos mais frequentemente em trios de pacientes portadores de fendas orofaciais em comparação com os trios de indivíduos saudáveis. Especificamente em relação aos fenótipos autoimunes, Bronstad et al. (2011) descreveram que em pacientes com doença de Addison, a duplicação no *ADAM3A* foi associada com o aumento do risco para desenvolver essa condição. Sugere-se, portanto, que a participação do gene *ADAM3A* apresenta diferente contribuição para o desenvolvimento de doenças autoimunes e ressalta a necessidade de avaliar o papel de CNVs comuns em genes imuno-regulatórios como fatores de risco para o desenvolvimento da autoimunidade.

Em suma, duplicações no gene *ADAM3A* e no gene *FCGR3B* foram descritas aqui como fatores de proteção ao desenvolvimento do LES. Somado ao fato de que a análise comparativa entre casos de LES e controles apontou evidências de que os indivíduos saudáveis apresentam maior número de duplicações em relação aos pacientes com LES ($p = 0,001$; OR = 0,7 [IC 95% 0,6-0,9]), este trabalho mostra o possível envolvimento de ganhos de segmentos de DNA como proteção ao desenvolvimento do LES.

Outras CNVRs que podem estar envolvidas no aumento da susceptibilidade ao LES foram identificadas nos genes *HHAT*, *BTNL3*, *CYP2A6/CYP2A7* a partir da comparação entre casos e controles, embora a avaliação em maior grupo amostral não tenha sido realizada para confirmar a associação com a doença.

Ainda que variantes de ocorrência comum estejam relacionadas a autoimunidade, outras evidências mostram que elas não esclarecem toda a susceptibilidade genética às doenças autoimunes (Bronstad et al., 2011; Craddock et al., 2010). Um segundo modelo que explica associação variante-fenótipo envolve CNVs raras, que apresentam frequência > 1% na população, geralmente surgem por recombinação homóloga não alélica e ocorrem de novo (Green et al., 2016; Itsara et al., 2009; Malhotra e Sebat, 2012; Shi et al., 2016).

Definir o limiar sobre qual o efeito de uma CNV pode estar associada a cada tipo de patologia é um grande desafio (Choy et al., 2010; Palmer et al., 2014; South et al., 2013). No presente trabalho foi estabelecido um protocolo de análise para identificação de CNVs raras. As razões que motivaram avaliar com mais detalhes como é feita a classificação das CNVs raras partiram da dificuldade de inferir a patogenicidade de variantes estruturais encontrada nos pacientes portadores de doenças poligênicas.

A *American College of Medical Genetics (ACMG)* propõe critérios a serem seguidos para a classificação de CNVs em três níveis: benignas, variantes de significado desconhecido e patogênicas (Kearney et al., 2011). Entretanto, tais diretrizes de análise de variantes sejam úteis em um ambiente específico, há dificuldade em aplicar os critérios propostos a todas as doenças (Richards et al., 2015; South et al., 2013). Embora a ACMG recomende não utilizar o tamanho da CNV como critério de exclusão, permitindo assim incluir variantes de qualquer tamanho, sugere-se que as CNVs que não apresentam gene no intervalo, não sejam consideradas na análise. Nesse caso, uma exceção poderia ser considerada nos casos em que a CNV apresenta tamanho atípico. Mesmo assim, em última instância, seria classificada como potencialmente benigna. Seguir tal critério de exclusão não parece ser o mais indicado especificamente em relação ao LES, visto que Kim et al. (2013) relataram CNVR do tipo deleção associada ao LES em uma região do cromossomo 10 que não afeta genes. Logo, mesmo as CNVs que não sobrepõem genes devem ser avaliadas, visto que podem ter função na etiologia do LES. Todavia, diante de outras CNVs que contêm genes previamente relatados como envolvidos na etiologia do LES ou de outros fenótipos autoimunes, parece plausível que essas tenham preferência na seleção de variantes para validação por metodologia alvo-específica.

Outro aspecto da classificação é baseado na análise da frequência da CNV em nível populacional. Atualmente, o DGV é o catálogo mais abrangente de variação estrutural encontrada em indivíduos saudáveis (MacDonald et al., 2014). Em muitos casos, os controles foram definidos em relação a uma determinada doença, logo, outras possíveis condições podem não ter sido descartadas. Outra questão é em relação à definição de tamanho das CNVs, em que muitas se encontram superestimadas devido à baixa resolução das plataformas de detecção usadas inicialmente. Isso motivou a inclusão de comparação das CNVs encontradas no presente trabalho com os grupos externos, como africanos (YRI) e europeus (CEU) do projeto *HapMap*, além do DGV e dos próprios controles brasileiros.

Análise do genoma em larga escala é susceptível ao aparecimento de artefatos técnicos, que ocasionam erros de genotipagem, levando a falsos-positivos. Com o intuito de confirmar os genótipos obtidos, diversos estudos demonstram a necessidade de validação por metodologia alvo-específica (Astring et al., 2014; Bronstad et al., 2011; Ellingford et al., 2017; Mafra et al., 2016).

Para aprimorar a seleção de variantes anterior à etapa de validação, a utilização de mais de um algoritmo para a identificação de CNVs tem sido descrita (Haraksingh et al., 2017; Uddin et al., 2015). Isso tem demonstrado uma maneira eficaz de filtrar as variantes que não configuram artefatos da metodologia utilizada. No presente trabalho, foi utilizado o *software Nexus Copy Number* como algoritmo alternativo na etapa de chamada de variantes raras.

A aplicação do protocolo de análise considerando os aspectos acima explicados, resultou na identificação de 13 deleções e 8 duplicações raras. O genótipo das duas CNVs raras selecionadas para validação foram confirmadas por metodologia altamente específica na quantificação do número de cópias, *ddPCR*. Isso mostra que o uso de dois algoritmos diferentes para seleção de CNVs para validação é um bom preditor para identificar variantes verdadeiras em estudos de *microarray*.

Na análise de CNVs em genes com relevância funcional, o número mínimo de marcadores consecutivos para determinar uma CNV reduzido para 15. Das CNVs selecionadas para etapa de validação, a deleção heterozigótica no gene *STAT4* não foi confirmada por *ddPCR*, e portanto, foi tratada como artefato da técnica de *Cytoscan HD array*. O número de marcadores igual a 16 em que a

deleção no *STAT4* foi identificada, inferior ao mínimo recomendado pelo fabricante na determinação de deleções, que corresponde a 25, pode ter resultado na ausência de concordância entre o genótipo determinado pelo *array* e o genótipo obtido por metodologia alvo-específica. Por outro lado, a deleção no gene *HLA-DPB2* determinada por 18 marcadores consecutivos, bem como aquelas encontradas nos genes *CFHR4*, *CFHR5* identificadas por respectivamente, 60 e 26 marcadores, foram confirmadas.

O complexo principal de histocompatibilidade (*major histocompatibility complex*, *MHC*), também denominado sistema de antígeno leucocitário humano (*human leucocyte antigen*, *HLA*) é uma região localizada no cromossomo 6 em humanos, com alta densidade gênica (Horton et al., 2004). É dividido em três classes: I, II e III. As classes I (*HLA-A*, *-B*, *-C*) e classe II (*HLA-DR*, *-DQ*, *-DP*) codificam as moléculas clássicas apresentadoras de antígenos, enquanto a classe III (*C4A*, *C4B*, *C2*) codifica genes relacionados à processos imunes, tais como os que compõem o sistema complemento, fatores de necrose tumoral e proteínas de choque térmico 70 (Neefjes et al., 2011).

Os genes do *MHC* estão associados com aumento da susceptibilidade a diversas doenças autoimunes, incluindo o LES (Chung et al., 2014; Khatun et al., 2017; Kim K et al., 2016; Schaschl et al., 2009). No presente estudo, relatamos a presença de deleção heterozigótica de 11,9 Kb no pseudogene *HLA-DPB2* no paciente *LES018*. Os alelos *HLA-DPB1* e *HLA-DPB2* apresentam DL relativamente baixo com outros haplótipos do *MHC* e embora não tenham sido amplamente estudados, a função de apresentação de antígenos na ativação de células TCD4+ foi documentada (Sabeti et al., 2002; Zhou X et al., 2009). Adicionalmente, variantes em genes do *MHC* de classe II foram associadas a outras doenças autoimunes, tais como artrite reumatoide (McHugh et al., 2006) e diabetes mellitus tipo 1 (Sanchez et al., 2006), e outras doenças, como esclerose sistêmica (Zhou X et al., 2009), câncer cervical (Chen et al., 2013) e asma (Zhao J et al., 2009).

Especificamente em relação às *CNVs*, duplicação no gene *HLA-DRB5* foi associada ao desenvolvimento de LES em pacientes chineses (Wu et al., 2014). Por outro lado, Kim et al. (2013) relataram *CNVRs* multialélicas com frequência maior que 5% em indivíduos saudáveis e mulheres coreanas portadoras de LES

englobando outros genes pertencentes ao *MHC* de classe II: *HLA-DRB1*, *HLA-DRB5*, *HLA-DRB6*.

No contexto do *MHC*, o componente 4 do complemento está mapeado na região de classe III no cromossomo 6p21.3 e existem dois isotipos de produtos de genes *C4*, *C4A* e *C4B* (Law et al., 1984). *C4* desempenha um papel crucial na ativação de defesa e na depuração de complexos imunes e de detritos apoptóticos (Holers, 2014). A ocorrência de variação estrutural, tal como deficiência fragmentária ou deleção no *C4* pode causar uma resposta imune deficiente aberrante e uma perturbação autoimune ou inflamatória (Mahmoudi et al., 2017; Yih Chen et al., 2016). Deleções no gene *C4* tem sido associadas com o aumento do risco para desenvolver LES e outros fenótipos autoimunes em diversas populações (Li et al., 2017; Liphaus et al., 2015; Saxena et al., 2009).

Levando em consideração a associação já descrita de *CNVs* no *C4*, no *HLA-DRB5* e a deleção no *HLA-DRB2* descrita no presente trabalho, é interessante avaliar de forma mais detalhada se há efeito sinérgico da presença de *CNVs* em genes do *MHC* de classe I e de classe III e como essas variantes podem contribuir para a patogênese do LES.

Adicionalmente, em pacientes com LES, alterações no padrão de ativação de outros genes do complemento leva ao dano inflamatório. A deficiência hereditária dos elementos *C1q*, *C1r*, *C1s*, *C4* ou *C2* da via clássica do complemento prejudica a depuração de complexos imunes e detritos das células apoptóticas, que predispõem à susceptibilidade ao LES (Leffler et al., 2014). Além da flutuação do *C3* sérico ser um biomarcador clínico comumente utilizado para avaliar a atividade da doença (Birmingham et al., 2010), variantes comuns nos genes *C3* também têm sido associadas ao risco de desenvolver LES (Miyagawa et al., 2008; Rhodes et al., 2009). Coletivamente, tais achados indicam o papel do complemento no desenvolvimento do LES.

O fator H do complemento (*complemente fator H*, *CFH*), regulador chave da via alternativa do complemento, desempenha função na modulação de resposta imune inata aos microrganismos, no controle da ativação *C3* e na prevenção de lesão inflamatória ao próprio tecido (Córdoba et al., 2008). Estruturalmente, apresenta agrupamento de cinco genes dispostos em tandem no cromossomo 1q32: *CFH*, *CFHR3*, *CFHR1*, *CFHR4*, *CFHR2* e *CFHR5* (Rodríguez de Córdoba et al., 2004). Zhao J. et al. (2011) descreveram variantes em tais genes em

associação com o LES em populações euro/afro-americanas e asiáticas. Além disso, foi sugerido que deleções envolvendo os genes *CFHR3-CFHR1* podem estar associadas ao LES. Em relação a outras doenças, deleção homo ou heterozigótica em *CFHR3-CFHR1* está associada tanto com o aumento do risco para desenvolver síndrome hemolítica-urêmica atípica (Moore et al., 2010), quanto com diminuição na susceptibilidade à degeneração macular relacionada à idade (Hageman et al., 2006).

No nosso estudo, foram encontradas deleções heterozigóticas nos genes *CFHR4* e *CFHR5*, respectivamente nos pacientes *LES002* e *LES005*. Funcionalmente, o gene *CFHR4* desempenha papel chave na regulação do complemento. A proteína *CFHR4* se liga à proteína C-reativa (*C-reactive protein, CRP*) e a recruta para a superfície das células necróticas, aumentando a remoção das mesmas ou facilitando a ligação e ativação do complemento *C1q* (Mihlan et al., 2009). Tem sido proposto que *CFHR4* limita a inflamação, aumentando a atividade do cofator do *CFH* e a deposição de complemento pela ligação à *CRP*, que auxilia na fagocitose de microrganismos e de células necróticas em tecidos inflamados (Hebecker et al., 2011). Uma vez que *CFHR4* regula a ativação do complemento e a opsonização em superfícies biológicas por meio de interação com *CRP* (Cantsilieris et al., 2012), a potencial deleção desse gene poderia levar à redução na ligação à proteína e assim limitaria a sua capacidade de inibir a inflamação, conduzindo a avanços no desenvolvimento do LES.

A glomerulonefrite C3 familiar tem sido associada com rearranjos que afetam os genes relacionados ao fator H do complemento, que incluem um gene híbrido *CFHR3-1* e uma duplicação interna no gene *CFHR5*. A duplicação de 6,3 Kb englobando os éxons 1 e 2 do gene *CFHR5* causa nefropatia *CFHR5*, transtorno monogênico relacionado à regulação do complemento (Gale e Pickering, 2011). A doença é caracterizada por hematúria, glomerulonefrite e insuficiência renal, o que ressalta o envolvimento da proteína *CFHR5* na regulação do complemento no rim. A condição é herdada como uma característica autossômica dominante com penetrância maior do que 90%, primeiramente descrito como endêmica da população da República do Chipre, ao leste do Mar Mediterrâneo (Deltas et al., 2013; Gale et al., 2010). Medjeral-Thomas et al. (2014) descreveram associação da mesma proteína alterada encontrada na nefropatia *CFHR5* em indivíduos portadores de glomerulonefrite C3 familiar sem ancestralidade cipriota.

O envolvimento de *CFHR5* em doenças renais é de particular interesse tendo em vista que o paciente com LES que apresentou deleção nesse gene apresenta nefrite, o que sugere que perdas nesse gene podem comprometer o funcionamento renal e a indica a possibilidade de tal variação estar relacionada ao desenvolvimento de nefrite lúpica.

Ressalta-se que apesar de *SNPs* nos genes *CFHR4* e *CFHR5* terem sido associados ao LES, deleções em tais genes ainda não haviam sido descritas em pacientes de LES, embora exista relatos da presença da deleção no gene *CFHR4* em banco de dados de variantes genômicas de indivíduos saudáveis (Zarrei et al., 2015). Em relação à deleção no gene *CFHR5*, não há relatos em banco de dados de variantes genômicas da presença de perda de segmento genômico nessa região. Tais achados fornecem evidências de que variantes no número de cópias em genes do fator H do complemento podem ter relação com o LES e pontuam a necessidade de compreender as vias pelas quais as deleções nessas regiões podem atuar na etiologia da doença. Visto que cada gene *CFH-CFHR* representa um componente funcional diferente para a regulação do complemento, sugere-se que o equilíbrio homeostático entre os cinco genes é necessário para manter a regulação da ativação do complemento (Zhao J. et al., 2011).

Existe um espectro contínuo de efeitos fenotípicos da *CNV*, desde características adaptativas, até letalidade embrionária (Wain e Tobin, 2011). O impacto de uma variante estrutural, em particular na expressão de genes sobrepostos pela *CNV*, pode ir além do efeito direto do número de cópias alterado. Por exemplo, o número de cópias do gene *AMY1*, que codifica a amilase salivar, pode chegar até 13 (Carpenter et al., 2015). Duplicações nesse gene geram níveis mais elevados da proteína na saliva, o que pode estar associado a dieta a nível populacional. Isto sugere que cópias extras do gene *AMY1* são expressas levando a concentrações mais elevadas do produto na saliva (Perry et al., 2007). Por outro lado, ganhos no gene *OPN1MW*, envolvido com pigmentação de cor de olho, não parece levar ao aumento da expressão proteica. Aparentemente, as variações fenotípicas na cor dos olhos resultam apenas de mutações de sequência que afetam a primeira cópia de *OPN1MW* (Deeb, 2005).

Especificamente em relação aos genes que já foram associados ao LES, os estudos que correlacionaram variação no número de cópias e o nível de

expressão demonstraram associação entre o deleções no gene *CCL3L1* e a redução da expressão da quimiocina *CCL3L1* (Gonzalez et al., 2005; Townson et al., 2002). De forma semelhante, o número de cópias gene *FCGR3B* correlaciona-se com o nível de expressão da proteína e a capacidade dos neutrófilos de se localizarem em complexos imunes (Willcocks et al., 2008).

Diante do exposto, a forma como a arquitetura genômica contribui para alterar a expressão gênica necessita ser amplamente investigada. Nesse sentido, a análise da expressão seria uma estratégia plausível para avaliar a contribuição das novas *CNVs* descritas nesse trabalho na patogênese do LES.

7.3 PERFIL DE *LOH*

Pesquisas sobre *LOH* começaram a ganhar destaque, como é evidenciado pelo número crescente de publicações após o estudo de Gibson et al. (2006) que descreveu pela primeira vez a abundância de regiões de perda de heterozigose no genoma de populações não endogâmicas. Outros aspectos genéticos de *LOH* em indivíduos saudáveis têm sido investigados, assim como estudos do tipo caso-controle tem sido realizados para identificar *LOH* associadas com traços complexos e doenças (Gamsiz et al., 2013; Ghani et al., 2015; Pippucci et al., 2014).

A distribuição de *LOH* ao longo do genoma não é uniforme, além de apresentar padrões continentais distintos, que pode ser explicado, em parte, pelas propriedades de recombinação e seleção natural (Pemberton et al., 2010). No presente trabalho, foi encontrada correlação positiva entre o número e extensão de segmentos *LOH* por indivíduo em ambos os grupos de pacientes com LES ($p = 0,96$) e controles ($p = 0,76$). Esse fato corrobora com relatos de que o número total de segmentos *LOH* por indivíduo tem padrões semelhantes aos observados para os comprimentos totais dos segmentos (Pemberton et al., 2010). O número médio de *LOH* por indivíduo foi maior nos pacientes com LES do que nos controles (8/7 segmentos), assim como a extensão de *LOH* (29/20 Mb), no entanto, as diferenças não foram estatisticamente significativas. Em relação à distribuição nos cromossomos, foram encontradas evidências de que os pacientes com LES apresentam maior número e extensão de *LOH* nos cromossomos 6 e 12.

A contribuição de *LOH* para o desenvolvimento de doenças tem sido testada em diversas populações, e tem se obtido diferentes resultados. O excesso de *LOH* em relação aos indivíduos saudáveis foi documentado por Lencz et al. (2007) em pacientes com esquizofrenia. No entanto, tal associação não foi replicada quando amostras de 13 países pertencentes ao consórcio *Psychiatric Genomics* foram analisadas (Johnson et al., 2016).

Gamsiz et al. (2013) descreveram a associação do aumento no número e na extensão de blocos de *LOH* com desenvolvimento de autismo em mulheres, quando comparado a indivíduos saudáveis. Usando o mesmo filtro de tamanho aplicado ao presente trabalho (*LOH* > 2 Mb), Ghani et al. (2015) mostrou que a carga de *LOH* é maior em pacientes com doença de Alzheimer em relação aos respectivos controles afro-americanos, o que pode refletir tanto o efeito cumulativo de múltiplas em homozigose na susceptibilidade à doença ou a contribuição de *loci* específicos que abrigam mutações recessivas e haplótipos do risco em um subconjunto dos pacientes. Entretanto, estudos genômicos em larga escala da carga global de *LOH* em populações norte-americanas e europeias não mostraram associação significativa com a doença de Alzheimer (Nalls et al., 2009; Sims et al., 2011).

Em doenças autoimunes, o perfil de perda de heterozigose ainda é pouco estudado, o que dificulta a comparação com outros trabalhos em que o desenvolvimento de fenótipos autoimunes seria o ponto em comum com o LES (Ku et al., 2011). Outra limitação nas análises de *LOH* é a insuficiência de bancos dados disponíveis publicamente que sirvam de referência para comparação com outras populações e entre doenças (Hauck et al., 2013). Na tentativa de contornar esse aspecto, além do grupo de indivíduos brasileiros, populações externas foram adicionadas à critério de comparação com perfil de *LOH* encontrado em pacientes com LES.

Nesse contexto, o desafio é revelar o significado patogênico de eventos múltiplos de *LOH*, que envolvem vários *loci* ao longo do genoma (Pippucci et al., 2014; Szpiech et al., 2013). O caráter exclusivo dos segmentos *LOH* restritos ao grupo de pacientes com LES, tanto os que foram compartilhados por mais de um indivíduo, quanto aqueles encontrados em apenas um paciente, pode ser confirmado diante da comparação com diversos indivíduos saudáveis pertencentes às populações brasileira ($n = 110$), africana ($n = 89$) e europeia (n

= 85) (Tabelas 11 e 12). Mesmo perante ao número reduzido de pacientes com LES (n = 23) em que foram realizadas as análises de perda de heterozigose, a comparação com grupos externos fornece evidências de que a ocorrência de tais segmentos *LOH* é rara na população geral.

A relevância desse tipo de análise é suportada pelo fato de que regiões genômicas em que a ocorrência de *LOH* é rara podem ser enriquecidas para *loci* de função crítica ou *loci* que abrigam variantes recessivas deletérias (Szpiech et al., 2013). Assim, em estudos de mapeamento de homozigose, eventos de *LOH* que ocorrem em regiões incomuns devem ser priorizados. Somado a isto, os padrões de distribuição não uniforme de *LOH* ao longo do genoma em diferentes populações mostram regiões em que eventos de *LOH* são frequentes, conhecidas como *hotspots* ou ilhas de *LOH*, bem como segmentos em que a frequência de regiões em homozigose é muito baixa, *coldspots* ou desertos de *LOH* (Pemberton et al., 2010).

Os *coldspots* de *LOH* representam regiões com maior diversidade e altas taxas de recombinação. Tais regiões são enriquecidas para variantes que em homozigose, tem impacto negativo severo na aptidão e, portanto, há tendência da homozigose não ser mantida, ou regiões que abrigam *loci* em que heterozigotos apresentam vantagem seletiva, favorecendo a alta diversidade haplotípica (Curik et al., 2014; Pemberton et al., 2010).

A partir da análise do perfil de *LOH* em populações da África, Europa, Oriente Médio, Ásia, Oceania e Américas, foram identificados *hotspots* e *coldspots* de *LOH* considerando todas as populações estudadas (N = 64) e também cada subgrupo (Pemberton et al., 2010). O subgrupo correspondente às Américas incluiu cinco populações nativo-americanas do projeto *HGDP-CEPH*: Maya, Pima, Colombian, Karitiana, Surui; e duas populações da fase III do projeto *HapMap*: indivíduos de ancestralidade africana no sudoeste dos EUA (ASW) e mexicanos (MXL). O principal *coldspot* das Américas está localizado no cromossomo 3q13.32, tem 144 Kb inclui os genes *HEG1*, *SLC12A8* e apresenta frequência estimada de 0,043–0,028%. No nosso estudo, o paciente *LES017* apresenta um segmento *LOH* de 23 Mb na região 3q13.32–3q23 que inclui inteiramente a região determinada como *coldspot* e contém 201 genes, 121 deles presentes no *OMIM*.

Segmentos longos de *LOH*, tal como o que foi identificado nesse paciente, são conhecidos por abrigarem maior quantidade de homozigotos deletérios do que seria esperado com base na cobertura total de *LOH* no genoma e na distribuição de homozigotos não deletérios (Szpiech et al., 2013). Somado ao fato de ser um *coldspot*, a ocorrência de *LOH* nessa região sugere potencial patogênico na susceptibilidade ao desenvolvimento de características autoimunes.

Os *hotspots* de *LOH* por sua vez, são regiões de baixa recombinação, que contêm poucos alelos recessivos deletérios e, assim, são capazes de manter a homozigose, também podem representar regiões que abrigam alvos de seleção positiva, que experimentaram redução global na diversidade genética e aumento na homozigose em torno dos *loci* selecionados (Pemberton et al., 2010).

No presente trabalho, foi encontrado um *hotspot* de *LOH* no cromossomo 16p11.2–16p11.1 em que a frequência é superior a 70% nos pacientes com LES (96%) e controles brasileiros (75%). Em europeus, a frequência é alta (94%), já em africanos, corresponde a 43%. Como, em parte, a formação de *hotspots* é decorrente de seleção positiva em determinadas regiões, por conseguinte, outros *hotspots* identificados, como o descrito nesse trabalho, podem fornecer evidências de seleção positiva recente e sugerem que a seleção natural desempenha papel na modulação dos padrões de homozigose genômica.

Especificamente em relação ao LES, Singh et al. (2014) mostraram que em população anormal de neutrófilos, os *LDGs*, pacientes com LES apresentam aumento da frequência de *LOH*, além de compartilharem um segmento *LOH* na região 5q23.3–5q31.1 entre 4/13 pacientes analisados no estudo, que contém diversos genes envolvidos com regulação da resposta imune. No nosso trabalho, o paciente *LES023* apresenta *LOH* na região 5q23.3–5q31.1. Entretanto, esse segmento não foi incluído na lista de *LOH* restritas ao grupo de pacientes com LES por estarem presentes em 9/110 controles brasileiros, 1/89 africanos, 14/85 europeus.

A análise de *LOH* aqui mostrada tem relevância por se tratar de dados novos, cujo assunto é pouco explorado na literatura. A avaliação futura dos genes, bem como a predição de patogenicidade das variantes encontradas em homozigose, poderão contribuir para o entendimento de como as regiões de perda de heterozigose influenciam a etiologia do LES.

8 CONCLUSÕES

Os resultados aqui descritos evidenciam que novos *loci* de susceptibilidade ao LES podem ser encontrados quando se estuda outras populações, principalmente as com grau de mistura, como a brasileira.

O painel de 345 *SNP-AIMs*, elaborado no presente trabalho a partir de marcadores presentes no *Cytoscan HD array*, mostrou-se eficiente para inferência de ancestralidade tri-híbrida em amostras de caso-controle em população brasileira. As proporções de composição ancestral mostraram que ambos os grupos de pacientes com LES e indivíduos saudáveis possuem maior contribuição europeia, seguida de africana e ameríndia. Embora o componente europeu tenha sido o mais representativo em ambos os grupos, diferenças significativas nas ancestralidades europeia e africana foram detectadas entre pacientes com LES e indivíduos saudáveis, o que explica a subestrutura detectada entre os grupos na análise de diferenciação populacional.

O perfil de *CNVs* mostrou que grande parte das variantes identificadas possui entre 10–100 Kb de tamanho e encontra-se distribuídas na maioria dos cromossomos, embora a sua distribuição no genoma não seja uniforme. As deleções são mais frequentes do que as duplicações, entretanto, os ganhos envolvem segmentos maiores no genoma. Evidências de que os indivíduos saudáveis apresentam maior quantidade e tamanho de duplicações do que os pacientes com LES foram encontradas, o que sugere possível papel contributivo da carga total de duplicações na proteção ao desenvolvimento da doença.

O protocolo desenvolvido nesse trabalho para identificação de variantes raras, que incluiu comparação com indivíduos saudáveis brasileiros, de grupos externos, de banco de dados, e análise de dados por algoritmo alternativo, resultou na identificação de 21 *CNVs* raras em pacientes com LES descritas pela primeira vez.

A associação da deleção no *FCGR3B* com o LES foi descrita, enquanto duplicações nesse gene e no *ADAM3A* foram relatadas como fator de proteção ao desenvolvimento da doença. Além disso, mostrou-se o efeito sinérgico da presença da deleção em ambos os *loci* *FCGR3B* e *ADAM3A* no aumento do risco

para desenvolver LES. Deleções em pacientes com LES envolvendo os genes *CFHR4*, *CFHR5* e *HLA-DPB2*, previamente descritos em associação com o LES na literatura, foram identificadas e confirmadas por metodologia independente, cuja investigação de sua relação com a patogênese pode contribuir para a compreensão da base genética do LES.

Em relação ao perfil de *LOH* considerando o genoma total, não foram encontradas evidências de que o número médio e a extensão dos segmentos seja diferente entre pacientes e indivíduos saudáveis. No entanto, pacientes com LES apresentam maior quantidade e maior tamanho de segmentos *LOH* nos cromossomos 6 e 12 quando comparados ao grupo controle. A análise de *LOH* permitiu a identificação de um conjunto de segmentos *LOH* restrito ao grupo de pacientes com LES, em comparação com os indivíduos saudáveis da população brasileira e de outros grupos externos.

9 REFERÊNCIAS BIBLIOGRÁFICAS

- Alarcón-Riquelme ME, Ziegler JT, Molineros J, Howard TD, Moreno-Estrada A, Sanchez-Rodriguez E, et al. Genome-wide association study in an Amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of European admixture. *Arthritis & Rheumatism*. 2016 Apr;68(4):932-43.
- Alarcón-Segovia D, Alarcón-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis & Rheumatism*. 2005;52(4):1138-47.
- Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*. 2009;19(9):1655-64.
- Almal SH, Padh H. Implications of gene copy-number variation in health and diseases. *Journal of Human Genetics*. 2012;57(1):6-13.
- Amur S, Parekh A, Mummaneni P. Sex differences and genomics in autoimmune diseases. *Journal of Autoimmunity*. 2012.
- Armstrong D, Zidovetzki R, Alarcón-Riquelme M, Tsao B, Criswell L, Kimberly R, et al. GWAS identifies novel SLE susceptibility genes and explains the association of the HLA region. *Genes and Immunity*. 2014.
- Asting AG, Lagerstedt KK, Kristiansson E, Lönnroth C, Andersson M, Rekabdar E, et al. Genetic copy number variations in colon mucosa indicating risk for colorectal cancer. *Journal of Cancer Therapy*. 2014;5(14):1354.
- Barbosa FB. CNVs em pacientes com lúpus eritematoso sistêmico [dissertação]: Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo; 2013.
- Bennett RJ, Forche A, Berman J. Rapid mechanisms for generating genome diversity: whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harbor Perspectives in Medicine*. 2014;4(10):a019604.
- Bi JH, Qiu ZW, Gazdar AF, Song K. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer. *Genes, Chromosomes and Cancer*. 2017.
- Bianchi I, Lleo A, Gershwin ME, Invernizzi P. The X chromosome and immune associated genes. *Journal of Autoimmunity*. 2012;38(2):J187-J92.
- Birmingham D, Irshaid F, Nagaraja H, Zou X, Tsao B, Wu H, et al. The complex nature of serum C3 and C4 as biomarkers of lupus renal flare. *Lupus*. 2010.
- Birney E, Smith GD, Greally JM. Epigenome-wide association studies and the interpretation of disease-omics. *PLoS Genetics*. 2016;12(6):e1006105.

- Bocquet B, Marzouka NaD, Hebrard M, Manes G, Sénéchal A, Meunier I, et al. Homozygosity mapping in autosomal recessive retinitis pigmentosa families detects novel mutations. 2013; (19):2487-500.
- Borchers AT, Naguwa SM, Shoenfeld Y, Gershwin ME. The geoepidemiology of systemic lupus erythematosus. *Autoimmunity Reviews*. 2010;9(5):A277-A87.
- Bouts YM, Wolthuis DFGJ, Dirkx MFM, Pieterse E, Simons EMF, Van Boekel AM, et al. Apoptosis and NET formation in the pathogenesis of SLE. *Autoimmunity*. 2012;(8):597-601.
- Breunis WB, van Mirre E, Geissler J, Laddach N, Wolbink G, van der Schoot E, et al. Copy number variation at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B. *Human Mutation*. 2009;30(5):E640-E50.
- Bronstad I, Wolff AS, Lovas K, Knappskog PM, Husebye ES. Genome-wide copy number variation (CNV) in patients with autoimmune Addison's disease. *BMC Med Genet*. 2011;12:111.
- Brown EE, Edberg JC, Kimberly RP. Fc receptor genes and the systemic lupus erythematosus diathesis. *Autoimmunity*. 2007 Dec;40(8):567-81.
- Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194(2):459-71.
- Cagnin NF. SNPs do gene MC1R e a susceptibilidade ao lúpus eritematoso sistêmico [tese]: Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo; 2015.
- Callegari-Jacques SM, Salzano FM. Brazilian Indian/non Indian interactions and their effects. *Ciência e Cultura (São Paulo)*. 1999;51(3/4):166-74.
- Cantsilieris S, White SJ, Richardson AJ, Guymer RH, Baird PN. Comprehensive analysis of copy number variation of genes at chromosome 1 and 10 loci associated with late age related macular degeneration. *PloS One*. 2012;7(4):e35255.
- Cardena MM, Ribeiro-dos-Santos A, Santos S, Mansur AJ, Pereira AC, Fridman C. Assessment of the relationship between self-declared ethnicity, mitochondrial haplogroups and genomic ancestry in Brazilian individuals. 2013;8(4):e62005.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *The Lancet*. 2003;361(9357):598-604.
- Carpenter D, Dhar S, Mitchell LM, Fu B, Tyson J, Shwan NA, et al. Obesity, starch digestion and amylase: association between copy number variants at human salivary (AMY1) and pancreatic (AMY2) amylase genes. *Human Molecular Genetics*. 2015;24(12):3472-80.
- Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nature Reviews Genetics*. 2016 Apr;17(4):224-38.
- Carvalho CM, Pfundt R, King DA, Lindsay SJ, Zuccherato LW, Macville MV, et al. Absence of heterozygosity due to template switching during replicative

rearrangements. *The American Journal of Human Genetics*. 2015 Apr 02;96(4):555-64.

Cerosaletti K, Habib T, James R, Chen J, Pickett M, Funk A, et al. Impact of genetic variants on B-cell development and function in systemic lupus erythematosus. *Arthritis Research & Therapy*. 2014;16(Suppl 1):A8.

Chakraborty R. Genetic microdifferentiation in human and others animal population. *Indian Anthropological Association*. 1985:171 – 80.

Chambers JM. Graphical methods for data analysis. 1983.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM & Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4(7).

Chen D, Juko-Pecirep I, Hammer J, Ivansson E, Enroth S, Gustavsson I, et al. Genome-wide association study of susceptibility loci for cervical cancer. *Journal of the National Cancer Institute*. 2013;105(9):624-33.

Chitnis S, Monterio J, Glass D, Apatoff B, Salmon J, Concannon P, et al. The role of X-chromosome inactivation in female predisposition to autoimmunity. *Arthritis Research*. 2000;2(5):399-406.

Choy KW, Setlur SR, Lee C, Lau TK. The impact of human copy number variation on a new era of genetic testing. *BJOG: An International Journal of Obstetrics and Gynaecology*. 2010 Mar;117(4):391-8.

Christofidou P, Nelson CP, Nikpay M, Qu L, Li M, Loley C, et al. Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. *The American Journal of Human Genetics*. 2015;97(2):228-37.

Chung SA, Brown EE, Williams AH, Ramos PS, Berthier CC, Bhangale T, et al. Lupus nephritis susceptibility Loci in women with systemic lupus erythematosus. *Journal of the American Society of Nephrology*. 2014;25(12):2859-70.

Cleynen I, Konings P, Robberecht C, Laukens D, Amininejad L, Théâtre E, et al. Genome-wide copy number variation scan identifies complement component C4 as novel susceptibility gene for Crohn's disease. *Inflammatory Bowel Diseases*. 2016;22(3):505-15.

Coelho A, Moura R, Cavalcanti C, Guimarães R, Sandrin-Garcia P, Crovella S, et al. A rapid screening of ancestry for genetic association studies in an admixed population from Pernambuco, Brazil. *Genetics and Molecular Research*. 2015;14(1):2876.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, et al. Origins and functional impact of copy number variation in the human genome. *Nature*. 2009;464(7289):704-12.

Cooney CM, Bruner GR, Aberle T, Namjou-Khales B, Myers LK, Feo L, et al. 46, X, del (X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE. *Genes and Immunity*. 2009;10(5):478-81.

- Coop G, Przeworski M. An evolutionary view of human recombination. *Nature Reviews Genetics*. 2007;8(1):23-34.
- Cooper GM, Nickerson DA, Eichler EE. Mutational and selective effects on copy-number variants in the human genome. *Nature Genetics*. 2007;39:S22-S9.
- Córdoba D, Rodriguez S, De Jorge EG. Translational mini-review series on complement factor H: Genetics and disease associations of human complement factor H. *Clinical & Experimental Immunology*. 2008;151(1):1-13.
- Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010 [cited 464 7289]; 713-20].
- Cui Y, Sheng Y, Zhang X. Genetic susceptibility to SLE: recent progress from GWAS. *Journal of Autoimmunity*. 2013;41:25-33.
- Curik I, Ferenčaković M, Sölkner J. Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*. 2014;166:26-34.
- Curtin PD. *The Atlantic slave trade: a census*: Univ of Wisconsin Press 1972.
- Curtis D. Extended homozygosity is not usually due to cytogenetic abnormality. *BMC Genetics*. 2007;8(1):67.
- Danchenko N, Satia J, Anthony M. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus*. 2006;15(5):308-18.
- Dang J, Shan S, Li J, Zhao H, Xin Q, Liu Y, et al. Gene-gene interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. *Tissue Antigens*. 2014;83(6):401-8.
- de Andrade Ramos BR, Mendes ND, Tanikawa AA, Amador MAT, dos Santos NPC, dos Santos SEB, et al. Ancestry informative markers and selected single nucleotide polymorphisms in immunoregulatory genes on preterm labor and preterm premature rupture of membranes: a case control study. *BMC Pregnancy and Childbirth*. 2016;16(1):30.
- Deeb S. The molecular basis of variation in human color vision. *Clinical Genetics*. 2005;67(5):369-77.
- Deltas C, Gale D, Cook T, Voskarides K, Athanasiou Y, Pierides A. C3 glomerulonephritis/CFHR5 nephropathy is an endemic disease in Cyprus: clinical and molecular findings in 21 families. *Complement Therapeutics*: Springer 2013:189-96.
- Deng Y, Tsao BP. Genetic susceptibility to systemic lupus erythematosus in the genomic era. *Nature Reviews Rheumatology*. 2010;6(12):683-92.
- Di Gaetano C, Fiorito G, Ortu MF, Rosa F, Guarrera S, Pardini B, et al. Sardinians genetic background explained by runs of homozygosity and genomic regions under positive selection. *PloS One*. 2014;9(3):e91237.

Dorner T, Giesecke C, Lipsky PE. Mechanisms of B cell autoimmunity in SLE. *Arthritis Research & Therapy*. 2010;13(5):243.

Duan J, Zhang J-G, Deng H-W, Wang Y-P. Comparative studies of copy number variation detection methods for next-generation sequencing technologies. *PloS One*. 2013;8(3):e59128.

Egan KJ, Campos Santos H, Beijamini F, Duarte NE, Horimoto AR, Taporoski TP, et al. Amerindian (but not African or European) ancestry is significantly associated with diurnal preference within an admixed Brazilian population. *Chronobiology International*. 2017:1-4.

Ellingford JM, Campbell C, Barton S, Bhaskar S, Gupta S, Taylor RL, et al. Validation of copy number variation analysis for next-generation sequencing diagnostics. *European Journal of Human Genetics*. 2017.

Enciso-Mora V, Hosking FJ, Houlston RS. Risk of breast and prostate cancer is not associated with increased homozygosity in outbred populations. *European Journal of Human Genetics*. 2010;18(8):909-14.

Fanciulli M, Norsworthy PJ, Petretto E, Dong R, Harper L, Kamesh L, et al. FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nature Genetics*. 2007;39(6):721-3.

Fanciulli M, Petretto E, Aitman T. Gene copy number variation and common human disease. *Clinical Genetics*. 2010;77(3):201-13.

Farhi A, Cohen AD, Shovman O, Comaneshter D, Amital H, Amital D. Bipolar disorder associated with rheumatoid arthritis: A case-control study. *Journal of Affective Disorders*. 2016;189:287-9.

Fortuna G, Brennan MT. Systemic lupus erythematosus: epidemiology, pathophysiology, manifestations, and management. *Dental Clinics of North America*. 2013 Oct;57(4):631-55.

Frieri M. Mechanisms of disease for the clinician: systemic lupus erythematosus. *Annals of Allergy, Asthma & Immunology*. 2013 Apr;110(4):228-32.

Galanter JM, Fernandez-Lopez JC, Gignoux CR, Barnholtz-Sloan J, Fernandez-Rozadilla C, Via M, et al. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genetics*. 2012;8(3):e1002554.

Gale DP, De Jorge EG, Cook HT, Martinez-Barricarte R, Hadjisavvas A, McLean AG, et al. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. *The Lancet*. 2010;376(9743):794-801.

Gale DP, Pickering MC. Regulating complement in the kidney: insights from CFHR5 nephropathy. *Disease Models & Mechanisms*. 2011;4(6):721-6.

Galupa R, Heard E. X-chromosome inactivation: new insights into cis and trans regulation. *Current Opinion in Genetics & Development*. 2015;31:57-66.

Gamazon ER, Stranger BE. The impact of human copy number variation on gene expression. *Briefings in Functional Genomics*. 2015 Sep;14(5):352-7.

Gamsiz ED, Viscidi EW, Frederick AM, Nagpal S, Sanders SJ, Murtha MT, et al. Intellectual disability is associated with increased runs of homozygosity in simplex autism. *The American Journal of Human Genetics*. 2013 Jul 11;93(1):103-9.

Gandin I, Faletra F, Faletra F, Carella M, Pecile V, Ferrero GB, et al. Excess of runs of homozygosity is associated with severe cognitive impairment in intellectual disability. *Genetics in Medicine*. 2015 May;17(5):396-9.

Garcia-Ortiz H, Velazquez-Cruz R, Espinosa-Rosales F, Jimenez-Morales S, Baca V, Orozco L. Association of TLR7 copy number variation with susceptibility to childhood-onset systemic lupus erythematosus in Mexican population. *Annals of the Rheumatic Diseases*. 2010 Oct;69(10):1861-5.

Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature Genetics*. 2009;41(11):1228-33.

Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, et al. Association of long runs of homozygosity with Alzheimer disease among african american individuals. *JAMA Neurology*. 2015 Nov;72(11):1313-23.

Gibson J, Morton NE, Collins A. Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*. 2006 Mar 01;15(5):789-95.

Gomes KFB, Santos AS, Semzezem C, Correia MR, Brito LA, Ruiz MO, et al. The influence of population stratification on genetic markers associated with type 1 diabetes. *Scientific Reports*. 2017;7.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science Signalling*. 2005;307(5714):1434.

Green E, Rees E, Walters JTR, Smith K, Forty L, Grozeva D, et al. Copy number variation in bipolar disorder. *Molecular Psychiatry*. 2016;21(1):89-93.

Gross A, Tönjes A, Kovacs P, Veeramah KR, Ahnert P, Roshyara NR, et al. Population-genetic comparison of the Sorbian isolate population in Germany with the German KORA population using genome-wide SNP arrays. *BMC Genetics*. 2011;12(1):67.

Guerra SG, Vyse TJ, Graham DSC. The genetics of lupus: a functional perspective. *Arthritis Research & Therapy*. 2012;(14)2:3.

Hageman GS, Hancox LS, Taiber AJ, Gehrs KM, Anderson DH, Johnson LV, et al. Extended haplotypes in the complement factor H (CFH) and CFH-related (CFHR) family of genes protect against age-related macular degeneration: Characterization, ethnic distribution and evolutionary implications. *Annals of Medicine*. 2006;38(8):592-604.

Halder I, Shriver M, Thomas M, Fernandez JR, Frudakis T. A panel of ancestry informative markers for estimating individual biogeographical ancestry and

admixture from four continents: utility and applications. *Human Mutation*. 2008;29(5):648-58.

Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, et al. Large multiallelic copy number variations in humans. *Nature Genetics*. 2015 Mar;47(3):296-303.

Haraksingh RR, Abyzov A, Urban AE. Comprehensive performance comparison of high-resolution array platforms for genome-wide Copy Number Variation (CNV) analysis in humans. *BMC Genomics*. 2017;18(1):321.

Harley ITW, Kaufman KM, Langefeld CD, Harley JB, Kelly JA. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nature Reviews Genetics*. 2009;10(5):285-90.

Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nature Genetics*. 2008;40(2):204-10.

Harville HM, Held S, Diaz-Font A, Davis EE, Diplas BH, Lewis RA, et al. Identification of 11 novel mutations in 8 BBS genes by high-resolution homozygosity mapping. *Journal of Medical Genetics*. 2009.

Hauck F, Magerus-Chatinet A, Vicca S, Rensing-Ehl A, Roesen-Wolff A, Roesler J, et al. Somatic loss of heterozygosity, but not haploinsufficiency alone, leads to full-blown autoimmune lymphoproliferative syndrome in 1 of 12 family members with FAS start codon mutation. *Clinical Immunology*. 2013;147(1):61-8.

Hebecker M, Okemefuna AI, Perkins SJ, Mihlan M, Huber-Lang M, JÃ³zsi M. Molecular basis of C-reactive protein binding and modulation of complement activation by factor H-related protein 4. *Molecular Immunology*. 2011;47(6):1347-55.

Henrichsen CN, Chaignat E, Reymond A. Copy number variants, diseases and gene expression. *Human Molecular Genetics*. 2009;18(R1):R1-R8.

Hieronymus H, Schultz N, Gopalan A, Carver BS, Chang MT, Xiao Y, et al. Copy number alteration burden predicts prostate cancer relapse. *Proceedings of the National Academy of Sciences*. 2014;111(30):11139-44.

Hinch AG, Altemose N, Noor N, Donnelly P, Myers SR. Recombination in the human Pseudoautosomal region PAR1. *PLoS Genetics*. 2014;10(7):e1004503.

Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis & Rheumatism*. 1997;40(9):1725.

Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis & Rheumatology*. 1997;40(9):1725-.

Holers VM. Complement and its receptors: new insights into human disease. *Annual Review of Immunology*. 2014;32:433-59.

- Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. *Nature Reviews Genetics*. 2004 Dec;5(12):889-99.
- Hosking FJ, Papaemmanuil E, Sheridan E, Kinsey SE, Lightfoot T, Roman E, et al. Genome-wide homozygosity signatures and childhood acute lymphoblastic leukemia risk. *Blood*. 2010;115(22):4472-7.
- Humphreys K, Grankvist A, Leu M, Hall P, Liu J, Ripatti S, et al. The genetic structure of the Swedish population. *PloS One*. 2011;6(8):e22547.
- Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nature Genetics*. 2004;36(9):949-51.
- Invernizzi P, Miozzo M, Oertelt-Prigione S, Meroni PL, Persani L, Selmi C, et al. X monosomy in female systemic lupus erythematosus. *Annals of the New York Academy of Sciences*. 2007;1110(1):84-91.
- Itsara A, Cooper GM, Baker C, Girirajan S, Li J, Absher D, et al. Population analysis of large copy number variants and hotspots of human genetic disease. *The American Journal of Human Genetics*. 2009;84(2):148-61.
- Jiménez S, Cervera R, Font J, Ingelmo M. The epidemiology of systemic lupus erythematosus. *Clinical Reviews in Allergy & Immunology*. 2003;25(1):3-11.
- Johnson EC, Bjelland DW, Howrigan DP, Abdellaoui A, Breen G, Borglum A, et al. No Reliable Association between Runs of Homozygosity and Schizophrenia in a Well-Powered Replication Study. *PLoS Genetics*. 2016 Oct;12(10):e1006343.
- Kamen DL. Environmental influences on systemic lupus erythematosus expression. *Rheumatic Disease Clinics of North America*. 2014;40(3):401-12.
- Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance C. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genetics in Medicine*. 2011 Jul;13(7):680-5.
- Kelley J, Johnson MR, Alarcon GS, Kimberly RP, Edberg JC. Variation in the relative copy number of the TLR7 gene in patients with systemic lupus erythematosus and healthy control subjects. *Arthritis & Rheumatism*. 2007 Oct;56(10):3375-8.
- Khatun S, Saleh AA, Roy CK, Khan S, Pal S, Sattar H. Molecular Analysis of HLA-DR and their association with systemic lupus erythematosus. *Bangladesh Medical Research Council Bulletin*. 2017;42(3):111-9.
- Kidd JM, Cooper GM, Donahue WF, Hayden HS, Samps N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. *Nature*. 2008;453(7191):56-64.
- Kim JH, Jung SH, Bae JS, Lee HS, Yim SH, Park SY, et al. Deletion variants of RABGAP1L, 10q21. 3, and C4 are associated with the risk of systemic lupus erythematosus in Korean women. *Arthritis & Rheumatism*. 2013;65(4):1055-63.

- Kim K, Bang S-Y, Yoo DH, Cho S-K, Choi C-B, Sung Y-K, et al. Imputing variants in HLA-DR beta genes reveals that HLA-DRB1 is solely associated with rheumatoid arthritis and systemic lupus erythematosus. *PloS One*. 2016;11(2):e0150283.
- Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G, et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*. 2009;30(1):69-78.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Research*. 2009;19(9):1639-45.
- Ku CS, Naidoo N, Teo SM, Pawitan Y. Regions of homozygosity and their impact on complex diseases and traits. *Human Genetics*. 2011 Jan;129(1):1-15.
- Kukurba KR, Parsana P, Balliu B, Smith KS, Zappala Z, Knowles DA, et al. Impact of the X Chromosome and sex on regulatory variation. *Genome Research*. 2016;26(6):768-77.
- LaFramboise T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research*. 2009:gkp552.
- Law S, Dodds A, Porter R. A comparison of the properties of two classes, C4A and C4B, of the human complement component C4. *The EMBO Journal*. 1984;3(8):1819.
- Lee YH, Bae SC, Seo YH, Kim JH, Choi SJ, Ji JD, et al. Association between FCGR3B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. *Inflammatory Research*. 2015 Dec;64(12):983-91.
- Leffler J, Bengtsson AA, Blom AM. The complement system in systemic lupus erythematosus: an update. *Annals of the Rheumatic Diseases*. 2014;73(9):1601-6.
- Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences*. 2007;104(50):19942-7.
- Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, et al. Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. *The American Journal of Human Genetics*. 2016;99(3):540-54.
- Li N, Zhang J, Liao D, Yang L, Wang Y, Hou S. Association between C4, C4A, and C4B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. *Scientific Reports*. 2017 Feb 16;7:42628.
- Liang D, Peng Y, Lv W, Deng L, Zhang Y, Li H, et al. Copy number variation sequencing for comprehensive diagnosis of chromosome disease syndromes. *The Journal of Molecular Diagnostics*. 2014;16(5):519-26.

- Lin D, Zeng D. Correcting for population stratification in genomewide association studies. *Journal of the American Statistical Association*. 2011;106(495):997-1008.
- Lin PI, Kuo PH, Chen CH, Wu JY, Gau SS, Wu YY, et al. Runs of homozygosity associated with speech delay in autism in a taiwanese han population: evidence for the recessive model. *PLoS One*. 2013;8(8):e72056.
- Lins TC, Vieira RG, Abreu BS, Gentil P, Moreno-Lima R, Oliveira RJ, et al. Genetic heterogeneity of self-reported ancestry groups in an admixed Brazilian population. *Journal of Epidemiology*. 2011;21(4):240-5.
- Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *American Journal of Human Biology*. 2010;22(2):187-92.
- Liphaus BL, Umetsu N, Jesus AA, Bando SY, Silva CA, Carneiro-Sampaio M. Molecular characterization of the complement C1q, C2 and C4 genes in Brazilian patients with juvenile systemic lupus erythematosus. *Clinics*. 2015;70(3):220-7.
- Liu J, Lewinger JP, Gilliland FD, Gauderman WJ, Conti DV. Confounding and heterogeneity in genetic association studies with admixed populations. *American Journal of Epidemiology*. 2013;177(4):351-60.
- Liu K, Kurien BT, Zimmerman SL, Kaufman KM, Taft DH, Kottyan LC, et al. X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased Prevalence of 47, XXX in Systemic Lupus Erythematosus and Sjögren's Syndrome. *Arthritis & Rheumatology*. 2016;68(5):1290-300.
- Liu KC. X chromosome gene dosage in autoimmune disease susceptibility and B cell development [tese]. University of Cincinnati; 2016.
- Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*. 2001;25(4):402-8.
- Lv J, Yang Y, Zhou X, Yu L, Li R, Hou P, et al. FCGR3B copy number variation is not associated with lupus nephritis in a Chinese population. *Lupus*. 2010;19(2):158-61.
- MacDonald JR, Ziman R, Yuen RK, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*. 2014;42(D1):D986-D92.
- Mafra F, Mazzotti D, Pellegrino R, Bianco B, Barbosa CP, Hakonarson H, et al. Copy number variation analysis reveals additional variants contributing to endometriosis development. *Journal of Assisted Reproduction and Genetics*. 2016:1-8.
- Mahmoudi M, Nilsson PH, Mollnes TE, Roos D, Sullivan KE. Complement Deficiencies. *Primary Immunodeficiency Diseases*: Springer 2017:437-60.
- Maidhof W, Hilar O. Lupus: An Overview of the Disease And Management Options. *Pharmacy and Therapeutics*. 2012;37(4):240.

- Malhotra D, Sebat J. CNVs: harbingers of a rare variant revolution in psychiatric genetics. *Cell*. 2012;148(6):1223-41.
- Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, et al. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythematosus. *Annals of the Rheumatic Diseases*. 2007.
- Marescalco MS, Capizzi C, Condorelli DF, Barresi V. Genome-wide analysis of recurrent copy-number alterations and copy-neutral loss of heterozygosity in head and neck squamous cell carcinoma. *Journal of Oral Pathology & Medicine*. 2014 Jan;43(1):20-7.
- Margery-Muir AA, Bundell C, Nelson D, Groth DM, Wetherall JD. Gender balance in patients with systemic lupus erythematosus. *Autoimmunity Reviews*. 2017 Mar;16(3):258-68.
- Marshall C, Howrigan D, Merico D, Thiruvahindrapuram B, Wu W, Greer D, et al. A contribution of novel CNVs to schizophrenia from a genome-wide study of 41,321 subjects. *Nature Genetics*. 2016; (49):27-35.
- Martin LS, Eskin E. Population Structure in Genetic Studies: Confounding Factors and Mixed Models. *bioRxiv*. 2016:092106.
- Maxwell A. Using ancestry-informative markers to identify fine structure across 15 populations of European origin. *European Journal of Human Genetics*. 2014;22(10):1190-1200.
- Mayadas TN, Tsokos GC, Tsuboi N. Mechanisms of immune complex-mediated neutrophil recruitment and tissue injury. *Circulation*. 2009;120(20):2012-24.
- McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nature Genetics*. 2007;39:S37-S42.
- McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nature Genetics*. 2008;40(10):1166-74.
- McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*. 2008;9(5):356-69.
- McHugh NJ, Owen P, Cox B, Dunphy J, Welsh K. MHC class II, tumour necrosis factor α , and lymphotoxin α gene haplotype associations with serological subsets of systemic lupus erythematosus. *Annals of the Rheumatic Diseases*. 2006;65(4):488-94.
- McQuillan R, Leutenegger AL, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *The American Journal of Human Genetics*. 2008 Sep;83(3):359-72.
- McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. The fine-scale structure of recombination rate variation in the human genome. *Science*. 2004;304(5670):581-4.

Medjeral-Thomas N, Malik TH, Patel MP, Toth T, Cook HT, Tomson C, et al. A novel CFHR5 fusion protein causes C3 glomerulopathy in a family without Cypriot ancestry. *Kidney International*. 2014 Apr;85(4):933-7.

Mezzavilla M, Vozzi D, Badii R, Alkowari MK, Abdulhadi K, Girotto G, et al. Increased rate of deleterious variants in long runs of homozygosity of an inbred population from Qatar. *Human Heredity*. 2015;79(1):14-9.

Mihlan M, Hebecker M, Dahse HM, Hälbig S, Huber-Lang M, Dahse R, et al. Human complement factor H-related protein 4 binds and recruits native pentameric C-reactive protein to necrotic cells. *Molecular Immunology*. 2009;46(3):335-44.

Mikulasova A, Smetana J, Wayhelova M, Janyskova H, Okubote SA, Hajek R, et al. Whole genome amplification effect on segmental copy-number changes and copy-number neutral loss of heterozygosity analysis by oligonucleotide-based array-comparative genomic hybridization in human myeloma cell line. *International Journal of Clinical and Experimental Pathology*. 2016;9(7):6965-+.

Miller S, Dykes D, Polesky H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*. 1988;16(3):1215.

Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, et al. Mapping copy number variation by population-scale genome sequencing. *Nature*. 2011;470(7332):59-65.

Miyagawa H, Yamai M, Sakaguchi D, Kiyohara C, Tsukamoto H, Kimoto Y, et al. Association of polymorphisms in complement component C3 gene with susceptibility to systemic lupus erythematosus. *Rheumatology*. 2008;47(2):158-64.

Mizoguchi M, Kuga D, Guan Y, Hata N, Nakamizo A, Yoshimoto K, et al. Loss of heterozygosity analysis in malignant gliomas. *Brain Tumor Pathology*. 2011 Jul;28(3):191-6.

Mohan C, Putterman C. Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nature Reviews Nephrology*. 2015 Jun;11(6):329-41.

Moore I, Strain L, Pappworth I, Kavanagh D, Barlow PN, Herbert AP, et al. Association of factor H autoantibodies with deletions of CFHR1, CFHR3, CFHR4, and with mutations in CFH, CFI, CD46, and C3 in patients with atypical hemolytic uremic syndrome. *Blood*. 2010;115(2):379-87.

Muniz YCN, Ferreira LB, Mendes-Junior CT, Wiezel CEV, Simões AL. Genomic ancestry in urban Afro-Brazilians. *Annals of Human Biology*. 2008;35(1):104-11.

Nakashima CAK, Silva JFMd, Galhardo AP, Leite MFS, Santos ABdSd, Fiorenzano GR, et al. Incidência e aspectos clínico-laboratoriais do lúpus eritematoso sistêmico em cidade do sul do Brasil. *Revista Brasileira de Reumatologia*. 2011:235-9.

- Nalls M, Guerreiro R, Simon-Sanchez J, Bras J, Traynor B, Gibbs J, et al. Extended tracts of homozygosity identify novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics*. 2009;10(3):183-90.
- Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*. 2011;11(12):823-36.
- Nimmerjahn F, Ravetch JV. Fcγ receptors as regulators of immune responses. *Nature Reviews Immunology*. 2008;8(1):34-47.
- Olgiati S, Quadri M, Fang M, Rood JP, Saute JA, Chien HF, et al. DNAJC6 Mutations Associated With Early-Onset Parkinson's Disease. *Annals of Neurology*. 2016.
- Özbalkan Z, Bagislar S, Kiraz S, Akyerli CB, Özer HT, Yavuz Å, et al. Skewed X chromosome inactivation in blood cells of women with scleroderma. *Arthritis & Rheumatism*. 2005;52(5):1564-70.
- Ozcelik T, Uz E, Akyerli CB, Bagislar S, Mustafa CA, Gursoy A, et al. Evidence from autoimmune thyroiditis of skewed X-chromosome inactivation in female predisposition to autoimmunity. *European Journal of Human Genetics*. 2006;14(6):791-7.
- Palmer E, Speirs H, Taylor PJ, Mullan G, Turner G, Einfeld S, et al. Changing interpretation of chromosomal microarray over time in a community cohort with intellectual disability. *American Journal of Medical Genetics Part A*. 2014;164(2):377-85.
- Pang J, Zhang S, Yang P, Hawkins-Lee B, Zhong J, Zhang Y, et al. Loss-of-function mutations in HPSE2 cause the autosomal recessive urofacial syndrome. *The American Journal of Human Genetics*. 2010;86(6):957-62.
- Peiffer DA, Le JM, Steemers FJ, Chang W, Jenniges T, Garcia F, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Research*. 2006;16(9):1136-48.
- Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ. Genomic patterns of homozygosity in worldwide human populations. *The American Journal of Human Genetics*. 2010;91(2):275-92.
- Pena SD, Bastos-Rodrigues L, Pimenta J, Bydlowski S. DNA tests probe the genomic ancestry of Brazilians. *Brazilian Journal of Medical and Biological Research*. 2009;42(10):870-6.
- Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*. 2007;39(10):1256-60.
- Petri M. Epidemiology of systemic lupus erythematosus. *Best Practice & Research Clinical Rheumatology*. 2002;16(5):847-58.
- Pippucci T, Magi A, Gialluisi A, Romeo G. Detection of runs of homozygosity from whole exome sequencing data: State of the art and perspectives for clinical, population and epidemiological studies. *Human Heredity*. 2014;77:63-72.

- Pons-Estel GJ, Alarcón GS, Scofield L, Reinlib L, Cooper GS. Understanding the epidemiology and progression of systemic lupus erythematosus. *Seminars in Arthritis and Rheumatism*. 2010;39(4):257-68.
- Ptacek T, Li X, Kelley J, Edberg J. Copy number variants in genetic susceptibility and severity of systemic lupus erythematosus. *Cytogenetic and Genome Research*. 2008;123(1-4):142-7.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*. 2007;81(3):559-75.
- Qin P, Li Z, Jin W, Lu D, Lou H, Shen J, et al. A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese. *European Journal of Human Genetics*. 2014;22(2):248-53.
- Rahbari R, Zuccherato LW, Tischler G, Chihota B, Ozturk H, Saleem S, et al. Understanding the genomic structure of copy-number variation of the low-affinity Fcγ receptor region allows confirmation of the association of FCGR3B deletion with rheumatoid arthritis. *Human Mutation*. 2017;38(4):390-9.
- Ravetch JV, Bolland S. IgG Fc receptors. *Annual Review of Immunology*. 2001;19(1):275-90.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. *Nature*. 2006;444(7118):444-54.
- Rhodes B, Hunnankul S, Morris D, Hsaio L, Graham DC, Nitsch D, et al. The heritability and genetics of complement C3 expression in UK SLE families. *Genes and Immunity*. 2009;10(5):525-30.
- Rhodes B, Vyse T. The genetics of SLE: an update in the light of genome-wide association studies. *Rheumatology*. 2008;47(11):1603-11.
- Ribeiro D. O povo brasileiro: a formação e o sentido do Brasil: Global Editora e Distribuidora Ltda 2015.
- Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*. 2015 May;17(5):405-24.
- Rodrigues RN. Os africanos no Brasil. São Paulo: Editora Nacional. 1932.
- Rodriguez de Córdoba S, Esparza-Gordillo J, de Jorge EG, Lopez-Trascasa M, Sánchez-Corral P. The human complement factor H: functional roles, genetic variations and disease associations. *Molecular Immunology*. 2004;41(4):355-67.
- Rolim H, Cronemberger S, Rangel H, Batista WD, Bastos-Rodrigues L, De Marco L. The role of genetic ancestry in Brazilian patients with primary congenital glaucoma. *Journal of Glaucoma*. 2016;25(1):e24-e8.

- Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419(6909):832-7.
- Sacks FM, Hermans MP, Fioretto P, Valensi P, Davis T, Horton E, et al. Association Between Plasma Triglycerides and High-Density Lipoprotein Cholesterol and Microvascular Kidney Disease and Retinopathy in Type 2 Diabetes Mellitus A Global Case-Control Study in 13 Countries. *Circulation*. 2014;129(9):999-1008.
- Safieh LA, Aldahmesh M, Shamseldin H, Hashem M, Shaheen R, Alkuraya H, et al. Clinical and molecular characterisation of Bardet-Biedl syndrome in consanguineous populations: the power of homozygosity mapping. *Journal of Medical Genetics*. 2010;47(4):236-41.
- Salzano FM. Em busca das raízes. *Ciência Hoje*. 1986;5(25):48-53.
- Salzano FM, Bortolini MC. The evolution and genetics of Latin American populations: Cambridge University Press 2005.
- Sanchez E, Torres B, Vilches J, Lopez-Nevot M, Ortego-Centeno N, Jimenez-Alonso J, et al. No primary association of MICA polymorphism with systemic lupus erythematosus. *Rheumatology*. 2006;45(9):1096-100.
- Saxena K, Kitzenmiller KJ, Wu YL, Zhou B, Esack N, Hiremath L, et al. Great genotypic and phenotypic diversities associated with copy-number variations of complement C4 and RP-C4-CYP21-TNX (RCCX) modules: a comparison of Asian-Indian and European American populations. *Molecular Immunology*. 2009;46(7):1289-303.
- Schaschl H, Aitman TJ, Vyse TJ. Copy number variation in the human genome and its implication in autoimmunity. *Clinical & Experimental Immunology*. 2009 Apr;156(1):12-6.
- Schrider DR, Hahn MW. Gene copy-number polymorphism in nature. *Proceedings of the Royal Society of London B: Biological Sciences*. 2010;277(1698):3213-21.
- Scofield RH, Bruner GR, Namjou B, Kimberly RP, Ramsey-Goldman R, Petri M, et al. Klinefelter's syndrome (47, XXY) in male systemic lupus erythematosus patients: Support for the notion of a gene-dose effect from the X chromosome. *Arthritis & Rheumatism*. 2008;58(8):2511-7.
- Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science*. 2004 Jul 23;305(5683):525-8.
- Shi J, Zhou W, Zhu B, Hyland PL, Bennett H, Xiao Y, et al. Rare germline copy number variations and disease susceptibility in familial melanoma. *Journal of Investigative Dermatology*. 2016;136(12):2436-43.
- Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN. Mapping of disease-associated variants in admixed populations. *Genome Biology*. 2011;12(5):223.

- Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. *Human Genetics*. 2003;112(4):387-99.
- Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, et al. Ethnic-affiliation estimation by use of population-specific DNA markers. *The American Journal of Human Genetics*. 1997;60(4):957.
- Sims R, Dwyer S, Harold D, Gerrish A, Hollingworth P, Chapman J, et al. No evidence that extended tracts of homozygosity are associated with Alzheimer's disease. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*. 2011;156(7):764-71.
- Singh N, Traisak P, Martin KA, Kaplan MJ, Cohen PL, Denny MF. Genomic alterations in abnormal neutrophils isolated from adult patients with systemic lupus erythematosus. *Arthritis Research & Therapy*. 2014;16(4):R165.
- Siraj AK, Khalak HG, Sultana M, Al-Rasheed M, Bavi P, Al-Sanea N, et al. Colorectal cancer risk is not associated with increased levels of homozygosity in Saudi Arabia. *Genetics in Medicine*. 2012;14(8):720-8.
- South ST, Lee C, Lamb AN, Higgins AW, Kearney HM, Working Group for the American College of Medical G, et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genetics in Medicine*. 2013 Nov;15(11):901-9.
- Stephen J, Girisha KM, Dalal A, Shukla A, Shah H, Srivastava P, et al. Mutations in patients with osteogenesis imperfecta from consanguineous Indian families. *European Journal of Medical Genetics*. 2015;58(1):21-7.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, Thorne N, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science*. 2007;315(5813):848-53.
- Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. *Science*. 2010;330(6004):641-6.
- Sudmant PH, Mallick S, Nelson BJ, Hormozdiari F, Krumm N, Huddleston J, et al. Global diversity, population stratification, and selection of human copy-number variation. *Science*. 2015a;349(6253):aab3761.
- Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015b;526(7571):75-81.
- Sun C, Molineres JE, Looger LL, Zhou XJ, Kim K, Okada Y, et al. High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nature Genetics*. 2016 Mar;48(3):323-30.
- Szpiech ZA, Xu J, Pemberton Trevor J, Peng W, Zöllner S, Rosenberg Noah A, et al. Long runs of homozygosity are enriched for deleterious variation. *The American Journal of Human Genetics*. 2013;93(1):90-102.

- Tan EM, Cohen AS, Fries JF, Masi AT, Mcshane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis & Rheumatism*. 1982;25(11):1271-7.
- Tang H, Quertermous T, Rodriguez B, Kardia SL, Zhu X, Brown A, et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *The American Journal of Human Genetics*. 2005;76(2):268-75.
- Terashima K, Yu A, Chow WY, Hsu WC, Chen P, Wong S, et al. Genome-wide analysis of DNA copy number alterations and loss of heterozygosity in intracranial germ cell tumors. *Pediatric Blood & Cancer*. 2014 Apr;61(4):593-600.
- Teruel M, Alarcón-Riquelme ME. The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *Journal of Autoimmunity*. 2016 Nov;74:161-75.
- Thomsen H, Chen B, Figlioli G, Elisei R, Romei C, Cipollini M, et al. Runs of homozygosity and inbreeding in thyroid cancer. *BMC Cancer*. 2016;16(1):227.
- Townson JR, Barcellos LF, Nibbs RJ. Gene copy number regulates the production of the human chemokine CCL3-L1. *European Journal of Immunology*. 2002;32(10):3016-26.
- Tyner C, Barber GP, Casper J, Clawson H, Diekhans M, Eisenhart C, et al. The UCSC Genome Browser database: 2017 update. *Nucleic Acids Research*. 2016:gkw1134.
- Uddin M, Thiruvahindrapuram B, Walker S, Wang Z, Hu P, Lamoureux S, et al. A high-resolution copy-number variation resource for clinical and population genetics. *Genetics in Medicine*. 2015.
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*. 2007;35(suppl 2):W71-W4.
- Vilar MP, Sato E. Estimating the incidence of systemic lupus erythematosus in a tropical region (Natal, Brazil). *Lupus*. 2002;11(8):528-32.
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *The American Journal of Human Genetics*. 2012;90(1):7-24.
- Wain LV, Tobin MD, eds. Copy number variation. *Methods in Molecular Biology*. 2010/12/15 ed 2011.
- Weckerle CE, Niewold TB. The unexplained female predominance of systemic lupus erythematosus: clues from genetic and cytokine studies. *Clinical Reviews in Allergy & Immunology*. 2011;40(1):42-9.
- Weir BS, Cockerham C. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc, Sunderland, MA, USA. 1996.
- Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, Newland SA, et al. Copy number of FCGR3B, which is associated with systemic lupus

erythematosus, correlates with protein expression and immune complex uptake. *The Journal of Experimental Medicine*. 2008;205(7):1573-82.

Woods CG, Cox J, Springell K, Hampshire DJ, Mohamed MD, McKibbin M, et al. Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *The American Journal of Human Genetics*. 2006;78(5):889-96.

Wu L, Guo S, Yang D, Ma Y, Ji H, Chen Y, et al. Copy number variations of HLA-DRB5 is associated with systemic lupus erythematosus risk in Chinese Han population. *Acta Biochimica et Biophysica Sinica*. 2014;46(2):155-60.

Yang ML, Gee AJP, Gee RJ, Zurita-Lopez CI, Khare S, Clarke SG, et al. Lupus autoimmunity altered by cellular methylation metabolism. *Autoimmunity*. 2012.

Yang TL, Guo Y, Zhang LS, Tian Q, Yan H, Papasian CJ, et al. Runs of homozygosity identify a recessive locus 12q21. 31 for human adult height. *The Journal of Clinical Endocrinology & Metabolism*. 2010;95(8):3777-82.

Yang ZF, Ho DW, Lau CK, Lam CT, Lum CT, Poon RT, et al. Allograft inflammatory factor-1 (AIF-1) is crucial for the survival and pro-inflammatory activity of macrophages. *International Immunology*. 2005;17(11):1391-7.

Yih Chen J, Ling Wu Y, Yin Mok M, Jan Wu YJ, Lintner KE, Wang CM, et al. Effects of complement C4 gene copy number variations, size dichotomy, and C4A deficiency on genetic risk and clinical presentation of systemic lupus erythematosus in east asian populations. *Arthritis & Rheumatology*. 2016;68(6):1442-53.

Yim SH, Jung SH, Chung B, Chung YJ. Clinical implications of copy number variations in autoimmune disorders. *The Korean Journal of Internal Medicine*. 2015;30(3):294-304.

Younkin SG, Scharpf RB, Schwender H, Parker MM, Scott AF, Marazita ML, et al. A genome-wide study of inherited deletions identified two regions associated with nonsyndromic isolated oral clefts. *Birth Defects Research Part A: Clinical and Molecular Teratology*. 2015;103(4):276-83.

Yu B, Guan M, Peng Y, Shao Y, Zhang C, Yue X, et al. Copy number variations of interleukina-17F, interleukina-21, and interleukina-22 are associated with systemic lupus erythematosus. *Arthritis & Rheumatism*. 2011;63(11):3487-92.

Yu B, Shao Y, Li P, Zhang J, Zhong Q, Yang H, et al. Copy number variations of the human histamine H4 receptor gene are associated with systemic lupus erythematosus. *Br J Dermatol*. 2010 Nov;163(5):935-40.

Yu C, Gershwin ME, Chang C. Diagnostic criteria for systemic lupus erythematosus: a critical review. *Journal of Autoimmunity*. 2014 Feb-Mar;48-49:10-3.

Yuan J, Zhao D, Wu L, Xu X, Pang Y, Zhang J, et al. FCGR3B copy number loss rather than gain is a risk factor for systemic lupus erythematosus and lupus nephritis: a meta-analysis. *International Journal of Rheumatic Diseases*. 2015.

Zanda M, Onengut-Gumuscu S, Walker N, Shtir C, Gallo D, Wallace C, et al. A genome-wide assessment of the role of untagged copy number variants in type 1 diabetes. *PLoS Genetics*. 2014;10(5):e1004367.

Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nature Reviews Genetics*. 2015.

Zhao J, Gu M, Bai J, Song X, Sun Y, Qiu C. Association study between HLA-DRB genes and asthmatic children. *Zhonghua Yi Xue Za Zhi*. 2009;89(32):2257-60.

Zhao J, Wu H, Khosravi M, Cui H, Qian X, Kelly JA, et al. Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility. *PLoS Genetics*. 2011;7(5):e1002079.

Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: A genome-wide association study in Koreans with replication in North Americans. *Arthritis & Rheumatism*. 2009;60(12):3807-14.

Zhou XJ, Cheng FJ, Lv JC, Luo H, Yu F, Chen M, et al. Higher DEFB4 genomic copy number in SLE and ANCA-associated small vasculitis. *Rheumatology (Oxford)*. 2012 Jun;51(6):992-5.

Zuccherato LW, Schneider S, Tarazona-Santos E, Hardwick RJ, Berg DE, Bogle H, et al. Population genetics of immune-related multilocus copy number variation in Native Americans. *Journal of The Royal Society Interface*. 2017;14(128):20170057.

10 APÊNDICE

APÊNDICE A – Características clínicas dos pacientes com LES

Apêndice A – Características clínicas e laboratoriais dos pacientes com LES analisados no presente trabalho (n = 135).

ID LES	Sexo	Idade diagn*	EUR	AFR	AME	Nefrite lúpica	Artrite	Rash malar	Úlceras orais/nasais	Fotossensibilidade	Convulsões e psicoses	Alterações hematológicas	Alterações imunológicas
LES001	F	27	97%	1%	2%								
LES002	F	19	70%	14%	15%								
LES003	F	40	33%	51%	16%								
LES004	F	35	43%	29%	28%								
LES005	F	44	59%	29%	13%								
LES006	F	36	31%	0%	69%								
LES007	F	47	65%	13%	22%								
LES008	F	58	37%	24%	39%								
LES009	F	21	40%	53%	8%								
LES010	F	47	53%	28%	18%								
LES011	F	55	68%	7%	25%								
LES012	F	31	49%	49%	2%								
LES013	F	24	64%	25%	11%								
LES014	F	36	67%	23%	10%								
LES015	F	16	58%	21%	21%								
LES016	F	27	41%	34%	25%								
LES017	F	24	57%	7%	36%								
LES018	F	18	60%	13%	27%								
LES019	M	24	78%	12%	11%								
LES020	M	15	66%	5%	29%								
LES021	M	17	63%	8%	29%								
LES022	M	29	38%	34%	28%								
LES023	M	16	62%	23%	15%								
LES024	F	56	80%	14%	6%								
LES025	F	27	22%	71%	7%								
LES026	F	14	97%	1%	1%								
LES027	F	27	96%	2%	2%								

LES028	F	19	98%	2%	1%								
LES029	F	NA	NA	NA	NA								
LES030	F	33	94%	2%	4%								
LES031	F	NA	56%	5%	39%								
LES032	F	21	56%	26%	19%								
LES033	F	41	68%	2%	30%								
LES034	F	34	95%	2%	4%								
LES035	F	19	91%	2%	7%								
LES036	M	17	5%	91%	3%								
LES037	F	35	84%	1%	15%								
LES038	F	39	35%	14%	51%								
LES039	F	39	97%	2%	2%								
LES040	F	12	87%	3%	10%								
LES041	F	26	93%	3%	4%								
LES042	F	45	87%	3%	10%								
LES043	F	57	96%	2%	2%								
LES044	F	36	3%	94%	3%								
LES045	F	46	98%	1%	1%								
LES046	F	46	94%	5%	2%								
LES047	F	26	97%	2%	1%								
LES048	F	51	50%	22%	27%								
LES049	F	24	96%	3%	2%								
LES050	F	32	96%	1%	3%								
LES051	F	20	68%	4%	28%								
LES052	F	31	82%	1%	17%								
LES053	F	21	97%	2%	1%								
LES054	F	25	63%	3%	34%								
LES055	F	55	84%	3%	14%								
LES056	F	29	19%	13%	68%								
LES057	F	40	67%	5%	28%								
LES058	F	30	96%	3%	2%								
LES059	F	27	90%	2%	8%								
LES060	F	19	47%	46%	7%								
LES061	M	53	43%	3%	54%								
LES062	F	45	84%	9%	7%								
LES063	F	46	89%	3%	8%								
LES064	F	43	96%	3%	1%								
LES065	F	40	96%	1%	3%								
LES066	F	53	37%	59%	3%								

LES067	F	33	85%	8%	7%								
LES068	F	23	69%	18%	14%								
LES069	F	28	41%	2%	57%								
LES070	F	62	1%	96%	3%								
LES071	F	24	59%	3%	38%								
LES072	F	27	84%	2%	15%								
LES073	F	NA	76%	20%	4%								
LES074	F	18	14%	77%	10%								
LES075	F	27	5%	94%	1%								
LES076	F	38	87%	4%	9%								
LES077	F	37	96%	1%	3%								
LES078	F	55	98%	1%	1%								
LES079	F	23	84%	13%	3%								
LES080	F	40	95%	3%	2%								
LES081	F	23	79%	19%	3%								
LES082	F	14	48%	8%	45%								
LES083	F	10	80%	18%	3%								
LES084	F	13	25%	72%	3%								
LES085	F	52	20%	74%	6%								
LES086	F	41	62%	4%	35%								
LES087	F	15	95%	2%	3%								
LES088	F	23	74%	2%	25%								
LES089	F	18	45%	6%	50%								
LES090	F	35	6%	81%	13%								
LES091	M	27	88%	2%	10%								
LES092	F	19	72%	18%	10%								
LES093	F	21	94%	3%	3%								
LES094	F	31	18%	77%	5%								
LES095	F	15	33%	33%	34%								
LES096	F	24	95%	2%	3%								
LES097	F	15	94%	4%	2%								
LES098	F	21	4%	89%	7%								
LES099	F	49	86%	7%	7%								
LES100	M	44	95%	2%	3%								
LES101	F	38	94%	5%	1%								
LES102	M	49	94%	1%	4%								
LES103	F	37	96%	2%	2%								
LES104	F	31	95%	4%	1%								
LES105	F	32	97%	1%	2%								

LES106	F	45	87%	9%	4%								
LES107	F	49	11%	33%	56%								
LES108	F	12	90%	9%	1%								
LES109	F	13	96%	2%	2%								
LES110	F	39	89%	9%	2%								
LES111	F	14	93%	6%	1%								
LES112	F	40	28%	68%	4%								
LES113	F	23	94%	4%	2%								
LES114	F	31	93%	5%	2%								
LES115	F	20	94%	4%	3%								
LES116	F	40	91%	7%	2%								
LES117	F	37	71%	3%	26%								
LES118	F	40	78%	2%	20%								
LES119	F	58	92%	2%	7%								
LES120	F	46	69%	2%	29%								
LES121	F	38	6%	91%	3%								
LES122	F	23	25%	71%	5%								
LES123	F	23	52%	31%	17%								
LES124	F	32	79%	19%	2%								
LES125	F	36	39%	50%	11%								
LES126	F	24	67%	2%	31%								
LES127	F	43	88%	9%	3%								
LES128	F	17	97%	1%	2%								
LES129	F	47	76%	13%	11%								
LES130	F	26	92%	1%	7%								
LES131	F	37	47%	38%	15%								
LES132	F	29	58%	37%	5%								
LES133	F	45	86%	6%	8%								
LES134	F	19	97%	3%	1%								
LES135	F	25	96%	3%	1%								

ID = identificação; F = feminino; M = masculino; Idade diagn = idade em que foi diagnosticado LES no pacientes; NA = não apresenta; EUR = europeu; AFR = africano; AME = ameríndio.

APÊNDICE B – PAINEL DE 345 SNP-AIMs

Apêndice B – PAINEL DE 345 *SNP-AIMs* selecionados para inferência de ancestralidade em pacientes com LES e indivíduos saudáveis genotipados por *Cytoscan HD array*.

ID AIM	Sonda Cytoscan HD array	Chr	#rs dbSNP	Alelo (A1A2)	Frequência (Alelo A1)			δ			Ref.	<i>p</i> EHW LES	<i>p</i> EHW BRA	F_{st}
					AFR	AME	EUR	AFR-AME	AFR-EUR	AME-EUR				
AIM001	S-4FSJS	1	rs6685064	CT	0,422	0,244	0,930	0,178	0,508	0,686	[1]	0,609	0,773	-0,004
AIM002	S-3ONKX	1	rs12085319	TG	0,985	0,653	0,183	0,332	0,802	0,470	[1]	0,689	0,104	0,080
AIM003	S-3NSWA	1	rs4920310	AG	0,883	0,695	0,181	0,188	0,702	0,514	[1]	0,611	0,301	0,000
AIM004	S-3KCDP	1	rs7515867	CT	0,039	0,062	0,613	0,023	0,574	0,551	[1]	0,652	0,249	0,038
AIM005	S-3CQWZ	1	rs11264115	AG	0,383	0,332	0,953	0,051	0,570	0,621	[1]	1,000	0,283	-0,008
AIM006	S-3XZFH	1	rs1469344	CT	0,902	0,981	0,613	0,079	0,289	0,368	[3]	0,519	0,297	0,062
AIM007	S-3AASD	1	rs12130799	GA	0,010	0,510	0,070	0,500	0,060	0,440	[2]	1,000	1,000	-0,008
AIM008	S-4NTRB	1	rs2984915	TC	0,853	0,011	0,090	0,842	0,763	0,079	[1]	0,251	1,000	-0,009
AIM009	S-3KGQG	1	rs12743599	CT	0,107	0,151	0,726	0,044	0,619	0,575	[1]	0,401	0,040	0,009
AIM010	S-3HKCM	1	rs12727814	AG	0,922	0,943	0,248	0,021	0,674	0,695	[1]	0,105	0,809	0,087
AIM011	S-4MAKQ	1	rs6681719	GA	0,733	0,054	0,895	0,679	0,162	0,841	[1]	1,000	0,456	0,006
AIM012	S-3USVJ	1	rs12065716	CG	0,864	0,611	0,115	0,253	0,749	0,496	[1]	0,128	0,152	-0,011
AIM013	S-4ILLQ	1	rs12021830	TC	0,995	0,107	0,838	0,888	0,157	0,731	[1]	1,000	0,755	0,031
AIM014	S-4TKSP	1	rs12123239	AG	0,015	0,023	0,608	0,008	0,593	0,585	[1]	0,399	0,850	0,017
AIM015	S-3GBBA	1	rs2065161	TC	0,564	0,078	0,947	0,486	0,383	0,869	[3]	1,000	0,416	-0,012
AIM016	S-3ABMR	1	rs2065160	AG	0,487	0,137	0,917	0,350	0,430	0,780	[4]	0,538	0,483	-0,003
AIM017	S-3XMHO	1	rs10779334	AG	0,163	0,992	0,903	0,829	0,740	0,089	[1]	1,000	0,309	-0,014
AIM018	S-4GYDH	1	rs6677662	GA	0,029	0,271	0,752	0,242	0,723	0,481	[1]	0,405	0,838	0,065
AIM019	S-3QRBj	2	rs798443	AG	0,030	0,680	0,830	0,650	0,800	0,150	[2]	1,000	0,518	0,018
AIM020	S-4PHLB	2	rs181130	CG	0,981	0,179	0,868	0,802	0,113	0,689	[1]	1,000	0,190	-0,002
AIM021	S-3MSPM	2	rs2140701	AG	0,063	0,966	0,414	0,903	0,351	0,552	[1]	1,000	0,838	0,037
AIM022	S-3KLAF	2	rs1869380	CG	0,164	0,636	0,362	0,472	0,198	0,274	[3]	0,162	0,671	-0,014

AIM023	S-3ENUH	2	rs1367158	CT	0,189	0,080	0,720	0,109	0,531	0,640	[1]	0,405	0,696	-0,013
AIM024	S-3WAEB	2	rs2710684	TC	0,146	0,935	0,217	0,789	0,071	0,718	[1]	0,662	0,802	0,012
AIM025	S-4MROR	2	rs6711746	TC	0,874	0,102	0,809	0,772	0,065	0,707	[1]	1,000	0,347	-0,010
AIM026	S-3YEIW	2	rs305163	GA	0,063	0,172	0,712	0,109	0,649	0,540	[1]	0,682	0,166	-0,009
AIM027	S-4JBVC	2	rs11124754	CT	0,704	0,787	0,081	0,083	0,623	0,706	[1]	0,609	0,238	0,001
AIM028	S-3GRXU	2	rs6740875	GA	0,175	0,321	0,848	0,146	0,673	0,527	[1]	1,000	1,000	-0,013
AIM029	S-3QOJF	2	rs2595214	AG	1,000	0,168	0,852	0,832	0,148	0,684	[1]	0,662	1,000	0,044
AIM030	S-3MLSJ	2	rs4643526	AG	0,917	0,019	0,177	0,898	0,740	0,158	[1]	1,000	0,637	0,004
AIM031	S-3HFEY	2	rs4671675	TC	0,194	0,298	0,843	0,104	0,649	0,545	[1]	0,611	0,330	-0,012
AIM032	S-3UIKO	2	rs13400937	GT	0,910	0,230	0,210	0,680	0,700	0,020	[2]	0,684	0,661	0,090
AIM033	S-3EXHX	2	rs7601949	TC	0,990	0,973	0,477	0,017	0,513	0,496	[1]	0,643	0,017	0,005
AIM034	S-3MYZM	2	rs7594727	AC	0,869	0,008	0,136	0,861	0,733	0,128	[1]	0,654	0,030	0,044
AIM035	S-4QSLF	2	rs1257084	TC	0,068	0,212	0,768	0,144	0,700	0,556	[1]	1,000	0,247	-0,003
AIM036	S-3DAZC	2	rs7598069	AG	0,029	0,989	0,390	0,960	0,361	0,599	[1]	0,402	0,441	-0,006
AIM037	S-4GSHV	2	rs1567803	CT	0,971	0,981	0,358	0,010	0,613	0,623	[1]	1,000	1,000	0,023
AIM038	S-4RLLP	2	rs6431064	CA	0,015	0,088	0,587	0,073	0,572	0,499	[1]	1,000	0,444	-0,012
AIM039	S-3KMRY	2	rs1036543	GA	0,650	0,718	0,061	0,068	0,589	0,657	[1]	0,538	0,359	-0,015
AIM040	S-4QNSO	2	rs7589619	CT	0,922	0,050	0,218	0,872	0,704	0,168	[1]	1,000	0,344	-0,004
AIM041	S-3KVKY	2	rs6429990	CG	0,976	0,977	0,430	0,001	0,546	0,547	[1]	0,399	0,034	-0,005
AIM042	S-3DFKH	2	rs298259	CT	0,131	0,160	0,761	0,029	0,630	0,601	[1]	0,676	0,558	0,011
AIM043	S-3AQPO	2	rs6759814	CT	0,704	0,737	0,055	0,033	0,649	0,682	[1]	1,000	1,000	0,012
AIM044	S-4NCLI	2	rs2044770	AG	0,971	0,076	0,824	0,895	0,147	0,748	[1]	1,000	0,024	-0,001
AIM045	S-4DXOC	2	rs3927978	AG	0,010	0,973	0,712	0,963	0,702	0,261	[1]	1,000	0,214	0,050
AIM046	S-3UNFQ	2	rs7568054	AG	0,995	0,947	0,349	0,048	0,646	0,598	[1]	0,670	0,021	0,017
AIM047	S-3ZDST	2	rs6707773	CT	0,791	0,057	0,764	0,734	0,027	0,707	[1]	1,000	0,521	-0,011
AIM048	S-3REUC	2	rs1569175	TC	0,260	0,640	0,040	0,380	0,220	0,600	[2]	0,414	0,666	-0,011
AIM049	S-4LNEW	2	rs16851773	AG	0,976	0,874	0,353	0,102	0,623	0,521	[1]	0,684	1,000	-0,010
AIM050	S-4FLNZ	3	rs6768750	GA	0,976	0,073	0,695	0,903	0,281	0,622	[1]	0,650	0,419	-0,012
AIM051	S-3DIBQ	3	rs4955903	TC	0,956	0,023	0,647	0,933	0,309	0,624	[1]	0,414	0,515	0,054

AIM052	S-4CTGG	3	rs6803290	CT	0,010	0,023	0,636	0,013	0,626	0,613	[1]	1,000	0,451	-0,008
AIM053	S-4LYVG	3	rs9833943	GA	0,029	0,931	0,309	0,902	0,280	0,622	[1]	0,619	0,344	-0,013
AIM054	S-3TGZQ	3	rs3925004	AC	0,044	0,885	0,193	0,841	0,149	0,692	[1]	0,519	0,137	-0,013
AIM055	S-3VIKH	3	rs17203	CG	0,785	0,819	0,133	0,034	0,652	0,686	[4]	0,365	0,161	0,003
AIM056	S-4FJSI	3	rs13097560	GA	0,932	0,954	0,342	0,022	0,590	0,612	[1]	0,421	0,444	-0,007
AIM057	S-3OMXJ	3	rs7644167	CT	0,296	0,351	0,921	0,055	0,625	0,570	[1]	0,662	0,501	0,083
AIM058	S-3UDNS	3	rs2317212	TG	0,660	0,927	0,122	0,267	0,538	0,805	[1]	1,000	0,134	0,031
AIM059	S-4KQCI	3	rs7630522	TC	0,738	0,942	0,213	0,204	0,525	0,729	[1]	0,652	0,386	-0,010
AIM060	S-4BIVN	3	rs4855697	AG	0,913	0,011	0,597	0,902	0,316	0,586	[1]	0,401	0,844	-0,009
AIM061	S-4FPNO	3	rs7630111	AC	0,117	0,186	0,756	0,069	0,639	0,570	[1]	0,684	0,833	0,014
AIM062	S-3MRWI	3	rs2937673	GA	0,044	0,019	0,553	0,025	0,509	0,534	[1]	0,379	0,850	0,007
AIM063	S-4ABFO	3	rs16832787	GA	0,947	0,095	0,823	0,852	0,124	0,728	[1]	1,000	0,353	0,003
AIM064	S-3HHVO	3	rs6791346	TC	0,917	0,000	0,151	0,917	0,766	0,151	[1]	0,345	0,809	-0,004
AIM065	S-4GVKR	3	rs2855557	AT	0,029	0,725	0,849	0,696	0,820	0,124	[1]	0,652	0,809	0,013
AIM066	S-4AQTD	3	rs1586861	TC	0,863	0,046	0,899	0,817	0,036	0,853	[1]	1,000	0,325	-0,012
AIM067	S-4KKKL	3	rs9832471	CA	0,811	0,008	0,057	0,803	0,754	0,049	[1]	1,000	1,000	-0,003
AIM068	S-4MLOY	3	rs734873	AG	0,030	0,560	0,090	0,530	0,060	0,470	[2]	1,000	1,000	-0,013
AIM069	S-4HTQS	3	rs2106124	GA	0,155	0,031	0,739	0,124	0,584	0,708	[1]	1,000	0,161	0,075
AIM070	S-3OCTL	3	rs6806083	GA	0,995	0,733	0,254	0,262	0,741	0,479	[1]	0,650	0,404	-0,012
AIM071	S-3YHCB	3	rs715790	TC	0,846	0,425	0,580	0,421	0,266	0,155	[3]	0,652	0,432	-0,010
AIM072	S-4IRHV	3	rs13327370	AG	0,078	0,989	0,770	0,911	0,692	0,219	[1]	0,597	0,436	-0,005
AIM073	S-4PNLP	4	rs9291090	AC	0,970	0,540	0,940	0,430	0,030	0,400	[2]	1,000	1,000	0,007
AIM074	S-3GBRY	4	rs7665516	AG	0,112	0,977	0,363	0,865	0,251	0,614	[1]	0,676	1,000	-0,014
AIM075	S-3EKKC	4	rs719776	CG	0,961	0,225	0,119	0,736	0,842	0,106	[1]	0,662	0,779	0,032
AIM076	S-4TCHU	4	rs11051	GA	0,971	0,080	0,262	0,891	0,709	0,182	[1]	0,662	1,000	-0,012
AIM077	S-4PGVC	4	rs12650513	GA	0,976	0,235	0,958	0,741	0,018	0,723	[1]	1,000	1,000	0,024
AIM078	S-3AGID	4	rs1369093	CT	0,330	0,720	0,050	0,390	0,280	0,670	[2]	0,568	0,688	0,040
AIM079	S-3UTQS	4	rs6855458	AG	0,035	0,134	0,654	0,099	0,619	0,520	[1]	1,000	0,565	0,007
AIM080	S-4SHUP	4	rs4693573	TC	0,112	0,989	0,829	0,877	0,717	0,160	[1]	1,000	0,778	0,054

AIM081	S-4FTBC	4	rs17014118	TC	0,155	0,317	0,848	0,162	0,693	0,531	[1]	0,220	0,448	0,091
AIM082	S-4ERLK	4	rs2089990	CA	0,053	0,919	0,181	0,866	0,128	0,738	[1]	0,035	1,000	-0,014
AIM083	S-4STTN	4	rs2851060	CT	0,102	0,927	0,229	0,825	0,127	0,698	[1]	1,000	0,169	-0,014
AIM084	S-3ZAAW	4	rs920559	GC	0,789	0,725	0,144	0,064	0,645	0,581	[1]	0,652	0,228	0,015
AIM085	S-3MKBE	4	rs6533426	AG	0,141	0,988	0,853	0,847	0,712	0,135	[1]	0,071	0,592	-0,015
AIM086	S-3NMRB	4	rs4541508	TC	0,917	0,866	0,320	0,051	0,597	0,546	[1]	0,379	1,000	0,061
AIM087	S-3CAYG	4	rs7664927	GA	0,180	0,893	0,050	0,713	0,130	0,843	[1]	1,000	0,208	0,051
AIM088	S-4QKLN	4	rs17015014	GC	0,947	0,115	0,869	0,832	0,078	0,754	[1]	0,208	1,000	0,112
AIM089	S-3UNFA	4	rs1391099	AG	0,291	0,174	0,848	0,117	0,557	0,674	[1]	0,619	0,260	-0,014
AIM090	S-3GVIM	4	rs7664076	GA	0,966	0,656	0,197	0,310	0,769	0,459	[1]	0,674	0,071	0,024
AIM091	S-3UNCT	4	rs17090615	CT	0,782	0,626	0,120	0,156	0,662	0,506	[1]	1,000	0,639	-0,013
AIM092	S-4LMNW	4	rs2702414	GA	0,900	0,290	0,940	0,610	0,040	0,650	[2]	1,000	0,346	-0,014
AIM093	S-4TIYF	4	rs3911610	CT	0,388	0,347	0,936	0,041	0,548	0,589	[1]	0,544	0,584	-0,010
AIM094	S-3UXOS	5	rs10071261	CT	0,097	0,893	0,914	0,796	0,817	0,021	[1]	1,000	0,205	0,041
AIM095	S-4BYDP	5	rs316598	CT	0,990	0,430	0,270	0,560	0,720	0,160	[2]	1,000	0,154	0,051
AIM096	S-4CZCY	5	rs1875017	AG	0,301	0,286	0,894	0,015	0,593	0,608	[1]	0,544	1,000	-0,013
AIM097	S-3AVTC	5	rs10042283	GT	0,956	0,931	0,337	0,025	0,619	0,594	[1]	1,000	0,441	-0,014
AIM098	S-3HETS	5	rs10068162	TC	0,995	0,206	0,876	0,789	0,119	0,670	[1]	0,538	0,469	-0,001
AIM099	S-3MJAQ	5	rs10461813	AG	0,951	0,046	0,237	0,905	0,714	0,191	[1]	0,662	0,639	0,001
AIM100	S-4DQSN	5	rs4867557	TC	0,801	0,874	0,239	0,073	0,562	0,635	[1]	0,367	0,536	-0,011
AIM101	S-4AYSM	5	rs37369	CT	0,330	0,410	0,910	0,080	0,580	0,500	[2]	0,279	0,309	0,001
AIM102	S-3BATE	5	rs1125508	CT	0,243	0,750	0,690	0,507	0,447	0,060	[3]	0,662	0,533	-0,014
AIM103	S-4EFUK	5	rs353373	AG	0,010	0,809	0,137	0,799	0,127	0,672	[1]	1,000	0,283	0,001
AIM104	S-3DDMX	5	rs3846662	AG	0,036	0,525	0,579	0,489	0,543	0,054	[3]	0,198	0,560	-0,009
AIM105	S-3GBAB	5	rs349323	AT	0,985	0,969	0,371	0,016	0,614	0,598	[1]	0,674	1,000	-0,004
AIM106	S-3YBEY	5	rs6866970	AG	0,142	0,985	0,940	0,843	0,798	0,045	[1]	0,107	0,044	-0,005
AIM107	S-3YDUJ	5	rs17157450	CT	0,903	0,145	0,864	0,758	0,039	0,719	[1]	0,609	1,000	0,018
AIM108	S-3UUFH	5	rs13354224	TC	0,097	0,992	0,889	0,895	0,792	0,103	[1]	0,251	0,526	-0,012
AIM109	S-4BXZC	5	rs10079352	GA	0,039	0,969	0,403	0,930	0,364	0,566	[1]	0,654	0,171	-0,002

AIM110	S-4QEPN	5	rs2227282	CG	0,995	0,897	0,270	0,098	0,725	0,627	[1]	1,000	1,000	-0,012
AIM111	S-3DESH	5	rs7732591	TC	0,010	0,687	0,852	0,677	0,842	0,165	[1]	1,000	0,794	0,077
AIM112	S-3TXPY	5	rs4145160	GA	0,850	0,127	0,953	0,723	0,103	0,826	[1]	0,561	1,000	0,058
AIM113	S-3OZWN	5	rs11960137	CG	0,942	0,122	0,875	0,820	0,067	0,753	[1]	0,568	1,000	-0,008
AIM114	S-3KSYA	5	rs6556352	CT	0,020	0,520	0,730	0,500	0,710	0,210	[2]	1,000	0,103	0,014
AIM115	S-3DHUZ	5	rs1107154	GA	0,015	0,939	0,610	0,924	0,595	0,329	[1]	0,421	0,239	0,014
AIM116	S-3RQMY	5	rs11743823	TC	0,053	0,897	0,186	0,844	0,133	0,711	[1]	0,617	1,000	-0,010
AIM117	S-3BZJF	6	rs2504853	TC	0,030	0,420	0,650	0,390	0,620	0,230	[2]	0,670	1,000	0,000
AIM118	S-3HYNA	6	rs7745461	GA	0,660	0,240	0,850	0,420	0,190	0,610	[2]	1,000	0,398	-0,013
AIM119	S-3LXMG	6	rs2754777	CA	0,995	0,973	0,484	0,022	0,511	0,489	[1]	1,000	0,109	-0,013
AIM120	S-3IKDM	6	rs9357427	AG	0,083	0,943	0,798	0,860	0,715	0,145	[1]	1,000	0,818	0,027
AIM121	S-4FRWJ	6	rs559035	GT	0,464	0,552	0,853	0,088	0,389	0,301	[3]	0,127	0,818	-0,012
AIM122	S-4BGBD	6	rs2397060	TC	0,200	0,910	0,880	0,710	0,680	0,030	[2]	0,643	0,341	-0,007
AIM123	S-3BJFX	6	rs1325476	TC	0,165	0,424	0,921	0,259	0,756	0,497	[1]	0,662	1,000	0,078
AIM124	S-4NVYP	6	rs718268	CT	0,236	0,568	0,316	0,332	0,080	0,252	[3]	1,000	0,344	-0,009
AIM125	S-3XKMR	6	rs7752055	CT	0,893	0,038	0,119	0,855	0,774	0,081	[1]	1,000	0,783	0,059
AIM126	S-3JXIH	6	rs2503770	TA	0,898	0,779	0,235	0,119	0,663	0,544	[1]	0,684	0,400	0,007
AIM127	S-3DSTZ	6	rs9384878	CT	0,848	0,015	0,649	0,833	0,199	0,634	[1]	0,091	1,000	0,030
AIM128	S-3NCFG	6	rs218862	CT	0,097	0,153	0,837	0,056	0,740	0,684	[1]	0,171	0,670	-0,011
AIM129	S-4RKTA	6	rs512342	CG	0,053	0,977	0,333	0,924	0,280	0,644	[1]	0,626	0,328	0,015
AIM130	S-3PWEY	6	rs9689446	CT	0,811	0,008	0,059	0,803	0,752	0,051	[1]	1,000	0,458	0,017
AIM131	S-4GWCP	6	rs9321456	CT	0,840	0,977	0,340	0,137	0,500	0,637	[1]	0,421	0,703	-0,008
AIM132	S-4IESO	6	rs7750326	AG	0,078	0,038	0,746	0,040	0,668	0,708	[1]	0,676	0,688	0,005
AIM133	S-4EHSQ	6	rs2065664	TC	0,099	0,039	0,639	0,060	0,540	0,600	[1]	0,220	0,244	-0,014
AIM134	S-4TBIW	6	rs4598087	GA	0,932	0,027	0,241	0,905	0,691	0,214	[1]	0,411	0,484	0,004
AIM135	S-4KMPS	6	rs7771341	CG	0,015	0,851	0,149	0,836	0,134	0,702	[1]	1,000	0,683	0,006
AIM136	S-4BCKO	6	rs6605534	GA	0,874	0,058	0,855	0,816	0,019	0,797	[1]	1,000	0,520	-0,005
AIM137	S-3EEVN	7	rs731257	AG	0,024	0,840	0,120	0,816	0,096	0,720	[1]	0,098	1,000	-0,013
AIM138	S-3GBUR	7	rs2080161	AC	0,976	0,099	0,798	0,877	0,178	0,699	[1]	1,000	1,000	-0,004

AIM139	S-3GLXC	7	rs7808899	GT	0,854	0,702	0,170	0,152	0,684	0,532	[1]	1,000	0,801	0,006
AIM140	S-4MLCV	7	rs2106362	GC	0,868	0,713	0,181	0,155	0,687	0,532	[1]	0,662	0,335	0,001
AIM141	S-4MDOL	7	rs7794503	TC	0,039	0,992	0,772	0,953	0,733	0,220	[1]	0,401	0,278	0,024
AIM142	S-4FPEZ	7	rs2598391	GA	0,039	0,996	0,700	0,957	0,661	0,296	[1]	0,676	1,000	0,021
AIM143	S-4OIJB	7	rs37268	CT	0,942	0,103	0,222	0,839	0,720	0,119	[1]	1,000	0,477	0,076
AIM144	S-4NVAH	7	rs4717865	GA	0,990	0,360	0,890	0,630	0,100	0,530	[2]	1,000	1,000	0,011
AIM145	S-4QWQC	7	rs11763970	AG	0,976	0,966	0,453	0,010	0,523	0,513	[1]	0,670	0,847	-0,013
AIM146	S-3TOPP	7	rs2384982	CT	0,398	0,187	0,938	0,211	0,540	0,751	[1]	1,000	0,204	-0,009
AIM147	S-4DCUC	7	rs7797239	CT	0,141	0,308	0,826	0,167	0,685	0,518	[1]	0,647	0,464	0,004
AIM148	S-4LASG	7	rs2299542	AG	0,864	0,031	0,644	0,833	0,220	0,613	[1]	0,387	0,680	-0,012
AIM149	S-3OPQX	7	rs350655	AC	0,888	0,011	0,097	0,877	0,791	0,086	[1]	1,000	0,346	-0,011
AIM150	S-3PKJC	7	rs7803075	GA	0,010	0,100	0,710	0,090	0,700	0,610	[2]	1,000	0,842	-0,013
AIM151	S-3KSUQ	7	rs2707575	GA	0,034	0,031	0,574	0,003	0,540	0,543	[1]	1,000	0,186	0,056
AIM152	S-3ELIA	7	rs6464211	CT	0,170	0,740	0,890	0,570	0,720	0,150	[2]	0,061	0,810	0,003
AIM153	S-3DXTR	8	rs11784513	TC	0,932	0,130	0,174	0,802	0,758	0,044	[1]	0,662	0,230	0,004
AIM154	S-4FHXX	8	rs10108270	AC	0,970	0,030	0,350	0,940	0,620	0,320	[2]	0,670	1,000	0,002
AIM155	S-4COUW	8	rs2924720	GA	0,112	0,946	0,305	0,834	0,193	0,641	[1]	1,000	0,518	-0,011
AIM156	S-4NGRU	8	rs1454284	CA	0,493	0,974	0,409	0,481	0,084	0,565	[3]	0,093	0,086	-0,008
AIM157	S-3FAGC	8	rs1039917	GA	0,978	0,487	0,625	0,491	0,353	0,138	[3]	0,279	0,665	0,004
AIM158	S-3FHVP	8	rs366178	AC	0,150	0,969	0,258	0,819	0,108	0,711	[1]	0,652	0,659	-0,002
AIM159	S-3LYSF	8	rs17150066	GA	0,985	0,088	0,858	0,897	0,127	0,770	[1]	1,000	0,761	0,005
AIM160	S-4MLSN	8	rs2736340	CT	0,879	0,089	0,792	0,790	0,087	0,703	[1]	1,000	0,663	-0,009
AIM161	S-4KMWQ	8	rs11778591	AC	0,418	0,057	0,908	0,361	0,490	0,851	[1]	0,186	0,604	0,071
AIM162	S-3BTMX	8	rs6586662	TA	0,849	0,897	0,276	0,048	0,573	0,621	[1]	0,228	0,842	0,010
AIM163	S-3LVKB	8	rs4871930	AG	0,078	0,977	0,751	0,899	0,673	0,226	[1]	0,093	0,268	-0,005
AIM164	S-3LPGZ	8	rs1436612	CG	0,898	0,011	0,178	0,887	0,720	0,167	[1]	1,000	0,779	-0,014
AIM165	S-3EBTA	8	rs956969	TC	0,049	0,034	0,623	0,015	0,574	0,589	[1]	0,421	0,571	-0,013
AIM166	S-3LBOB	8	rs3176921	AG	0,237	0,983	0,929	0,746	0,692	0,054	[5]	0,251	0,346	-0,011
AIM167	S-3UDQE	8	rs16880715	TG	0,859	0,019	0,091	0,840	0,768	0,072	[1]	1,000	0,759	0,003

AIM168	S-3HYGN	8	rs13280988	AG	0,058	0,031	0,581	0,027	0,523	0,550	[1]	0,674	0,054	-0,003
AIM169	S-3CVIA	8	rs9642819	GA	0,806	0,031	0,731	0,775	0,075	0,700	[1]	1,000	1,000	0,031
AIM170	S-4LUBH	8	rs7017679	AG	0,602	0,771	0,092	0,169	0,510	0,679	[1]	1,000	0,585	-0,010
AIM171	S-3PEZK	8	rs799889	TG	0,113	0,103	0,830	0,010	0,717	0,727	[1]	0,670	0,639	0,033
AIM172	S-3EGTE	8	rs2450219	CT	0,942	0,943	0,359	0,001	0,583	0,584	[1]	0,676	0,550	0,071
AIM173	S-3OKIQ	8	rs1357038	AT	0,874	0,000	0,161	0,874	0,713	0,161	[1]	1,000	1,000	-0,011
AIM174	S-4PXSD	8	rs7006331	CT	0,049	0,050	0,599	0,001	0,550	0,549	[1]	0,091	0,700	-0,013
AIM175	S-3QJGB	8	rs1876201	CT	0,151	0,095	0,770	0,056	0,619	0,675	[1]	0,220	1,000	-0,006
AIM176	S-3ETIE	8	rs2001907	TC	0,050	0,590	0,100	0,540	0,050	0,490	[2]	0,310	0,117	0,002
AIM177	S-4OBXF	8	rs10111852	TC	0,015	0,023	0,590	0,008	0,575	0,567	[1]	0,220	1,000	-0,009
AIM178	S-3UARE	9	rs10756267	GA	0,990	0,118	0,888	0,872	0,102	0,770	[1]	0,215	1,000	-0,007
AIM179	S-4NBBM	9	rs791654	AG	0,376	0,168	0,885	0,208	0,509	0,717	[1]	0,178	0,625	0,003
AIM180	S-3FUFN	9	rs10810942	GA	0,794	0,053	0,846	0,741	0,052	0,793	[1]	1,000	0,324	-0,005
AIM181	S-4NWLRL	9	rs10811305	GC	0,976	0,062	0,850	0,914	0,126	0,788	[1]	0,299	0,556	0,007
AIM182	S-3ZQSV	9	rs3780827	CG	0,704	0,618	0,077	0,086	0,627	0,541	[1]	0,128	0,791	-0,011
AIM183	S-4ICXB	9	rs3793451	TC	0,030	0,450	0,050	0,420	0,020	0,400	[2]	1,000	0,688	-0,003
AIM184	S-4MSDX	9	rs10981899	GT	0,985	0,181	0,897	0,804	0,088	0,716	[1]	0,251	0,691	0,008
AIM185	S-4GTKX	9	rs2809393	CT	0,350	0,103	0,867	0,247	0,517	0,764	[1]	0,345	0,815	-0,009
AIM186	S-4DNHJ	9	rs9032	CG	0,064	0,415	0,904	0,351	0,840	0,489	[3]	0,405	0,470	0,045
AIM187	S-3OQTJ	10	rs10751867	CT	0,825	0,649	0,084	0,176	0,741	0,565	[1]	0,171	0,260	0,049
AIM188	S-3JJDD	10	rs4881046	AT	0,122	0,985	0,882	0,863	0,760	0,103	[1]	1,000	0,743	0,009
AIM189	S-4NULN	10	rs4747860	CT	0,869	0,920	0,311	0,051	0,558	0,609	[1]	0,411	0,113	0,080
AIM190	S-3JJKH	10	rs4749305	GA	0,655	0,865	0,141	0,210	0,514	0,724	[1]	1,000	1,000	-0,011
AIM191	S-3CMAH	10	rs3123687	CG	0,966	0,027	0,072	0,939	0,894	0,045	[1]	1,000	0,325	-0,009
AIM192	S-3NVQM	10	rs4948800	TC	0,937	0,954	0,410	0,017	0,527	0,544	[1]	0,668	0,700	-0,005
AIM193	S-3DKJI	10	rs709616	GA	0,136	0,973	0,929	0,837	0,793	0,044	[1]	1,000	0,007	0,020
AIM194	S-4FKKX	10	rs2795918	CT	0,150	0,954	0,308	0,804	0,158	0,646	[1]	0,643	1,000	-0,003
AIM195	S-4IFJS	10	rs7911953	TG	0,884	0,157	0,906	0,727	0,022	0,749	[1]	0,540	0,021	-0,012
AIM196	S-3MVIK	10	rs11000445	TC	0,180	0,989	0,944	0,809	0,764	0,045	[1]	0,524	0,242	-0,013

AIM197	S-4GOVR	10	rs4746136	GA	0,990	0,220	0,870	0,770	0,120	0,650	[2]	1,000	0,687	-0,012
AIM198	S-3DMNR	10	rs7097617	GA	0,966	0,076	0,194	0,890	0,772	0,118	[1]	0,654	0,510	-0,012
AIM199	S-3DQTB	10	rs7902158	GA	0,961	0,953	0,237	0,008	0,724	0,716	[1]	0,684	0,049	0,073
AIM200	S-4BXWZ	10	rs11186960	AC	0,039	0,008	0,564	0,031	0,525	0,556	[1]	0,652	0,124	0,007
AIM201	S-3GBZE	10	rs4244304	CT	0,981	0,164	0,897	0,817	0,084	0,733	[1]	1,000	1,000	-0,004
AIM202	S-4IXCW	10	rs4918326	TC	0,053	0,939	0,262	0,886	0,209	0,677	[1]	1,000	0,611	-0,014
AIM203	S-4TBYJ	10	rs4918842	CT	0,090	0,930	0,150	0,840	0,060	0,780	[2]	0,215	0,260	0,020
AIM204	S-3MYFW	10	rs7087634	TC	0,136	0,313	0,901	0,177	0,765	0,588	[1]	0,208	0,013	0,123
AIM205	S-3JUAC	10	rs1419138	CT	0,888	0,691	0,081	0,197	0,807	0,610	[1]	0,670	0,403	0,075
AIM206	S-4NURC	10	rs12358575	CT	0,869	0,893	0,276	0,024	0,593	0,617	[1]	1,000	0,423	0,025
AIM207	S-4OKYQ	10	rs10764919	GA	0,981	0,954	0,430	0,027	0,551	0,524	[1]	0,182	0,564	0,014
AIM208	S-4IQCJ	11	rs10741584	GA	0,879	0,023	0,133	0,856	0,746	0,110	[1]	1,000	0,760	-0,002
AIM209	S-3OWPU	11	rs10833134	TC	0,078	0,859	0,950	0,781	0,872	0,091	[1]	0,128	0,596	0,095
AIM210	S-3WUZT	11	rs10741838	CT	0,485	0,154	0,933	0,331	0,448	0,779	[1]	0,652	0,736	0,098
AIM211	S-4OMKD	11	rs11029687	AG	0,024	0,118	0,663	0,094	0,639	0,545	[1]	0,401	1,000	0,033
AIM212	S-3ILNH	11	rs7927234	GC	0,148	0,988	0,890	0,840	0,742	0,098	[1]	0,342	0,063	0,126
AIM213	S-3ACFG	11	rs2993051	TC	0,180	0,930	0,225	0,750	0,045	0,705	[1]	1,000	0,651	-0,005
AIM214	S-3IUSL	11	rs11034734	TC	0,990	0,981	0,488	0,009	0,502	0,493	[1]	0,251	0,695	0,058
AIM215	S-4NAWE	11	rs10437653	CA	0,937	0,019	0,563	0,918	0,374	0,544	[1]	0,675	1,000	-0,005
AIM216	S-4IMNG	11	rs1790740	GA	0,830	0,775	0,071	0,055	0,759	0,704	[1]	0,182	1,000	0,084
AIM217	S-4STOT	11	rs7111814	TC	0,034	0,920	0,713	0,886	0,679	0,207	[1]	0,676	0,677	-0,011
AIM218	S-4CAVG	11	rs568789	TC	0,995	0,076	0,242	0,919	0,753	0,166	[1]	1,000	0,298	-0,006
AIM219	S-3HOUX	11	rs1042602	CA	0,005	0,034	0,449	0,029	0,444	0,415	[4]	1,000	1,000	-0,005
AIM220	S-3GJKY	11	rs282496	AG	0,457	0,958	0,728	0,501	0,271	0,230	[3]	1,000	0,468	-0,009
AIM221	S-3SGHW	11	rs520345	TA	0,947	0,031	0,294	0,916	0,653	0,263	[1]	0,405	0,281	0,012
AIM222	S-3VXOM	11	rs236919	TC	0,893	0,763	0,379	0,130	0,514	0,384	[3]	1,000	0,847	-0,009
AIM223	S-3OULB	11	rs10893032	GA	0,005	0,866	0,649	0,861	0,644	0,217	[1]	0,421	1,000	-0,001
AIM224	S-4GJUN	11	rs2510719	CA	0,961	0,966	0,426	0,005	0,535	0,540	[1]	1,000	0,130	-0,011
AIM225	S-3YFJC	11	rs4937920	TC	0,005	0,805	0,085	0,800	0,080	0,720	[1]	1,000	0,691	-0,012

AIM226	S-3PKLK	12	rs10848644	TC	0,019	0,015	0,533	0,004	0,514	0,518	[1]	0,674	0,700	-0,013
AIM227	S-4PFWF	12	rs7961436	GA	0,084	0,057	0,640	0,027	0,556	0,583	[1]	1,000	1,000	-0,014
AIM228	S-4HBWV	12	rs11048128	CG	0,201	0,977	0,226	0,776	0,025	0,751	[1]	0,626	0,409	-0,004
AIM229	S-3RDPA	12	rs1486341	CA	0,044	0,298	0,852	0,254	0,808	0,554	[1]	1,000	0,611	0,085
AIM230	S-4GCVW	12	rs2193662	TG	0,157	0,485	0,930	0,328	0,773	0,445	[1]	0,544	0,161	-0,011
AIM231	S-3WTRO	12	rs214678	CT	0,070	0,640	0,150	0,570	0,080	0,490	[2]	0,561	0,416	0,025
AIM232	S-4GOOV	12	rs10735825	TC	0,108	0,489	0,928	0,381	0,820	0,439	[1]	1,000	1,000	0,004
AIM233	S-3YAPY	12	rs10783486	AG	0,199	0,954	0,261	0,755	0,062	0,693	[1]	0,654	1,000	-0,014
AIM234	S-3VUEQ	12	rs893280	CT	0,927	0,004	0,255	0,923	0,672	0,251	[1]	1,000	1,000	0,009
AIM235	S-4CXIY	12	rs10492199	CT	0,020	0,950	0,313	0,930	0,293	0,637	[1]	0,676	0,840	-0,013
AIM236	S-3UWCQ	12	rs3759171	AG	0,015	0,034	0,732	0,019	0,717	0,698	[1]	0,010	0,840	-0,012
AIM237	S-3PYEC	12	rs1795504	GC	0,809	0,683	0,135	0,126	0,674	0,548	[1]	1,000	0,448	0,013
AIM238	S-4KLDY	12	rs10466960	AC	0,150	0,977	0,914	0,827	0,764	0,063	[1]	1,000	1,000	0,047
AIM239	S-3MFTO	12	rs10732643	GT	0,121	0,989	0,930	0,868	0,809	0,059	[1]	1,000	0,394	0,068
AIM240	S-4JLLT	12	rs11113995	CT	0,364	0,309	0,924	0,055	0,560	0,615	[1]	1,000	0,692	0,064
AIM241	S-4KRHA	12	rs4766807	AT	0,025	0,948	0,622	0,923	0,597	0,326	[5]	0,689	0,335	-0,005
AIM242	S-3RIBS	12	rs2694874	GC	0,211	0,038	0,734	0,173	0,523	0,696	[1]	0,220	0,153	0,045
AIM243	S-3WDCN	13	rs2585897	AG	0,116	0,924	0,143	0,808	0,027	0,781	[1]	1,000	0,779	-0,008
AIM244	S-3UIIC	13	rs7990216	CT	0,917	0,729	0,102	0,188	0,815	0,627	[1]	0,270	1,000	-0,009
AIM245	S-4ODKX	13	rs394307	AG	0,068	0,023	0,581	0,045	0,513	0,558	[1]	1,000	0,256	-0,012
AIM246	S-3BMKD	13	rs7328696	AG	0,709	0,580	0,042	0,129	0,667	0,538	[1]	1,000	0,734	-0,001
AIM247	S-3LLPQ	13	rs1337973	AG	0,903	0,191	0,946	0,712	0,043	0,755	[1]	1,000	1,000	-0,007
AIM248	S-4PJLQ	13	rs2151236	AT	0,024	0,915	0,852	0,891	0,828	0,063	[1]	0,365	1,000	0,007
AIM249	S-3QNJA	13	rs9533623	GA	0,990	0,950	0,448	0,040	0,542	0,502	[1]	0,386	1,000	-0,009
AIM250	S-3HTTK	13	rs9596011	GA	0,903	0,078	0,115	0,825	0,788	0,037	[1]	0,293	0,099	0,037
AIM251	S-3BRFX	13	rs9527650	GA	0,199	0,966	0,320	0,767	0,121	0,646	[1]	0,214	0,228	-0,011
AIM252	S-3IFOU	13	rs1924381	TC	0,922	0,649	0,133	0,273	0,789	0,516	[1]	0,063	0,165	-0,013
AIM253	S-3ELFL	13	rs4885162	TC	0,976	0,137	0,305	0,839	0,671	0,168	[1]	0,105	0,097	0,054
AIM254	S-3IAZY	13	rs9530435	CT	0,070	0,950	0,790	0,880	0,720	0,160	[2]	0,619	0,301	-0,001

AIM255	S-4SIUA	13	rs9523747	TA	0,794	0,969	0,227	0,175	0,567	0,742	[1]	0,399	1,000	0,022
AIM256	S-4KTCE	13	rs320075	TG	0,319	0,000	0,000	0,319	0,319	0,000	[3]	1,000	1,000	0,084
AIM257	S-4AUJA	13	rs2476230	CG	0,822	0,710	0,175	0,112	0,647	0,535	[1]	0,386	0,335	0,040
AIM258	S-4LLPT	14	rs1950420	AC	0,025	0,893	0,678	0,868	0,653	0,215	[1]	0,379	0,563	-0,006
AIM259	S-4LFDU	14	rs7151991	AG	0,131	0,935	0,188	0,804	0,057	0,747	[1]	1,000	0,759	0,013
AIM260	S-4OYDQ	14	rs1958053	AT	0,211	0,977	0,308	0,766	0,097	0,669	[1]	1,000	1,000	-0,011
AIM261	S-3ZKNB	14	rs1244379	CT	0,063	0,088	0,635	0,025	0,572	0,547	[1]	0,207	0,177	-0,005
AIM262	S-4IXHQ	14	rs11625446	AC	0,078	0,957	0,231	0,879	0,153	0,726	[1]	0,365	0,208	-0,004
AIM263	S-4OAFV	14	rs2357442	AC	0,760	0,270	0,870	0,490	0,110	0,600	[2]	1,000	1,000	-0,012
AIM264	S-4CROB	14	rs1188184	GA	0,073	0,031	0,658	0,042	0,585	0,627	[1]	0,405	0,701	0,006
AIM265	S-3YVDA	14	rs928108	AG	0,102	0,920	0,268	0,818	0,166	0,652	[1]	0,662	0,666	-0,011
AIM266	S-3PKOS	14	rs1889720	CT	0,853	0,008	0,072	0,845	0,781	0,064	[1]	0,279	0,113	0,058
AIM267	S-4DQQP	14	rs1950993	TG	0,980	0,720	0,370	0,260	0,610	0,350	[2]	0,074	0,344	0,010
AIM268	S-3XGTN	14	rs2296274	AG	0,039	0,927	0,802	0,888	0,763	0,125	[1]	0,670	0,507	0,011
AIM269	S-4GEHY	14	rs2251244	AC	0,015	0,283	0,819	0,268	0,804	0,536	[1]	0,643	0,822	-0,013
AIM270	S-4GWLH	14	rs10145908	CT	0,830	0,011	0,086	0,819	0,744	0,075	[1]	1,000	1,000	-0,013
AIM271	S-3NWPW	14	rs6573746	CA	0,257	0,488	0,944	0,231	0,687	0,456	[1]	1,000	0,687	0,083
AIM272	S-3YPFT	14	rs12435594	CG	0,063	0,927	0,295	0,864	0,232	0,632	[1]	1,000	0,196	-0,012
AIM273	S-4GUQY	14	rs7142344	CT	0,024	0,969	0,859	0,945	0,835	0,110	[1]	1,000	0,409	-0,013
AIM274	S-4ELDX	14	rs9323913	TC	0,134	0,950	0,307	0,816	0,173	0,643	[1]	1,000	0,837	-0,012
AIM275	S-4DBRH	14	rs200354	TG	0,680	0,870	0,170	0,190	0,510	0,700	[2]	0,399	0,230	0,038
AIM276	S-4DRMM	14	rs12434632	TC	0,024	0,855	0,206	0,831	0,182	0,649	[1]	0,544	0,562	-0,013
AIM277	S-4TGQG	14	rs11160805	TC	0,019	0,912	0,186	0,893	0,167	0,726	[1]	1,000	0,393	-0,010
AIM278	S-3NHFY	14	rs4983425	GA	0,820	0,054	0,050	0,766	0,770	0,004	[1]	0,609	0,683	0,054
AIM279	S-3CXKQ	15	rs7164838	GA	0,976	0,080	0,702	0,896	0,274	0,622	[1]	1,000	1,000	0,011
AIM280	S-4FMVW	15	rs8041147	CT	0,777	0,908	0,098	0,131	0,679	0,810	[1]	0,609	0,587	-0,012
AIM281	S-3WMSM	15	rs735480	CT	0,985	0,683	0,106	0,302	0,879	0,577	[1]	0,626	1,000	0,003
AIM282	S-3SDJD	15	rs1153858	TC	0,864	0,977	0,265	0,113	0,599	0,712	[1]	0,405	0,001	0,081
AIM283	S-4HPVA	15	rs1453858	AT	0,772	0,794	0,089	0,022	0,683	0,705	[1]	0,561	0,791	-0,014

AIM284	S-4NOVE	15	rs4646	AC	0,332	0,741	0,296	0,409	0,036	0,445	[4]	0,182	0,620	0,006
AIM285	S-3ZYNR	15	rs728244	GA	0,068	0,943	0,849	0,875	0,781	0,094	[1]	0,689	0,808	0,097
AIM286	S-3SGHO	15	rs972801	CT	0,843	0,763	0,219	0,080	0,624	0,544	[3]	1,000	0,834	0,015
AIM287	S-3EGKU	15	rs7176730	CT	0,879	0,019	0,580	0,860	0,299	0,561	[1]	0,014	0,124	0,005
AIM288	S-4RNQL	15	rs2415047	AC	0,131	0,351	0,896	0,220	0,765	0,545	[1]	1,000	0,785	-0,007
AIM289	S-3IERW	15	rs746655	AG	0,854	0,004	0,122	0,850	0,732	0,118	[1]	1,000	0,736	-0,007
AIM290	S-3TNMG	15	rs901130	GC	0,961	0,866	0,303	0,095	0,658	0,563	[1]	1,000	0,153	-0,007
AIM291	S-4FQEJ	15	rs7178655	TG	0,898	0,908	0,278	0,010	0,620	0,630	[1]	0,228	0,840	0,017
AIM292	S-4HUYE	15	rs11631761	AG	0,898	0,034	0,614	0,864	0,284	0,580	[1]	0,684	0,318	-0,006
AIM293	S-4HYBY	15	rs4402506	TC	0,985	0,950	0,412	0,035	0,573	0,538	[1]	0,674	0,441	-0,003
AIM294	S-4EFYD	15	rs7173885	CG	0,083	0,992	0,751	0,909	0,668	0,241	[1]	1,000	1,000	-0,007
AIM295	S-4IXOF	16	rs4787040	AT	0,961	0,920	0,286	0,041	0,675	0,634	[1]	0,421	0,686	0,011
AIM296	S-3MGVJ	16	rs9937557	CG	0,927	0,202	0,940	0,725	0,013	0,738	[1]	1,000	1,000	-0,010
AIM297	S-4QMUX	16	rs7185307	TC	0,121	0,088	0,751	0,033	0,630	0,663	[1]	0,093	0,386	0,012
AIM298	S-3ZTQU	16	rs1650999	CT	0,307	0,364	0,964	0,057	0,657	0,600	[3]	1,000	0,416	0,043
AIM299	S-4OKWU	16	rs30237	AG	0,097	0,916	0,838	0,819	0,741	0,078	[1]	0,609	0,605	-0,014
AIM300	S-3IJGP	16	rs4347621	GA	0,903	0,065	0,152	0,838	0,751	0,087	[1]	0,015	1,000	-0,014
AIM301	S-3OGOR	16	rs1787781	AG	0,879	0,015	0,163	0,864	0,716	0,148	[1]	1,000	1,000	0,126
AIM302	S-4CWQY	16	rs10871351	GA	0,937	0,019	0,230	0,918	0,707	0,211	[1]	0,652	0,214	0,016
AIM303	S-3TOOO	16	rs8043824	AG	0,015	0,794	0,747	0,779	0,732	0,047	[1]	0,676	1,000	0,130
AIM304	S-3NPHS	16	rs2967340	AG	0,092	0,225	0,848	0,133	0,756	0,623	[1]	0,652	0,498	-0,001
AIM305	S-3LXVV	16	rs2966849	AG	0,310	0,690	0,110	0,380	0,200	0,580	[2]	1,000	0,737	-0,007
AIM306	S-4PCGB	16	rs2086824	CA	0,015	0,038	0,565	0,023	0,550	0,527	[1]	0,684	0,084	-0,008
AIM307	S-3XYUU	17	rs4791868	CG	0,874	0,859	0,063	0,015	0,811	0,796	[1]	0,162	0,072	0,026
AIM308	S-3KLTM	17	rs8068853	TC	0,034	0,092	0,716	0,058	0,682	0,624	[1]	0,676	1,000	0,014
AIM309	S-3NMFO	17	rs212498	AG	0,044	0,629	0,772	0,585	0,728	0,143	[3]	0,689	1,000	0,061
AIM310	S-4EMKO	17	rs8081410	AG	0,907	0,870	0,309	0,037	0,598	0,561	[1]	1,000	0,173	-0,010
AIM311	S-3SCJX	17	rs7211426	AG	0,951	0,092	0,212	0,859	0,739	0,120	[1]	0,652	0,605	0,027

AIM312	S-3PKQH	17	rs1426741	TG	0,879	0,683	0,133	0,196	0,746	0,550	[1]	1,000	1,000	0,024
AIM313	S-4DSGJ	17	rs2052074	GA	0,966	0,118	0,177	0,848	0,789	0,059	[1]	0,089	0,025	0,040
AIM314	S-3GXQJ	17	rs4790939	GC	0,908	0,107	0,115	0,801	0,793	0,008	[1]	0,171	0,469	0,115
AIM315	S-4HQGO	17	rs918077	GA	0,752	0,679	0,120	0,073	0,632	0,559	[1]	0,528	0,348	0,004
AIM316	S-4SXD	17	rs4789193	TA	0,049	0,935	0,809	0,886	0,760	0,126	[1]	0,674	0,822	0,030
AIM317	S-4MWBG	18	rs680173	GC	0,180	0,962	0,314	0,782	0,134	0,648	[1]	1,000	0,838	-0,003
AIM318	S-3XQGJ	18	rs4798812	AG	0,270	0,800	0,230	0,530	0,040	0,570	[2]	1,000	0,836	0,029
AIM319	S-3MISJ	18	rs628419	CT	0,005	0,112	0,641	0,107	0,636	0,529	[1]	0,684	0,848	0,017
AIM320	S-4GPAL	18	rs12185395	CT	0,966	0,981	0,421	0,015	0,545	0,560	[1]	0,399	1,000	-0,003
AIM321	S-3HFNR	18	rs1893441	TC	0,966	0,080	0,805	0,886	0,161	0,725	[1]	0,544	1,000	-0,010
AIM322	S-3UUSX	18	rs7238445	AG	0,920	0,290	0,190	0,630	0,730	0,100	[2]	0,162	0,257	-0,011
AIM323	S-4JKJX	18	rs667508	AG	0,329	0,797	0,821	0,468	0,492	0,024	[3]	0,355	0,111	-0,007
AIM324	S-4GRDM	19	rs8113143	AC	0,990	0,710	0,320	0,280	0,670	0,390	[2]	0,676	0,571	0,010
AIM325	S-3YKGD	19	rs4884	AG	0,160	0,814	0,287	0,654	0,127	0,527	[4]	0,063	0,684	-0,005
AIM326	S-4IEYO	20	rs2889678	CT	0,942	0,969	0,297	0,027	0,645	0,672	[1]	1,000	0,401	-0,007
AIM327	S-4KPTY	20	rs6088466	AG	0,019	0,966	0,304	0,947	0,285	0,662	[1]	0,093	0,151	-0,001
AIM328	S-3HOGL	20	rs6101991	AG	0,147	0,962	0,896	0,815	0,749	0,066	[1]	1,000	0,736	0,012
AIM329	S-4OZMZ	20	rs1201686	GC	0,403	0,237	0,903	0,166	0,500	0,666	[1]	0,643	0,274	0,020
AIM330	S-3YDPD	20	rs3907047	CT	0,010	0,630	0,060	0,620	0,050	0,570	[2]	0,008	1,000	-0,015
AIM331	S-4SUKK	20	rs731803	TC	0,180	0,981	0,398	0,801	0,218	0,583	[1]	1,000	1,000	-0,011
AIM332	S-3GJJW	20	rs172982	GA	0,900	0,500	0,546	0,400	0,354	0,046	[3]	0,609	0,298	0,003
AIM333	S-3REEG	20	rs1877751	GA	0,131	0,966	0,331	0,835	0,200	0,635	[1]	0,399	1,000	-0,013
AIM334	S-3BGDZ	21	rs2242865	TG	0,068	0,989	0,844	0,921	0,776	0,145	[1]	1,000	1,000	0,016
AIM335	S-3ADKZ	21	rs2835367	TC	0,175	0,256	0,899	0,081	0,724	0,643	[1]	0,643	0,470	-0,004
AIM336	S-4NHJR	21	rs2836181	CT	0,199	0,984	0,951	0,785	0,752	0,033	[1]	0,414	1,000	-0,012
AIM337	S-4IKIV	21	rs12329755	GC	0,010	0,981	0,610	0,971	0,600	0,371	[1]	0,674	0,432	0,037
AIM338	S-4KEFU	21	rs762421	AG	0,937	0,038	0,634	0,899	0,303	0,596	[1]	0,171	1,000	-0,013
AIM339	S-3YZLO	22	rs5748014	CA	0,233	0,057	0,752	0,176	0,519	0,695	[1]	0,670	1,000	-0,013

AIM340	S-3OFND	22	rs5754506	TC	0,142	0,992	0,865	0,850	0,723	0,127	[1]	0,128	0,116	0,003
AIM341	S-3NNGV	22	rs132663	TA	0,107	0,973	0,316	0,866	0,209	0,657	[1]	0,674	0,686	-0,005
AIM342	S-4PGKG	22	rs5757362	CT	0,985	0,065	0,647	0,920	0,338	0,582	[1]	0,676	0,258	0,003
AIM343	S-4AMD	22	rs470113	AG	0,858	0,065	0,835	0,793	0,023	0,770	[1]	1,000	1,000	0,019
AIM344	S-3YXKR	22	rs9611566	TC	0,034	0,939	0,219	0,905	0,185	0,720	[1]	0,299	0,035	-0,006
AIM345	S-3YYQK	22	rs16990991	AG	0,015	0,874	0,222	0,859	0,207	0,652	[1]	0,002	1,000	0,008

ID = identificação; chr = cromossomo; AFR = africanos; AME = ameríndios; EUR = europeus; δ = diferencial de frequência; Ref = referências; pEHW = valor de p do equilíbrio de Hardy-Weinberg; BRA = controles brasileiros; Fst = estatística F de Wright.

Referências (Apêndice B)

- [1] Galanter, J.M., Fernandez-Lopez, J.C., Gignoux, C.R., Barnholtz-Sloan, J., Fernandez-Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A.V., Figueroa, L.U. & Raska, P. (2012) Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. *PLoS Genet*, 8, e1002554.
- [2] Kosoy, R., Nassir, R., Tian, C., White, P.A., Butler, L.M., Silva, G., Kittles, R., Alarcón-Riquelme, M.E., Gregersen, P.K. & Belmont, J.W. (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human mutation*, 30, 69-78.
- [3] Halder, I., Shriver, M., Thomas, M., Fernandez, J.R. & Frudakis, T. (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Human mutation*, 29, 648-658.
- [4] Lins, T.C., Vieira, R.G., Abreu, B.S., Grattapaglia, D. & Pereira, R.W. (2010) Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. *American Journal of Human Biology*, 22, 187-192.
- [5] Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A. & Cameron, N. (2003) Skin pigmentation, biogeographical ancestry and admixture mapping. *Human genetics*, 112, 387-399.

APÊNDICE C – Genotipagem por Cytoscan HD array

Apêndice C – Qualidade das amostras genotipadas por *Cytoscan HD array* segundo três parâmetros de qualidade (*SNPQC*, *MAPD*, *wavinessSD*).

ID LES	Qualidade do array			CNVs		LOH	
	<i>SNPQC</i> ≥ 12	<i>MAPD</i> ≤ 0,25	<i>wavinessSD</i> ≤ 0,12	<i>Call</i>	Tam (Kb)	<i>Call</i>	Tam (Kb)
LES001	20,11	0,18	0,08	16	1739	7	16456
LES002	16,13	0,18	0,09	16	2216	6	16465
LES003	15,93	0,18	0,08	17	1486	4	12555
LES004	20,12	0,22	0,1	27	1525	3	9734
LES005	16,06	0,17	0,07	15	1425	6	15768
LES006	21,46	0,18	0,09	20	1891	7	18146
LES007	16,45	0,21	0,09	17	2276	6	15875
LES008	20,31	0,19	0,09	16	676	4	10000
LES009	19,71	0,19	0,09	24	2224	2	5723
LES010	15,53	0,21	0,09	13	463	4	9391
LES011	18,95	0,2	0,09	22	1589	6	14638
LES012	15,5	0,19	0,08	18	1758	4	9979
LES013	14,2	0,19	0,09	18	844	9	22797
LES014	14,58	0,19	0,08	11	2310	9	21626
LES015	13,93	0,24	0,1	33	1699	7	19584
LES016	12,89	0,22	0,09	18	1310	18	102023
LES017	12,39	0,22	0,09	14	1576	30	184352
LES018	12,02	0,21	0,08	20	2062	10	26619
LES019	19,93	0,19	0,09	21	1416	10	26366
LES020	15,26	0,18	0,08	11	1043	11	30891
LES021	18,26	0,2	0,09	20	1723	5	10440
LES022	16,02	0,19	0,08	5	652	12	59909
LES023	18,69	0,18	0,1	55	5360	4	11147
BRA001	20,29	0,18	0,09	19	1813	7	19016
BRA002	20,83	0,18	0,09	19	3636	7	19337
BRA003	20,11	0,18	0,09	24	2325	5	12775
BRA004	23,18	0,17	0,08	15	1125	10	25536
BRA005	23,66	0,18	0,09	15	1254	6	14416
BRA006	21,36	0,18	0,08	20	1849	11	27958
BRA007	19,98	0,17	0,08	7	756	12	31156
BRA008	20,95	0,18	0,09	29	2434	9	23091
BRA009	18,14	0,18	0,07	13	1300	7	16252
BRA010	20,34	0,15	0,07	16	1093	5	13488
BRA011	17,71	0,15	0,06	15	1867	6	15387
BRA012	23,31	0,17	0,08	10	1774	3	6866
BRA013	23,79	0,14	0,06	11	1245	5	13224
BRA014	20,62	0,18	0,08	14	2290	11	27710

BRA015	19,14	0,17	0,08	24	1945	6	21903
BRA016	23,47	0,18	0,08	24	1995	6	14609
BRA017	19,11	0,19	0,09	21	1766	6	17481
BRA018	20,02	0,18	0,08	24	4275	9	23260
BRA019	21,83	0,17	0,08	23	2693	7	20045
BRA020	18,8	0,2	0,09	31	1582	1	2738
BRA021	17,62	0,18	0,07	16	1931	8	21127
BRA022	20,14	0,2	0,09	33	3010	5	12268
BRA023	18	0,19	0,08	27	3289	2	4850
BRA024	25,64	0,14	0,08	24	1580	6	15996
BRA025	17,87	0,21	0,09	37	4202	7	17380
BRA026	25,23	0,18	0,08	22	2479	7	17957
BRA027	21,95	0,19	0,1	20	2343	11	27508
BRA028	16,92	0,22	0,09	33	3569	7	18648
BRA029	24,17	0,17	0,09	22	2215	5	12946
BRA030	20,63	0,19	0,08	14	1316	6	16959
BRA031	21,55	0,18	0,09	22	1913	5	14273
BRA032	18,84	0,19	0,08	19	1671	4	11843
BRA033	19,52	0,18	0,1	42	4643	6	15601
BRA034	21,04	0,18	0,08	21	2973	7	16913
BRA035	24,17	0,17	0,09	23	3460	3	9050
BRA036	22,84	0,18	0,09	24	1550	11	31396
BRA037	15,97	0,18	0,07	12	967	7	17298
BRA038	15,12	0,2	0,1	37	3276	10	25254
BRA039	22,11	0,17	0,07	18	3727	5	11342
BRA040	22,3	0,18	0,09	29	2732	7	24752
BRA041	18,08	0,19	0,11	29	2316	5	13560
BRA042	16,42	0,24	0,11	42	3310	10	68052
BRA043	17,96	0,21	0,1	19	2447	7	16985
BRA044	17,12	0,22	0,1	20	2637	9	25877
BRA045	15,74	0,22	0,1	22	1685	18	69828
BRA046	18,38	0,18	0,11	34	4010	6	14933
BRA047	18,93	0,18	0,12	52	4302	6	16862
BRA048	17,36	0,23	0,12	67	4403	5	13301
BRA049	19,6	0,18	0,1	32	2282	7	19266
BRA050	20,59	0,18	0,09	17	1616	10	23910
BRA051	15,77	0,2	0,09	24	2074	5	14068
BRA052	17,03	0,19	0,12	51	10988	6	15542
BRA053	23,51	0,17	0,09	26	2803	8	66215
BRA054	22,21	0,18	0,11	42	3448	6	25479
BRA055	22,11	0,18	0,1	28	6068	6	51643
BRA056	18,41	0,18	0,09	28	4495	12	34897
BRA057	14,13	0,19	0,08	24	3472	5	14156
BRA058	12,1	0,19	0,07	11	1688	7	19172

BRA059	13,59	0,22	0,1	23	2106	2	4375
BRA060	14,34	0,22	0,12	37	3648	1	2617
BRA061	14,26	0,16	0,06	14	1190	8	22575
BRA062	13,91	0,21	0,1	20	1957	12	32078
BRA063	13,53	0,22	0,09	20	1598	7	17351
BRA064	14,62	0,22	0,1	36	3151	11	69512
BRA065	13,37	0,2	0,08	18	1860	5	12926
BRA066	20,12	0,18	0,1	15	1299	7	16096
BRA067	16,72	0,2	0,1	16	2316	8	20890
BRA068	18,43	0,19	0,09	12	1017	6	17948
BRA069	16,74	0,2	0,1	17	2460	12	30538
BRA070	16,79	0,21	0,1	20	1779	8	18444
BRA071	18,57	0,18	0,08	19	2842	5	13632
BRA072	23,75	0,16	0,09	22	3202	5	13630
BRA073	16,7	0,19	0,08	19	1635	5	13250
BRA074	15,66	0,21	0,11	74	8335	3	8496
BRA075	23,12	0,17	0,09	28	1659	0	0
BRA076	19,48	0,21	0,11	25	5851	5	11725
BRA077	20,07	0,17	0,08	16	2121	5	11526
BRA078	20,63	0,18	0,08	18	1349	7	19411
BRA079	16,91	0,18	0,08	13	1098	10	27534
BRA080	24,23	0,17	0,09	30	1077	0	0
BRA081	22,11	0,17	0,08	19	1582	9	20945
BRA082	21,63	0,16	0,08	15	1578	9	25930
BRA083	21,48	0,18	0,1	28	2973	10	25086
BRA084	22,55	0,16	0,09	21	1781	9	23106
BRA085	16,9	0,21	0,12	35	3322	7	17130
BRA086	17,43	0,17	0,09	22	2183	11	27113
BRA087	20,9	0,17	0,11	14	2680	3	8565
BRA088	19,1	0,19	0,1	29	2337	4	10101
BRA089	20,64	0,17	0,09	17	3258	6	15215
BRA090	22,04	0,17	0,09	17	1405	6	14335
BRA091	18,17	0,19	0,08	17	1676	12	31851
BRA092	21,82	0,18	0,11	25	4106	6	16348
BRA093	20,24	0,19	0,1	24	2983	5	14646
BRA094	20,4	0,18	0,09	20	3740	3	6578
BRA095	19,62	0,18	0,1	23	3668	5	13564
BRA096	19,09	0,17	0,08	19	2333	4	10956
BRA097	18,91	0,17	0,1	15	2043	6	16581
BRA098	14,87	0,18	0,07	15	3504	8	21861
BRA099	13,91	0,2	0,08	22	2220	8	19744
BRA100	13,61	0,23	0,12	30	3497	5	13025
BRA101	13,27	0,21	0,12	51	3880	13	32443
BRA102	13,92	0,19	0,09	14	658	4	10597

BRA103	12,61	0,21	0,08	13	708	13	36135
BRA104	12,24	0,18	0,07	12	1127	10	25064
BRA105	17,54	0,2	0,09	47	15242	6	17768
BRA106	15,1	0,21	0,12	29	6514	11	28707
BRA107	19,23	0,17	0,08	32	2008	5	18454
BRA108	21,27	0,17	0,07	21	5813	10	28006
BRA109	16,25	0,17	0,07	28	3125	9	22875
BRA110	24,04	0,16	0,07	29	2792	8	20806

ID = identificação; Call = número total de segmentos Tam = tamanho; BRA = controles saudáveis brasileiros.

APÊNDICE D – Listas de genes com relevância funcional para o LES

Apêndice D – Listas de genes com relevância funcional para o LES: genes que sobrepõem CNVs do tipo deleção ou duplicação descritas em associação com o LES (1), genes previamente associados ao LES em estudos de ligação gênica e/ou GWAS (2); genes relacionados com autoimunidade de forma geral (3).

Lista	Gene	Chr	Posição inicial	Posição final	Tam (Kb)	Referência
1	<i>FCGR3B</i>	1	161592987	161601252	8,3	Yuan et al., 2015
1	<i>RABGAP1L</i>	1	174128551	174927327	798,8	Kim JH et al., 2013
1	<i>IL21</i>	4	123533782	123542212	8,4	Yu et al., 2011
1	<i>C4A</i>	6	31949833	31970457	20,6	Li et al., 2017
1	<i>C4B</i>	6	31949833	31970458	20,6	Li et al., 2017
1	<i>HLA-DRB5</i>	6	32485153	32498006	12,9	Wu L et al., 2014
1	<i>IL17F</i>	6	52101483	52109298	7,8	Yu et al., 2011
1	<i>DEFB4B</i>	8	7272384	7274354	2,0	Zhou et al., 2012
1	<i>IL22</i>	12	68642024	68647281	5,3	Yu et al., 2011
1	<i>CCL3L1</i>	17	34522267	34524156	1,9	Mamtani et al., 2007
1	<i>HRH4</i>	18	22040592	22059921	19,3	Yu et al., 2010
1	<i>TLR7</i>	X	12885201	12908480	23,3	Garcia-Ortiz et al., 2010
2	<i>PTPN22</i>	1	114356432	114414375	57,9	Wang Y et al., 2015
2	<i>CRP</i>	1	159682078	159684379	2,3	Edberg et al., 2008
2	<i>FCGR2A</i>	1	161475204	161489360	14,2	Chu et al., 2004
2	<i>FCGR2B</i>	1	161632904	161648444	15,5	Kyogoku et al., 2002
2	<i>TNFSF4</i>	1	173152869	173176471	23,6	Graham et al., 2008
2	<i>NMNAT2</i>	1	183217371	183387634	170,3	Gateva et al., 2009
2	<i>NCF2</i>	1	183524696	183560056	35,4	Graham et al., 2011
2	<i>CFHR3</i>	1	196743929	196763203	19,3	Zhao et al., 2011
2	<i>CFHR1</i>	1	196788860	196801319	12,5	Zhao et al., 2011
2	<i>CFHR4</i>	1	196857143	196887843	30,7	Zhao et al., 2011
2	<i>CFHR2</i>	1	196912933	196928356	15,4	Zhao et al., 2011
2	<i>CFHR5</i>	1	196946666	196978803	32,1	Zhao et al., 2011
2	<i>IL10</i>	1	206940947	206945839	4,9	Gateva et al., 2009
2	<i>RASGRP3</i>	2	33661415	33789798	128,4	He et al., 2010
2	<i>IL1RN</i>	2	113875469	113891593	16,1	Huang et al., 2002
2	<i>CXCR4</i>	2	136871918	136873813	1,9	Wang A et al., 2010
2	<i>IFIH1</i>	2	163123588	163175039	51,5	Graham et al., 2011
2	<i>STAT1</i>	2	191833761	191878976	45,2	Kawasaki et al., 2008
2	<i>STAT4</i>	2	191894301	192016322	122,0	Yang et al., 2009
2	<i>CD28</i>	2	204571197	204603636	32,4	Wong et al., 2005
2	<i>PDCD1</i>	2	242792032	242801058	9,0	Thorburn et al., 2007
2	<i>TREX1</i>	3	48507228	48509044	1,8	Namjou et al., 2011
2	<i>PXK</i>	3	58318616	58410878	92,3	Harley JB et al., 2008
2	<i>BANK1</i>	4	102711763	102995969	284,2	Kozyrev et al., 2008
2	<i>IL2</i>	4	123372625	123377650	5,0	Solomou et al., 2001
2	<i>IL4</i>	5	132009677	132018370	8,7	Wu M et al., 2003

2	<i>TNIP1</i>	5	150409503	150444692	35,2	Gateva et al., 2009
2	<i>MIR146A</i>	5	159912358	159912457	0,1	Löfgren et al., 2012
2	<i>C2</i>	6	31868775	31913449	44,7	Bryan e Wu, 2014
2	<i>HLA-DRB1</i>	6	32546546	32557613	11,1	Fernando et al., 2007
2	<i>HLA-DQA1</i>	6	32605182	32611429	6,2	Marchini et al., 2003
2	<i>UHRF1BP1</i>	6	34759793	34845291	85,5	Gateva et al., 2009
2	<i>PRDM1</i>	6	106534194	106557814	23,6	Gateva et al., 2009
2	<i>ATG5</i>	6	106632351	106773695	141,3	Zhang et al., 2015
2	<i>TNFAIP3</i>	6	138188580	138204449	15,9	Kim S-K et al., 2014
2	<i>ICA1</i>	7	8153403	8301780	148,4	Vaughn et al., 2012
2	<i>IL6</i>	7	22766765	22771621	4,9	Solus et al., 2015
2	<i>IKZF1</i>	7	50344377	50472798	128,4	Graham et al., 2011
2	<i>IRF5</i>	7	128580771	128590088	9,3	Zervou et al., 2017
2	<i>XKR6</i>	8	10753656	11058875	305,2	Guerra et al., 2012
2	<i>FAM167A</i>	8	11278972	11324276	45,3	Harley ITW et al., 2009
2	<i>BLK</i>	8	11351520	11422108	70,6	Guerra et al., 2012
2	<i>IKBKB</i>	8	42128819	42190171	61,4	Vaughn et al., 2012
2	<i>LYN</i>	8	56792385	56925006	132,6	Vaughn et al., 2012
2	<i>WDFY4</i>	10	49893517	50191001	297,5	Harley ITW et al., 2009
2	<i>LRRC18</i>	10	50117528	50122280	4,8	Guerra et al., 2012
2	<i>PHRF1</i>	11	576482	612222	35,7	Harley ITW et al., 2009
2	<i>IRF7</i>	11	612554	615999	3,4	Harley ITW et al., 2009
2	<i>CD44</i>	11	35160416	35244867	84,5	Guerra et al., 2012
2	<i>ETS1</i>	11	128328655	128457453	128,8	He et al., 2010
2	<i>SLC15A4</i>	12	129277738	129308541	30,8	He et al., 2010
2	<i>ELF1</i>	13	41506054	41556418	50,4	Guerra et al., 2012
2	<i>PRKCB</i>	16	23847299	24231932	384,6	Vaughn et al., 2012
2	<i>ITGAM</i>	16	31271287	31344213	72,9	Harley JB et al., 2008
2	<i>IRF8</i>	16	85932773	85956211	23,4	Graham et al., 2011
2	<i>TYK2</i>	19	10461203	10491248	30,0	Graham et al., 2011
2	<i>CD40</i>	20	44746905	44758384	11,5	Vaughn et al., 2012
2	<i>SCUBE1</i>	22	43599228	43739394	140,1	Harley ITW et al., 2009
2	<i>UBE2L3</i>	22	21922018	21978323	56,3	Vaughn et al., 2012
2	<i>IRAK1</i>	X	153275956	153285342	9,4	Kaufman et al., 2012
2	<i>MECP2</i>	X	153287263	153363188	75,9	Kaufman et al., 2012
3	<i>MMEL1</i>	1	2522080	2564481	42,4	Smyk et al., 2012
3	<i>PADI4</i>	1	17634689	17690495	55,8	Gregersen e Behrens, 2006
3	<i>IL23R</i>	1	67632168	67725650	93,5	Gregersen e Behrens, 2006
3	<i>FCGR1C</i>	1	149369293	149378297	9,0	Bronstad et al., 2011
3	<i>ITLN1</i>	1	160846329	160854960	8,6	Lettre e Rioux, 2008
3	<i>RGS1</i>	1	192544856	192549159	4,3	Lettre e Rioux, 2008
3	<i>REL</i>	2	61108751	61150178	41,4	Gregersen e Behrens, 2006
3	<i>IL18RAP</i>	2	103035253	103069024	33,8	Lettre e Rioux, 2008
3	<i>LRP1B</i>	2	140988995	142889270	1900,3	Bronstad et al., 2011

3	<i>CTLA4</i>	2	204732510	204738683	6,2	Gregersen e Behrens, 2006
3	<i>ATG16L1</i>	2	234160216	234204320	44,1	Lettre e Rioux, 2008
3	<i>TBC1D5</i>	3	17198653	17784240	585,6	Bronstad et al., 2011
3	<i>CCR3</i>	3	46283871	46308197	24,3	Lettre e Rioux, 2008
3	<i>MST1</i>	3	49721379	49726196	4,8	Lettre e Rioux, 2008
3	<i>KIAA1257</i>	3	128689781	128712986	23,2	Bronstad et al., 2011
3	<i>RHO</i>	3	129247481	129254187	6,7	Bronstad et al., 2011
3	<i>H1FOO</i>	3	129262056	129270204	8,1	Bronstad et al., 2011
3	<i>PLXND1</i>	3	129274055	129325582	51,5	Bronstad et al., 2011
3	<i>IL12A</i>	3	159706622	159713806	7,2	Lettre e Rioux, 2008
3	<i>LPP</i>	3	187871662	188608460	736,8	Lettre e Rioux, 2008
3	<i>UGT2B28</i>	4	70146216	70160768	14,6	Bronstad et al., 2011
3	<i>NFKB1</i>	4	103422485	103538459	116,0	Smyk et al., 2012
3	<i>PTGER4</i>	5	40680031	40693837	13,8	Lettre e Rioux, 2008
3	<i>HSD17B4</i>	5	118788201	118878030	89,8	Bronstad et al., 2011
3	<i>IL12B</i>	5	158741790	158757481	15,7	Lettre e Rioux, 2008
3	<i>CDKAL1</i>	6	20534687	21232634	697,9	Lettre e Rioux, 2008
3	<i>HLA-F</i>	6	29691116	29694303	3,2	Singer et al., 1996
3	<i>HLA-F-AS1</i>	6	29694377	29716826	22,4	Singer et al., 1996
3	<i>HLA-G</i>	6	29794755	29798899	4,1	Singer et al., 1996
3	<i>HLA-H</i>	6	29855382	29858856	3,5	Singer et al., 1996
3	<i>HLA-A</i>	6	29910246	29913661	3,4	Singer et al., 1996
3	<i>HLA-J</i>	6	29973747	29977733	4,0	Singer et al., 1996
3	<i>HLA-L</i>	6	30227338	30234728	7,4	Singer et al., 1996
3	<i>HLA-E</i>	6	30457182	30461982	4,8	Singer et al., 1996
3	<i>HLA-C</i>	6	31236525	31239913	3,4	Singer et al., 1996
3	<i>HLA-B</i>	6	31321648	31324989	3,3	Singer et al., 1996
3	<i>HLA-DRA</i>	6	32407618	32412826	5,2	Todd et al., 1988
3	<i>HLA-DRB6</i>	6	32520489	32527779	7,3	Todd et al., 1988
3	<i>HLA-DQB1</i>	6	32627240	32634466	7,2	Todd et al., 1988
3	<i>HLA-DQA2</i>	6	32709162	32714664	5,5	Todd et al., 1988
3	<i>HLA-DQB2</i>	6	32723874	32731330	7,5	Todd et al., 1988
3	<i>HLA-DMB</i>	6	32902405	32908847	6,4	Todd et al., 1988
3	<i>HLA-DMA</i>	6	32916390	32920899	4,5	Todd et al., 1988
3	<i>HLA-DOA</i>	6	32971959	32977389	5,4	Todd et al., 1988
3	<i>HLA-DPA1</i>	6	33032345	33048555	16,2	Todd et al., 1988
3	<i>HLA-DPB1</i>	6	33043702	33057473	13,8	Todd et al., 1988
3	<i>HLA-DPB2</i>	6	33080292	33096890	16,6	Todd et al., 1988
3	<i>CCR6</i>	6	167525294	167552629	27,3	Lettre e Rioux, 2008
3	<i>BMPER</i>	7	33944522	34195484	251,0	Bronstad et al., 2011
3	<i>TYW1</i>	7	66461816	66704498	242,7	Lettre e Rioux, 2008
3	<i>NCF1</i>	7	74188308	74203659	15,4	Olsson e Holmdahl, 2012
3	<i>SCARA5</i>	8	27727398	27850369	123,0	Bronstad et al., 2011
3	<i>ADAM3A</i>	8	39308563	39380470	71,9	Bronstad et al., 2011

3	<i>JAK2</i>	9	4985244	5128183	142,9	Lettre e Rioux, 2008
3	<i>TRAF1</i>	9	123664670	123676850	12,2	Lettre e Rioux, 2008
3	<i>IL2RA</i>	10	6052656	6104333	51,7	Lettre e Rioux, 2008
3	<i>ZNF365</i>	10	64133915	64162217	28,3	Lettre e Rioux, 2008
3	<i>NKX2-3</i>	10	101292689	101296280	3,6	Lettre e Rioux, 2008
3	<i>INS</i>	11	2181008	2182439	1,4	Gregersen e Behrens, 2006
3	<i>CXCR5</i>	11	118754474	118766980	12,5	Smyk et al., 2012
3	<i>LRRK2</i>	12	40618812	40763086	144,3	Lettre e Rioux, 2008
3	<i>ERBB3</i>	12	56473808	56497291	23,5	Lettre e Rioux, 2008
3	<i>RASSF9</i>	12	86198330	86230318	32,0	Bronstad et al., 2011
3	<i>ANKS1B</i>	12	99128568	99288674	160,1	Grayson et al., 2010
3	<i>SH2B3</i>	12	111843751	111889427	45,7	Lettre e Rioux, 2008
3	<i>NFKBIA</i>	14	35870715	35873960	3,2	Smyk et al., 2012
3	<i>TSHR</i>	14	81421868	81575294	153,4	Gregersen e Behrens, 2006
3	<i>GOLGA6L6</i>	15	20737093	20747114	10,0	Grayson et al., 2010
3	<i>GOLGA8C</i>	15	20767673	20781026	13,4	Grayson et al., 2010
3	<i>POTEB</i>	15	21040700	21071977	31,3	Grayson et al., 2010
3	<i>CLEC16A</i>	16	11038344	11220201	181,9	Lettre e Rioux, 2008
3	<i>NOD2</i>	16	50731049	50766987	35,9	Gregersen e Behrens, 2006
3	<i>TRIM16</i>	17	15531279	15586193	54,9	Grayson et al., 2010
3	<i>ORMDL3</i>	17	38077295	38083884	6,6	Lettre e Rioux, 2008
3	<i>STAT3</i>	17	40465342	40540405	75,1	Lettre e Rioux, 2008
3	<i>PTPN2</i>	18	12785476	12884334	98,9	Lettre e Rioux, 2008
3	<i>SPIB</i>	19	50922194	50934309	12,1	Smyk et al., 2012
3	<i>ICOSLG</i>	21	45646721	45660834	14,1	Lettre e Rioux, 2008
3	<i>AIRE</i>	21	45705720	45718102	12,4	Liston et al., 2004
3	<i>FOXP3</i>	X	49106896	49121288	14,4	Gregersen e Behrens, 2006

chr = cromossomo; *Tam* = Tamanho.

Referências (Apêndice D)

Bronstad I, Wolff AS, Lovas K, Knappskog PM, Husebye ES. Genome-wide copy number variation (CNV) in patients with autoimmune Addison's disease. *BMC Medical Genetics*. 2011;12:111.

Bryan AR, Wu EY. Complement Deficiencies in Systemic Lupus Erythematosus. *Current Allergy and Asthma Reports*. 2014;14(7):1-7.

Chu Z, Tsuchiya N, Kyogoku C, Ohashi J, Qian Y, Xu S, et al. Association of Fcγ receptor IIb polymorphism with susceptibility to systemic lupus erythematosus in Chinese: a common susceptibility gene in the Asian populations. *Tissue Antigens*. 2004;63(1):21-7.

Edberg JC, Wu J, Langefeld CD, Brown EE, Marion MC, McGwin G, et al. Genetic variation in the CRP promoter: association with systemic lupus erythematosus. *Human Molecular Genetics*. 2008;17(8):1147-55.

Fernando MM, Stevens CR, Sabeti PC, Walsh EC, McWhinnie AJ, Shah A, et al. Identification of two independent risk factors for lupus within the MHC in United Kingdom families. *PLoS Genetics*. 2007;3(11):e192.

Garcia-Ortiz H, Velazquez-Cruz R, Espinosa-Rosales F, Jimenez-Morales S, Baca V, Orozco L. Association of TLR7 copy number variation with susceptibility to childhood-onset systemic lupus erythematosus in Mexican population. *Annals of the Rheumatic Diseases*. 2010 Oct;69(10):1861-5.

Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nature Genetics*. 2009;41(11):1228-33.

Graham DSC, Graham RR, Manku H, Wong AK, Whittaker JC, Gaffney PM, et al. Polymorphism at the TNF superfamily gene TNFSF4 confers susceptibility to systemic lupus erythematosus. *Nature Genetics*. 2008;40(1):83-9.

Graham DSC, Morris DL, Bhangale TR, Criswell LA, Syvänen A-C, Rönnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genetics*. 2011;7(10):e1002341.

Grayson BL, Smith ME, Thomas JW, Wang L, Dexheimer P, Jeffrey J, et al. Genome-wide analysis of copy number variation in type 1 diabetes. *PloS One*. 2010;5(11):e15393.

Gregersen PK, Behrens TW. Genetics of autoimmune diseases—disorders of immune homeostasis. *Nature Reviews Genetics*. 2006;7(12):917-28.

Guerra SG, Vyse TJ, Graham DSC. The genetics of lupus: a functional perspective. *Arthritis Research & Therapy*. 2012;(14)2:3.

Harley ITW, Kaufman KM, Langefeld CD, Harley JB, Kelly JA. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nature Reviews Genetics*. 2009;10(5):285-90.

Harley JB, Alarcón-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nature Genetics*. 2008;40(2):204-10.

He C, Liu Y, Cheng Y, Gao J, Pan T, Han J, et al. TNIP1, SLC15A4, ETS1, RasGRP3 and IKZF1 are associated with clinical features of systemic lupus erythematosus in a Chinese Han population. *Lupus*. 2010;19(10):1181-6.

Huang C-M, Wu M-C, Wu J-Y, Tsai F-J. Interleukin-1 receptor antagonist gene polymorphism in Chinese patients with systemic lupus erythematosus. *Clinical Rheumatology*. 2002;21(3):255-7.

Kaufman KM, Zhao J, Kelly JA, Hughes T, Adler A, Sanchez E, et al. Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. *Annals of the Rheumatic Diseases*. 2012.

Kawasaki A, Ito I, Hikami K, Ohashi J, Hayashi T, Goto D, et al. Role of STAT4 polymorphisms in systemic lupus erythematosus in a Japanese population: a

case-control association study of the STAT1-STAT4 region. *Arthritis Research and Therapy*. 2008;10(5):113.

Kim JH, Jung SH, Bae JS, Lee HS, Yim SH, Park SY, et al. Deletion variants of RABGAP1L, 10q21. 3, and C4 are associated with the risk of systemic lupus erythematosus in Korean women. *Arthritis & Rheumatism*. 2013;65(4):1055-63.

Kim S-K, Choe J-Y, Bae J, Chae S-C, Park D-J, Kwak SG, et al. TNFAIP3 gene polymorphisms associated with differential susceptibility to rheumatoid arthritis and systemic lupus erythematosus in the Korean population. *Rheumatology*. 2014;ket473.

Kozyrev SV, Abelson A-K, Wojcik J, Zaghlool A, Reddy MPL, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. *Nature Genetics*. 2008;40(2):211-6.

Kyogoku C, Dijstelbloem HM, Tsuchiya N, Hatta Y, Kato H, Yamaguchi A, et al. Fcγ receptor gene polymorphisms in Japanese patients with systemic lupus erythematosus: contribution of FCGR2B to genetic susceptibility. *Arthritis & Rheumatism*. 2002;46(5):1242-54.

Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Human molecular genetics*. 2008;17(R2):R116-R21.

Li N, Zhang J, Liao D, Yang L, Wang Y, Hou S. Association between C4, C4A, and C4B copy number variations and susceptibility to autoimmune diseases: a meta-analysis. *Sci Rep*. 2017 Feb 16;7:42628.

Liston A, Gray DH, Lesage S, Fletcher AL, Wilson J, Webster KE, et al. Gene dosage-limiting role of Aire in thymic expression, clonal deletion, and organ-specific autoimmunity. *Journal of Experimental Medicine*. 2004;200(8):1015-26.

Löfgren SE, Frostegård J, Truedsson L, Pons-Estel BA, D'Alfonso S, Witte T, et al. Genetic association of miRNA-146a with systemic lupus erythematosus in Europeans through decreased expression of the gene. *Genes and Immunity*. 2012;13(3):268-74.

Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, et al. CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythematosus. *Annals of the Rheumatic Diseases*. 2007.

Marchini M, Antonioli R, Lleò A, Barili M, Caronni M, Origgi L, et al. HLA class II antigens associated with lupus nephritis in Italian SLE patients. *Human Immunology*. 2003;64(4):462-8.

Namjou B, Kothari PH, Kelly JA, Glenn SB, Ojwang JO, Adler A, et al. Evaluation of the TREX1 gene in a large multi-ancestral lupus cohort. *Genes and immunity*. 2011;12(4):270-9.

Olsson LM, Holmdahl R. Copy number variation in autoimmunity—importance hidden in complexity? *European Journal of Immunology*. 2012;42(8):1969-76.

Singer DS, Mozes E, Kirshner S, Kohn LD. Role of MHC class I molecules in autoimmune disease. *Critical Reviews in Immunology*. 1996;17(5-6):463-8.

Smyk DS, Bogdanos DP, Mytilinaïou MG, Burroughs AK, Rigopoulou EI. Rheumatoid arthritis and primary biliary cirrhosis: cause, consequence, or coincidence? *Arthritis*. 2012;2012.

Solomou EE, Juang Y-T, Gourley MF, Kammer GM, Tsokos GC. Molecular basis of deficient IL-2 production in T cells from patients with systemic lupus erythematosus. *The Journal of Immunology*. 2001;166(6):4216-22.

Solus JF, Chung CP, Oeser A, Li C, Rho YH, Bradley KM, et al. Genetics of serum concentration of IL-6 and TNF α in systemic lupus erythematosus and rheumatoid arthritis: a candidate gene analysis. *Clinical Rheumatology*. 2015;34(8):1375-82.

Thorburn C, Prokunina-Olsson L, Sterba K, Lum R, Seldin MF, Alarcon-Riquelme ME, et al. Association of PDCD1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. *Genes and Immunity*. 2007;8(4):279-87.

Todd JA, Acha-Orbea H, Bell JI, Chao N, Fronek Z, Jacob CO, et al. A molecular basis for MHC class II-associated autoimmunity. *Science*. 1988;240(4855):1003-10.

Vaughn SE, Kottyan LC, Munroe ME, Harley JB. Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. *Journal of Leukocyte Biology*. 2012;92(3):577-91.

Wang A, Guilpain P, Chong BF, Chouzenoux S, Guillevin L, Du Y, et al. Dysregulated expression of CXCR4/CXCL12 in subsets of patients with systemic lupus erythematosus. *Arthritis & Rheumatology*. 2010;62(11):3436-46.

Wang Y, Ewart D, Crabtree JN, Yamamoto A, Baechler EC, Fazeli P, et al. PTPN22 Variant R620W Is Associated With Reduced Toll-like Receptor 7–Induced Type I Interferon in Systemic Lupus Erythematosus. *Arthritis & Rheumatology*. 2015;67(9):2403-14.

Wong C, Lit L, Tam L, Li E, Lam C. Aberrant production of soluble costimulatory molecules CTLA-4, CD28, CD80 and CD86 in patients with systemic lupus erythematosus. *Rheumatology*. 2005;44(8):989-94.

Wu L, Guo S, Yang D, Ma Y, Ji H, Chen Y, et al. Copy number variations of HLA-DRB5 is associated with systemic lupus erythematosus risk in Chinese Han population. *Acta Biochimica et Biophysica Sinica*. 2014;46(2):155-60.

Wu M, Huang C, Tsai JJ, Chen H, Tsai F. Polymorphisms of the interleukin-4 gene in Chinese patients with systemic lupus erythematosus in Taiwan. *Lupus*. 2003;12(1):21-5.

Yang W, Ng P, Zhao M, Hirankarn N, Lau C, Mok C, et al. Population differences in SLE susceptibility genes: STAT4 and BLK, but not PXX, are associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes and Immunity*. 2009;10(3):219-26.

Yu B, Guan M, Peng Y, Shao Y, Zhang C, Yue X, et al. Copy number variations of interleukina-17F, interleukina-21, and interleukina-22 are associated with systemic lupus erythematosus. *Arthritis & Rheumatism*. 2011;63(11):3487-92.

Yu B, Shao Y, Li P, Zhang J, Zhong Q, Yang H, et al. Copy number variations of the human histamine H4 receptor gene are associated with systemic lupus erythematosus. *Br J Dermatol*. 2010 Nov;163(5):935-40.

Yuan J, Zhao D, Wu L, Xu X, Pang Y, Zhang J, et al. FCGR3B copy number loss rather than gain is a risk factor for systemic lupus erythematosus and lupus nephritis: a meta-analysis. *International Journal of Rheumatic Diseases*. 2015.

Zervou M, Dorschner J, Ghodke-Puranik Y, Boumpas D, Niewold T, Goulielmos G. Association of IRF5 polymorphisms with increased risk for systemic lupus erythematosus in population of Crete, a southern-eastern European Greek island. *Gene*. 2017;610:9-14.

Zhang Y-m, Cheng F-j, Zhou X-j, Qi Y-y, Zhao M-h, Zhang H. Rare variants of ATG5 are likely to be associated with chinese patients with systemic lupus erythematosus. *Medicine*. 2015;94(22):e939.

Zhao J, Wu H, Khosravi M, Cui H, Qian X, Kelly JA, et al. Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility. *PLoS Genetics*. 2011;7(5):e1002079.

Zhou XJ, Cheng F-J, Lv JC, Luo H, Yu F, Chen M, et al. Higher DEFB4 genomic copy number in SLE and ANCA-associated small vasculitis. *Rheumatology*. 2012;51(6):992-

11 ANEXOS

ANEXO A – Protocolo Cytoscan HD array

As CNVs foram detectadas por meio de *Cytoscan HD array* (Affymetrix Inc., Santa Clara, CA, USA), seguindo as recomendações do fabricante. O *Cytoscan HD array* contém sondas de oligonucleotídeos com poder para detecção de alterações estruturais, com ampla cobertura do genoma, contendo 1,9 milhões de não polimórficas e 750 mil SNPs.

As 133 amostras foram processadas de acordo com o protocolo padrão do *Cytoscan HD array*. O ensaio consistiu nas etapas descritas a seguir.

A – 1 Digestão do DNA

Reagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Água livre de nucleases

Tampão TE Low EDTA

BSA 100X: concentração inicial 10 µg/µL.

Enzima *NspI*

Tampão da *NspI* 10X: Tris-HCl 10 mM; NaCl 50 mM; MgCl₂ 10 mM; DTT 1 mM.

Procedimento

Inicialmente, 5 µL do DNA a 50 ng/µL foram adicionados aos poços correspondentes de uma placa de 96 poços. Como controle positivo foi utilizado 5 µL do DNA fornecido no *kit* e como controle negativo, 5 µL de água livre de nucleases. Em seguida, a mistura da digestão foi preparada em um tubo de 1,5 mL contendo todos os reagentes: 11,55 µL de água livre de nucleases; 0,2 µL de BSA 100X; 2,0 µL de tampão da *NspI* 10X; 1,0 µL enzima *NspI* (quantidades para uma reação).

A mistura foi colocada no *Vortex* três vezes em alta velocidade, e distribuída igualmente em tubos conectados (strip). Com o auxílio de uma pipeta multicanal 14,75 µL foram adicionados aos poços de uma placa de 96 poços. A placa foi selada, colocada no *Vortex* uma vez em alta velocidade e centrifugada a 2000 rpm durante um minuto. A este passo seguiu-se o programa do termociclador listado em seguida.

Programa CytoScan Digest

37°C por 2 horas;
65°C por 20 minutos;
4°C indefinidamente.

A – 2 Ligação aos adaptadoresReagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Adaptador de *NspI*: concentração 50 µM.

Enzima *T4 DNA Ligase*

Tampão da *T4 DNA Ligase 10X*: Tris-HCl (pH 7,6) 660 mM; MgCl₂ 66 mM; DTT 100 mM; ATP 660 µM.

Procedimento

Os fragmentos de DNA anteriormente digeridos pela enzima *NspI* apresentam, independente do tamanho, extremidades livres de 4 pb de comprimento. Nesta etapa, adaptadores específicos são ligados a essas extremidades. Para que a reação ocorresse, a mistura da ligação foi preparada em um tubo de 1,5 mL contendo todos os reagentes: 0,75 µL de adaptador *NspI*; 2,5 µL tampão da *T4 DNA Ligase 10X*; 2,0 µL enzima *T4 DNA Ligase* (quantidades para uma reação).

Em seguida, a mistura foi colocada no *Vortex* três vezes em alta velocidade, e distribuída igualmente em tubos conectados (strip). Com o auxílio de um pipeta multicanal 5,25 µL foram adicionados aos poços de uma placa de 96 poços. A placa foi selada, colocada no *Vortex* uma vez em alta velocidade, e centrifugada a 2000 rpm durante 1 minuto. A este passo seguiu-se o programa do termociclador listado abaixo.

Programa CytoScan Ligate

16°C por 3 horas;
70°C por 20 minutos;
4°C indefinidamente.

A – 3 Amplificação dos fragmentos por PCRReagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Água livre de nucleases

Reagente GC-Melt

dNTP: concentração 2,5 mM de cada ácido nucléico.

PCR Primer 002

Clontech Takara Bio, Mountain View, CA, USA:

Enzima *Taq* DNA Polimerase TITANIUM™ 50X: Glicerol 1% 20 mM; Tris-HCl (pH 8,0) 0,4 mM; KCl 2 mM; EDTA (pH 8,0) 2,0 µM; Tween-20: 0,005%; Nonidet P-40: 0,005%.

Tampão da *Taq* TITANIUM™ 10X: Tricina-KOH (pH 8,0) 40 mM; KCl 16 mM; MgCl₂ 3,5 mM; BSA 3,75 µg/ml.

Procedimento

Inicialmente, a placa contendo as amostras foi centrifugada a 2000 rpm durante um minuto. As amostras foram diluídas em 75 µL de água livre de nuclease. A placa foi devidamente selada e colocada no *Vortex* duas vezes e centrifugada novamente nas condições similares às citadas anteriormente. Quatro alíquotas de 10 µL de cada amostra foram transferidas para outra placa de PCR.

Em seguida, a mistura da ligação foi preparada em um tubo de 50 mL contendo todos os reagentes: 39,5 µL de água livre de nucleases; 20,0 µL de reagente GC-Melt; 14,0 µL de dNTP; 4,5 µL de *primer* (PCR Primer 002); 10,0 µL tampão da *Taq* TITANIUM™ 50X; 2,0 µL enzima *Taq* DNA Polimerase TITANIUM™ (quantidades para uma reação).

A mistura foi colocada no *Vortex* três vezes em alta velocidade, e distribuída em um reservatório. Com o auxílio de um pipeta multicanal 90 µL foram adicionado aos poços das amostras. A placa foi selada, colocada no *Vortex* duas vezes em alta velocidade, e centrifugada a 2000 rpm durante um minuto. A este passo seguiu-se o programa do termociclador listado abaixo. As condições da PCR foram otimizadas pelo fabricante para amplificar preferencialmente fragmentos entre 150 e 2000 pb.

Programa CytoScan PCR

94°C por 3 minutos;	} 30X
94°C por 30 segundos	
60°C por 45 segundos	
68°C por 15 segundos	
68°C por 7 minutos;	
4°C indefinidamente.	

1º Ponto de Checagem

Para checar se as etapas de digestão, ligação e amplificação do material ocorreram de forma correta, foi feita eletroforese em gel de agarose 2% do

produto da PCR. Foi utilizado tampão de amostra (*Loading Dye*) *USB RapidRun™ 5X* e marcador (*ladder*) *USB 50-2000 pb*. Como as condições da PCR foram otimizadas para amplificar fragmentos entre 150 e 2000 pb, o padrão de bandas visualizado no gel deveria corresponder a esse intervalo, como mostrado na Figura A1.

Figura A – 1. Padrão de bandas do gel de agarose 2% dos produtos amplificados da PCR. A maioria do material deve estar entre 150 e 2000 pb. Fonte: Manual do protocolo *Cytoscan HD array* (Affymetrix®).

A– 4 Purificação dos produtos da PCR

Reagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Purification Beads

Tampão de lavagem (Purification Wash Buffer)

Tampão de eluição (Elution Buffer)

Procedimento

Os quatro produtos da PCR de cada amostra foram transferidos para tubos (tipo Eppendorf Safe-Lock) de 1,5 mL correspondentes. Foram adicionados 720 µL do reagente *Purification Beads* em cada amostra. Cada tubo foi tampado, misturado por inversão e incubado a temperatura ambiente durante dez minutos.

Os tubos foram centrifugados a 16100 rcf durante três minutos e transferidos para um suporte magnético (MagnaRack™ Life Technologies). Deixando os

tubos no suporte, o sobrenadante foi retirado e descartado, com auxílio de uma pipeta, sem que tocasse o precipitado.

Em seguida, 1,0 mL do tampão de lavagem foi adicionado a cada amostra. Os tubos foram tampados, e levados ao *Vortex* na velocidade máxima durante dois minutos. Centrifugou-se a 16100 rcf durante três minutos com as dobradiças dos tubos voltadas para fora na centrífuga. Os tubos foram levados de volta ao suporte magnético. Novamente o sobrenadante foi retirado e descartado, com auxílio de uma pipeta, sem que tocasse o precipitado. Centrifugou-se a 16100 rcf durante 30 segundos e quaisquer gotas do tampão de lavagem restantes foi removida do fundo do tubo com o auxílio de uma pipeta. Os tubos foram tirados do suporte magnético e deixados destampados durante dez minutos em temperatura ambiente, para assegurar que qualquer resíduo restante evaporasse.

Adicionou-se 52 μ L do tampão de eluição nas amostras. Os tubos foram tampados, colocados dentro de um adaptador de espuma e levados ao *Vortex* na velocidade máxima durante dez minutos para ressuspender o material aderido às *Beads* magnéticas. Centrifugou-se a 16100 rcf durante três minutos com as dobradiças dos tubos voltadas para fora na centrífuga. Os tubos foram colocados no suporte magnético por dez minutos, até que todas as *Beads* magnéticas aderissem a parede do tubo.

O volume de 47 μ L da amostra eluída foi transferido para o poço correspondente de uma placa de 96 poços. A placa foi devidamente selada, levada ao *Vortex* uma vez em alta velocidade, e centrifugada a 2000 rpm durante um minuto.

2º Ponto de Checagem:

Para certificar que a purificação dos produtos da PCR garantiu condições ideais para o prosseguimento dos experimentos foi feita quantificação do DNA. Na avaliação das leituras de DO, os padrões atingidos devem ser: (1) concentração: ≥ 2500 ng/ μ L, (2) razão A260/A280: 1,8 e 2,0, (3) medição A320: $< 0,1$.

A – 5 Fragmentação

Reagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Água livre de nuclease

Reagente de fragmentação

Tampão de fragmentação (Fragmentation Buffer)

Procedimento

Inicialmente, a placa contendo o produto purificado foi levada ao *Vortex* uma vez em alta velocidade e colocada sobre o bloco de refrigeração em gelo, onde permaneceu durante essa etapa.

Em seguida, a mistura da fragmentação foi preparada em um tubo de 1,5 mL contendo todos os reagentes: 122,4 µL de água livre de nuclease; 7,2 µL de reagente de fragmentação; 158,4 µL de tampão de fragmentação (quantidades para oito reações).

A mistura foi colocado no *Vortex* três vezes em alta velocidade, e distribuído igualmente em tubos conectados (strip). Com o auxílio de um pipeta multicanal, adicionou-se 10,0 µL aos poços de um placa de 96 poços. A placa foi selada, colocada no *Vortex* uma vez em alta velocidade e depois em uma centrífuga pré-refrigerada (4°C) a 2000 rpm durante um minuto. A este passo seguiu-se o programa do termociclador listado abaixo.

Programa CytoScan Fragment

37°C por 35 minutos;

95°C por 15 minutos;

4°C indefinidamente.

3º Ponto de Checagem

Para checar se a fragmentação ocorreu de forma esperada, foi feita eletroforese em gel de agarose 4% do produto fragmentado. Foi utilizado tampão de amostra (*Lodding Dye*) *USB RapidRun™ 5X* e marcador (*ladder*) *TrackIt™ 25 pb*. O padrão de bandas visualizado no gel deveria corresponder ao intervalo de 25 a 125 pb para confirmar que houve a fragmentação, como mostrado na Figura A2.

Figura A – 2. Padrão de bandas do gel de agarose 4% dos produtos fragmentados. A maioria do material deve estar entre 25 e 125 pb. Fonte: Manual do protocolo *Cytoscan HD array* (Affymetrix®).

A – 6 Pré-marcação

Reagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Reagente de marcação de DNA: concentração 30 mM.

Enzima *TdT*

Tampão da *TdT* 5X: Ácido cacodílico (pH 6,8) 500mM; CoCl₂ 5mM; DTT 0,5mM.

Procedimento

A placa contendo o produto purificado foi levada ao *Vortex* uma vez em alta velocidade, e colocada sobre o bloco de refrigeração em gelo, onde permaneceu durante essa etapa.

Em seguida, a mistura da marcação foi preparada em um tubo de 1,5 mL contendo todos os reagentes: 2,0 µL de reagente de marcação do DNA; 14,0 µL de tampão da *TdT* 5X; 3,5 µL de enzima *TdT* (quantidades para uma reação).

A mistura foi colocada no *Vortex* três vezes em alta velocidade, e distribuída igualmente em tubos conectados (strip). Com o auxílio de um pipeta multicanal 19,5 µL foram adicionados aos poços de uma placa de 96 poços. A placa foi selada, colocada no *Vortex* uma vez em alta velocidade e depois em uma centrífuga pré-refrigerada (4°C) a 2000 rpm durante um minuto. A este passo seguiu-se o programa do termociclador listado abaixo.

Programa CytoScan Label

37°C por 4 horas;
 95°C por 15 minutos;
 4°C indefinidamente.

A – 7 Hibridação nos chips Cytoscan HD arrayReagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Reagente Oligo Control 0100**Tampão de hibridação 1 (Hyb Buffer Part 1)****Tampão de hibridação 2 (Hyb Buffer Part 2)****Tampão de hibridação 3 (Hyb Buffer Part 3)****Tampão de hibridação 4 (Hyb Buffer Part 4)**Procedimento

Nesta fase, cada amostra foi hibridada em um *chip Cytoscan HD™* individual. Inicialmente, a mistura da hibridação foi preparada em um tubo de 15 mL contendo todos os reagentes: 165,0 µL de tampão de hibridação 1; 15,0 µL tampão de hibridação 2; 7,0 µL de tampão de hibridação 3; 1,0 µL de tampão de hibridação 4; 2,0 µL reagente Oligo Control 0100 (quantidades para uma reação).

A mistura foi colocado no *Vortex* três vezes em alta velocidade, e distribuído em um reservatório. Com o auxílio de um pipeta multicanal 190 µL foram adicionados aos poços das amostras. A placa foi selada, colocada no *Vortex* duas vezes em alta velocidade, e centrifugada a 2000 rpm durante um minuto. A este passo seguiu-se o programa do termociclador listado abaixo.

Programa CytoScan Hyb

95°C por 10 minutos;
 49°C indefinidamente.

Em seguida, deixando as amostras no termociclador, adicionou-se 200 µL de amostra em cada *chip* individual. Antes dos *chips* serem vedados firmemente com adesivo, qualquer excesso de líquido em torno dos septos foi removido cuidadosamente. Imediatamente, os *chips* foram levados ao forno de hibridação (*GeneChip® Hybridization Oven 640*; Affymetrix), onde permaneceram por um período de 16 horas a 50°C e 60 rpm.

A – 8 Lavagem e marcação

Reagentes e soluções

Affymetrix Inc., Santa Clara, CA, USA:

Tampão de lavagem A (Wash Buffer A)

Tampão de lavagem B (Wash Buffer B)

Tampão de coloração 1 (Stain Buffer 1)

Tampão de coloração 2 (Stain Buffer 2)

Tampão do *array* (Array Holding Buffer)

Procedimento

Imediatamente após os *chips* serem retirados do forno de hibridação, o conteúdo no *chip* foi removido e substituído por 200 µL de tampão do *array* (Array Holding Buffer). Em seguida, os *chips* foram inseridos na estação de lavagem (*GeneChip® Fluidics Station 450*; Affymetrix Inc.), quatro de cada vez. Na estação de lavagem foram executados dois passos: a remoção do excesso de marcador e coloração. Para tal, dois recipientes contendo os tampões de lavagem A e B foram posicionados na parte lateral do equipamento. Os tampões de coloração 1 e 2 e o tampão do *array* foram distribuídos em respectivos tubos do tipo Eppendorf de 1,5 mL. Cada conjunto de três tubos contendo os reagentes citados correspondeu a um *chip* e foi posicionado na parte frontal da estação. O programa para execução automática das etapas de lavagem e coloração dos *chips Cytoscan HD™* foi selecionado e teve duração de duas horas para cada conjunto de quatro *chips*.

A – 9 Escaneamento

Procedimento

O equipamento *GeneChip® Scanner 3000 7G* foi utilizado para gerar os resultados da varredura completa do genoma. São utilizados programas específicos para captação de imagem do *GeneChip Cytoscan HD array* analisado. A imagem dessa leitura pode ser observada com o programa de visualização *GeneChip® Operating software* (GCOS) (Affymetrix®), no qual pode-se verificar o padrão das sondas controle do sistema, dispostas nas extremidades e no centro de cada *chip*, de forma a se comprovar que as condições de hibridação foram adequadas.

ANEXO B – Aprovação do comitê de ética

B – 1 Aprovação comitê de ética HCFMRP-USP – Parte 1

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

Ribeirão Preto, 15 de agosto de 2012

Ofício nº 2880/2012
CEP/MGV

Prezados Senhores,

O trabalho intitulado **“CNVs EM PACIENTES COM LÚPUS ERITEMATOSO SISTÊMICO”**, foi analisado pelo Comitê de Ética em Pesquisa, em sua 351ª Reunião Ordinária realizada em 13/08/2012, e enquadrado na categoria: **APROVADO**, bem como a solicitação de dispensa do Termo de Consentimento Livre e Esclarecido, de acordo com o Processo HCRP nº 8050/2012.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 196/96 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

DRª. MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores

FERNANDA BUENO BARBOSA

PROF.DR. AGUINALDO LUIZ SIMÕES(Orientador)

Depto. de Genética – FMRP-USP

B – 2 Aprovação comitê de ética HCFMRP-USP – Parte 2

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

Ribeirão Preto, 03 de agosto de 2016

Ofício nº 2670/2016
CEP/MGV

Prezados Senhores,

O trabalho intitulado **“ASSOCIAÇÃO DE CNVS NA SUSCEPTIBILIDADE AO LÚPUS ERITEMATOSO SISTÊMICO”**, foi analisado pelo Comitê de Ética em Pesquisa, em sua 433ª Reunião Ordinária realizada em 01/08/2016 e enquadrado na categoria: APROVADO, bem como o Biorrepositório e a solicitação de dispensa do Termo de Consentimento Livre e Esclarecido, de acordo com o Processo HCRP nº 9158/2016.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (IGH-GCP), bem como a Resolução nº 466/12 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente,

DR. MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores

FERNANDA BUENO BARBOSA

PROF.DR.AGUINALDO LUIZ SIMÕES(Orientador)

Depto. de Genética da FMRP-USP

B – 3 Biorrepositório “LÚPUS E VARIANTES ESTRUTURAIS”***CARTA DE ENCAMINHAMENTO AO COMITÊ DE ÉTICA EM PESQUISA - CEP*****HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA DE RIBEIRÃO
PRETO DA UNIVERSIDADE DE SÃO PAULO****Ao Comitê de Ética em Pesquisa do HCFMRP-USP e da FMRP-USP**

Vem-se por meio desta solicitar a autorização do Comitê de Ética em Pesquisa do HCFMRP-USP e da FMRP-USP para a criação de um biorrepositório denominado: “LÚPUS E VARIANTES ESTRUTURAIS”. Esse biorrepositório será de responsabilidade do Prof. Dr. Aguinaldo Luiz Simões e da pós-graduanda Fernanda Bueno Barbosa, vinculados ao Departamento de Genética desta instituição, estando o referido Departamento ciente e de acordo com a criação.

O biorrepositório denominado “LÚPUS E VARIANTES ESTRUTURAIS” tem como objetivo o armazenamento de amostras de DNA para fins de pesquisa e análise científica, visto que há necessidade de confirmação das variantes no número de cópias (CNVs) encontradas em maior grupo amostral, vinculado ao projeto de pesquisa denominado: “Associação de *CNVs* na susceptibilidade ao lúpus eritematoso sistêmico”.

Tal armazenamento será realizado da seguinte forma:

- serão armazenadas 140 amostras de DNA de pacientes com lúpus eritematoso sistêmico que pertencem ao Banco de Amostras do Núcleo de Pesquisa em Imunogenética — BANPI (CEP HCMRP-USP: processo nº 7581/2007) sob responsabilidade do Prof. Dr. Eduardo Antônio Donadi (FMRP-USP);
- e, 160 amostras de DNA de indivíduos controle referentes ao processo CAAE: 35316314.9.1001.5404, sob responsabilidade da Profª. Drª. Vera Lúcia Gil da Silva Lopes (FCM-UNICAMP);

no Laboratório de Genética Bioquímica (Bloco B), no Departamento de Genética - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo Av. Bandeirantes, 3900 - Monte Alegre - Ribeirão Preto - SP. Todas as amostras serão armazenadas em tubos devidamente etiquetados, dentro de freezer a -20° C, no Laboratório

de Genética Bioquímica. Ao final do projeto, e com a devida autorização do CEP, as amostras componentes deste biorrepositório poderão retornar aos seus bancos de origem.

Fica sob a responsabilidade do Prof. Dr. Aguinaldo Luiz Simões e da pós-graduanda Fernanda Bueno Barbosa a coleta, o armazenamento, a catalogação, a sistematização, a guarda e o uso do material estocado. E se comprometem a identificar as amostras e os dados coletados de modo que garanta o sigilo e a confidencialidade dos doadores, identificando as amostras por meio de códigos.

Toda coleta e armazenamento de material serão precedidas de um "termo de consentimento de guarda de material" assinado pelo doador ou responsável pelo material a ser armazenado. Além disso, serão criados mecanismos que assegurem a possibilidade de contato com os doadores para fornecimento de informação de seu interesse.

Os responsáveis declaram que conhecem e se comprometem a seguir as resoluções do CNS vigentes no país referentes à normas éticas de pesquisa (Resolução 466/2012) e de criação e utilização de biobancos / biorrepositórios (Resolução nº 441/ 2011).

Os responsáveis pelo biorrepositório declaram que a pesquisa a ser feita com o material será submetida à aprovação do CEP desta instituição e, quando for o caso, da Comissão Nacional de Ética em Pesquisa (CONEP).

Prof. Dr. Aguinaldo Luiz Simões
Livre-Docente, Departamento de Genética
FMRP-USP

Ribeirão Preto, 23 de maio de 2016.

Me. Fernanda Bueno Barbosa
Doutoranda, Departamento de Genética
FMRP-USP

Ciência e concordância da chefia do Departamento:

Prof. Dr. Wilson Araújo da Silva Júnior
Chefe do Departamento de Genética
FMRP-USP

Wilson Araújo da Silva Júnior
Chefe do Departamento de Genética
FMRP-USP

Ribeirão Preto, 04 de julho de 2016.

ANEXO C – Manuscritos

C – 1 Comprovante de publicação artigo 01

Short Communication

doi: 10.1111/ahg.12208

Ancestry Informative Marker Panel to Estimate Population Stratification Using Genome-wide Human Array

Fernanda B. Barbosa¹, Natalia F. Cagnin¹, Milena Simioni², Allysson A. Farias³, Fábio R. Torres², Miriam C. Molck², Tânia K. Araujo², Vera L. Gil-Da-Silva-Lopes², Eduardo A. Donadi⁴ and Aguinaldo L. Simões^{1*}

¹Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil

²Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Brazil

³Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São Paulo, Brazil

⁴Division of Clinical Immunology, Department of Medicine, Ribeirão Preto Medical School, University of São Paulo, Brazil

Summary

Case-control studies are a powerful strategy to identify candidate genes in complex diseases. In admixed populations, association studies can be affected by population stratification, leading to spurious genetic associations. Ancestry informative markers (AIMs) can be used to minimise this effect. The aim of this work was to select a set of AIMs to estimate population stratification in a Brazilian case-control study performed using a genome-wide array. A total of 345 single nucleotide polymorphism (SNP) AIMs, selected from the Cytoscan HD array and based on previously reported panels, was used to discriminate between European, African, and Amerindian populations. These SNP-AIMs were used to infer ancestry in systemic lupus erythematosus (SLE) patients ($n = 23$) and in healthy subjects ($n = 110$). Moderate population substructure was observed between SLE and control groups ($F_{st} = 0.0113$). Although patients and controls have shown a major European genomic contribution, significant differences in the European ($P = 6.47 \times 10^{-5}$) and African ($P = 1.14 \times 10^{-3}$) ancestries were detected between the two groups. We performed a two-step validation of the 345 SNP-AIMs panel estimating the ancestral contributions using a panel of 12 AIMs and approximately 70K SNPs from the array. Evaluation of population substructure in case-control studies, avoiding spurious genetic associations, can be performed using our panel of 345 SNP-AIMs.

Keywords: AIMs, admixed population, case-control studies, systemic lupus erythematosus, Cytoscan HD array

C – 2 Artigo publicado**1 Ancestry informative marker panel to estimate population stratification using genome-
2 wide human array**

3 FERNANDA B BARBOSA¹, NATALIA F CAGNIN¹, MILENA SIMIONI², ALLYSSON A
4 DE FARIAS³, FÁBIO R TORRES², MIRIAM C MOLCK², TÂNIA K ARAUJO², VERA L
5 GIL-DA-SILVA-LOPES², EDUARDO A DONADI⁴, AGUINALDO L SIMÕES¹.

6 ¹Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil

7 ²Department of Medical Genetics, Faculty of Medical Sciences, University of Campinas, Brazil

8 ³Department of Genetics and Evolutionary Biology, Biosciences Institute, University of São
9 Paulo, Brazil

10 ⁴Division of Clinical Immunology, Department of Medicine, Ribeirão Preto Medical School,
11 University of São Paulo, Brazil

12 Corresponding author

13 Aguinaldo Luiz Simões, Department of Genetics, Ribeirão Preto Medical School, University
14 of São Paulo, Avenida Bandeirantes, 3900, 14090-190, Monte Alegre, Ribeirão Preto, SP,
15 Brazil.

16 Tel: +55 16 36023050

17 Email: alsimoes@fmrp.usp.br

18 Keywords: AIMs, admixed population, case-control studies, systemic lupus erythematosus,
19 Cytoscan HD array.

20

21

22 Summary

23 Case-control studies are a powerful strategy to identify candidate genes in complex diseases.
 24 In admixed populations, association studies can be affected by population stratification, leading
 25 to spurious genetic associations. Ancestry informative markers (AIMs) can be used to minimise
 26 this effect. The aim of this work was to select a set of AIMs to estimate population stratification
 27 in a Brazilian case-control study performed using a genome-wide array. A total of 345 single
 28 nucleotide polymorphism (SNP) AIMs, selected from the Cytoscan HD array and based on
 29 previously reported panels, was used to discriminate between European, African and
 30 Amerindian populations. These SNP-AIMs were used to infer ancestry in systemic lupus
 31 erythematosus (SLE) patients ($n = 23$) and in healthy subjects ($n = 110$). Moderate population
 32 substructure was observed between SLE and control groups ($F_{st} = 0.0113$). Although patients
 33 and controls have shown a major European genomic contribution, significant differences in the
 34 European ($P = 6.47 \times 10^{-5}$) and African ($P = 1.14 \times 10^{-3}$) ancestries were detected between the
 35 two groups. We performed a two-step validation of the 345 SNP-AIMs panel estimating the
 36 ancestral contributions using a panel of 12 AIMs and ~70K SNPs from the array. Evaluation
 37 of population substructure in case-control studies, avoiding spurious genetic associations, can
 38 be performed using our panel of 345 SNP-AIMs.

39

40 INTRODUCTION

41 Studies investigating genetic background have been performed to identify the genetic variants
 42 contributing to susceptibility for complex human diseases (Lin & Zeng, 2011). In the classic
 43 design of case-control studies, allele frequencies between disease and healthy subjects are
 44 compared (Farhi et al., 2016, Sacks et al., 2014); however, spurious associations not related to
 45 causative *loci* can be obtained by unmeasured population substructure (Coelho et al., 2015).

The issue of unobserved confounding effects appears when the investigated population is composed of several ancestral subpopulations with different allele frequencies and disease-risk not equally represented in cases and controls. This bias on allele frequencies can result in false-positive associations during statistical analyses (Balding, 2006, Cardon & Palmer, 2003, Shriner et al., 2011, Tang et al., 2005).

Ancestry informative markers (AIMs), defined as markers exhibiting differential allele frequencies (δ) greater than 30% in any pair of parental populations, have been used to minimise effects of population stratification in genetic studies (Bonilla et al., 2004, Enoch et al., 2006, Galanter et al., 2012).

In this context, the admixed Brazilian population poses a particular challenge due to five centuries of admixture between individuals from three continents: Europeans settlers, enslaved Africans and Brazilian Amerindians in different proportions throughout the Brazilian territory (Lins et al., 2010). The establishment of a set of AIMs is a critical step to perform reliable genome-wide association studies of complex diseases in admixed populations to infer ancestry and to adjust for population stratification in case-control studies.

Using a whole genome-wide DNA microarray, that encompasses 750,000 single nucleotide polymorphisms (SNPs), we reported 345 markers that are able to evaluate accurately the ancestral composition of the Brazilian population in case-control studies with no further experiments.

MATERIAL AND METHODS

Subjects

A patient group of 23 unrelated SLE individuals (18 women) was followed-up at the Collagen Disease Outpatient Clinic of the University Hospital of the Ribeirão Preto Medical School,

University of São Paulo, SP, Brazil. Patients were diagnosed according to the American College of Rheumatology revised criteria for SLE (Hochberg, 1997). In addition, a control group of 110 healthy unrelated subjects (70 women) from the same geographical region was also evaluated.

This study was approved by the local Research Ethics Committees (CAAE: 03199712.0.0000.5440 and 35316314.9.1001.5404).

Design of SNP-AIMs panel

DNA was extracted from blood samples using either salting-out (Miller et al., 1988) or a commercial kit (QIAGEN Inc, Valencia, CA, USA). In order to genotype patients and controls, a genome-wide screening using Cytoscan HD array (Affymetrix, Santa Clara, CA, USA) was performed according to the manufacturer's protocol. Quality control of the chip was carried out using the Chromosome Analysis Suite v. 3.0 (ChAS, Affymetrix). Single nucleotide polymorphisms (SNPs) were mapped according to the human genome reference assembly based on National Centre for Biotechnology Information build 37 (UCSC version hg19).

The design of the AIMs panel was performed by selecting autosomal SNPs from the Cytoscan HD array that are able to distinguish European, African and/or Amerindian populations and which were previously classified as AIMs (Galanter et al., 2012, Halder et al., 2008, Kosoy et al., 2009, Lins et al., 2010, Shriver et al., 2003).

Hardy-Weinberg equilibrium, population differentiation and ancestry composition

SNP-AIMs genotypes for each subject were exported from ChAS. Allele frequencies and deviations from Hardy-Weinberg equilibrium (HWE) were calculated using Plink v. 1.9 tool (Purcell et al., 2007). Population differentiation (Wright's F-statistics, F_{st}) was calculated using the GDA 1.1 software (Weir & Cockerham, 1996).

In order to estimate ancestry, two types of analysis were performed: (a) evaluation of group ancestry (case and controls) and (b) evaluation of individual ancestry. The ancestral composition of SLE and control groups was estimated by the Admix 95 software (Chakraborty, 1985), based on the allele frequency of all SNP-AIMs included in the panel using the European, African and Amerindian frequencies as parental reference populations. The individual ancestral composition based on the same SNP-AIMs set was estimated with the Admixture v. 1.23 software (Alexander et al., 2009), a Bayesian model-based algorithm used for clustering the genetic data. For the ancestral inference analysis, genotype frequencies of European and African parental populations were obtained from the HapMap project data (CEU – Utah residents exhibiting Northern and Western European ancestry and YRI – Yoruba in Ibadan, Nigeria). Brazilian Amerindian reference population was based on the Karitiana and Surui populations from the HGDP-CEPH Human Genome Diversity Cell Line Panel.

To evaluate if the ancestral proportions of European, African and Amerindian components are different between cases and controls, the Wilcoxon Rank sum test with continuity correction was performed using the information provided by Admixture software.

Individual ancestry data was also used to compare the self-declared ancestry with the ancestry determined by genetic markers in the SLE group. Patients declared themselves as belonging to one of three groups: White, *Pardo* – the intermediate phenotype between White and Black – or Black, according to *IBGE* (Brazilian Institute of Geography and Statistics).

Validation of the SNP-AIMs panel

To validate the SNP-AIMs panel, the ancestral composition results were compared with another panel of AIMs previously evaluated by our group in SLE patients (Cagnin et al, manuscript in preparation). This second panel encompass 12 AIMs previously described by Shriver et al.

(1997) that are able to distinguish European, African and/or Amerindian groups. Of these markers, five were genotyped by PCR-RFLP as SNPs (CKM rs4884:C>T, DRD-2 rs1079598:C>T, FY-Null rs2814778:A>G, LPL rs285:C>T, RB2300 rs2252544:C>T) and seven were genotyped by conventional PCRs, including *Alu* insertions (Sb19.3 rs3138524, APO rs3138522, PV92 rs3138523) and insertions/deletions polymorphisms (AT3-I/D rs3138521, MID-52 rs16344, MID-93 rs16383, MID-575 rs140864). The ancestry composition of SLE patients was estimated using the Admix 95 software and then compared with results achieved for the SNP-AIMs panel. European and African parental allele frequency for the 12 AIMs panel were based on data reported by Shriver et al. (1997). Parental frequency of Amerindians was calculated based on 82 Brazilian Amerindian individuals from the following groups: Baniwa (n = 40), Tikuna (n = 16), Kanamari (n = 14) and Kashinawa (n = 12), using the same set of 12 markers.

A second validation step was performed using the entire genome-wide dataset from Cytoscan HD array (~750K SNPs). Plink v2 (Chang et al., 2015) was used to control the quality and manage the merged data. The total number of SNPs retained after the first quality control using ChAS was 749,157 autosomal markers for 133 individuals (SLE and Brazilian control samples). After merging data with reference samples from HapMap and HGDP and filtering for call rate (< 99%), minor allele frequency (MAF < 0.01), Hardy Weinberg equilibrium deviation ($P < 10^{-8}$) and linkage disequilibrium ($r^2 > 0.8$), the final number of autosomal SNPs was 64,399 (~70K).

The individual ancestry estimates of European, African and Amerindian components determined by Admixture in SLE and control groups were compared with those obtained for the subset of selected SNP-AIMs.

Statistical analysis adjusting for population stratification

To illustrate how the association of SLE with specific *loci* could be affected by the admixture levels, statistical analyses adjusting for ancestry estimates using data from the SNP-AIMs panel were performed in SLE patients and controls.

This approach consisted in the selection of 30 additional SNP-AIMs ($\delta > 30\%$), not included among the 345 SNP-AIMs panel, and 371 SNPs among the ~750K SNPs from the Cytoscan HD array located in ten candidate genes for SLE: *ETS1*, *FAM167A*, *IRF5*, *JAZF1*, *MECP2*, *PTPN22*, *PXK*, *SCUBE1*, *STAT4*, *XKR6* (Dang et al., 2014, Gateva et al., 2009, Harley et al., 2009, Harley et al., 2008). Individual genotypes were extracted from ChAS. The allele frequency in SLE patients and controls was calculated and then compared by Fisher's Exact Test using Plink. Using a cutoff of $P < 0.05$, three SNP-AIMs and ten SNPs were selected for logistic regression analysis adjusting for ancestry estimates. Data about the individual proportions of ancestry obtained with the Admixture software for the 345 SNP-AIMs panel were added as a continuous quantitative independent variable into the logistic regression models examined using the R package (<http://www.R-project.org/>). Three independent models correcting for European (EUR), African (AFR) and Amerindian (AME) ancestries were tested. In order to identify false positive associations, adjusted P values for each ancestral component obtained by the multivariate analysis (regression models) were used for comparisons with results from the non-adjusted regression models.

RESULTS

SNP-AIMs panel

Searching the literature for biallelic SNP-AIMs that are able to differentiate the three Brazilian parental groups (European, African and Amerindian) resulted in the identification of five AIM panels composed of 806 SNP-AIMs (Galanter et al., 2012, Halder et al., 2008, Kosoy et al.,

2009, Lins et al., 2010, Shriver et al., 2003) (Table S1). Among these, 345 SNP-AIMs were included in the 750,000 SNPs from the Cytoscan HD array. This set of SNPs is able to discriminate the three population pairs (EUR-AFR, AFR-AME and AME-EUR) (Figure 1, Table S2).

The probe list for ChAS corresponding to each SNP-AIM included in the final panel of AIMs is available for download as a supplementary file (list_of_345_probes_to_ChAS.txt).

Population differentiation and ancestral composition

Deviations from HWE were detected in six *loci* for the SLE group and in 14 *loci* for control group ($P < 0.05$) (Table S3). However, after Bonferroni correction, no deviations from HWE were observed (significance threshold of 10^{-4}).

Population differentiation analyses revealed that patients and controls exhibited a moderate genetic differentiation ($F_{st} = 0.0113$, 95% CI 0.0081 – 0.0150).

Ancestral composition analysis using Admix 95 software revealed that the main component of patients and control group was European ($66.2\% \pm 0.11\%$ and $77.8\% \pm 0.07\%$), followed by African ($22.0\% \pm 0.09\%$ and $12.6\% \pm 0.06\%$) and Amerindian ($11.8\% \pm 0.10\%$ and $9.5\% \pm 0.06\%$). Comparisons using the proportions of each ancestral component obtained from Admixture software showed significant differences in European ($\text{median}_{\text{SLE}} = 66.1\%$, $\text{median}_{\text{BRZ}} = 79.3\%$, $P = 6.47 \times 10^{-5}$, 95% CI -0.203 – -0.079) and African ($\text{median}_{\text{SLE}} = 23.1\%$, $\text{median}_{\text{BRZ}} = 9.8\%$, $P = 1.14 \times 10^{-3}$, 95% CI 0.034 – 0.160) ancestries between SLE and control groups. No significant difference was found in the Amerindian ancestry ($\text{median}_{\text{SLE}} = 12.7\%$, $\text{median}_{\text{BRZ}} = 10.0\%$, $P = 0.09$, 95% CI -0.005 – 0.060).

Regarding the self-declared ancestry in SLE group subjects that self-declared as White showed a greater European contribution ($64.5\% \pm 14.0\%$). However, significant contributions from

African ($20.3\% \pm 13.2\%$) and Amerindian ($15.2\% \pm 9.5\%$) ancestries were identified in the same cohort of individuals. Subjects who self-declared as Pardo showed high proportions of African ($36.1\% \pm 17.8\%$) and European ($60.3\% \pm 14.8\%$) ancestries. The only SLE patient that self-declared as Black showed a greater African ancestry (52.2%); however, the patient also had a significant contribution of European (38.9%) ancestry.

Validation of the SNP-AIM panel

Similar results were observed between the ancestry estimates in the SLE group obtained using the 345 SNP-AIMs and 12 AIMs panels (Table 1). Individual ancestry estimates for the European, African and Amerindian components using ~70K SNPs from Cytoscan HD array showed similar results compared with ancestry contributions obtained with the subset of 345 SNP-AIMs (Figure 2).

Identification of spurious associations in SLE

The effect of admixture levels on the results of an association study was assessed using three additional SNP-AIMs that differentiate between European and African populations and 10 SNPs selected from a set of 371 SNPs located in candidate genes for SLE, whose case-control associations were significant in Fisher's Exact Test ($P < 0.05$). We ran ancestry-adjusted logistic regression models to account for the effect of population stratification due to admixture (non-adjusted model, model 1 = adjusted for African ancestry, model 2 = adjusted for European ancestry, model 3 = adjusted for Amerindian ancestry).

After adjusting the models to assess the influence of ancestry, we demonstrated that SNP-AIMs showing significant association to SLE in the non-adjusted regression model became not significant after correction for European and African ancestries. In relation to the SNPs located in candidate genes for SLE, significant associations still remained for 8/10 SNPs (model 1),

7/10 SNPs (model 2) and 10/10 SNPs (model 3) (Table 2). Therefore, using this approach, we identified possible false positive associations, highlighting the effect of population stratification in genetic analysis.

DISCUSSION

High-density microarray platforms, developed for simultaneous detection of copy number variations and SNPs, are able to genotype biallelic markers and to identify structural variations (Uddin et al., 2014). The inclusion of SNPs with high values of MAF among the Cytoscan HD array SNP markers, i.e. common variants in the general population, allowed the identification of 345 AIMs, previously reported as suitable to distinguish between European, African and Amerindian populations (Galanter et al., 2012, Halder et al., 2008, Kosoy et al., 2009, Lins et al., 2010, Shriver et al., 2003).

Case-control studies of complex diseases in admixed populations require the use of AIMs to avoid spurious associations (Shriner et al., 2011). The ancestry estimation in SLE patients and healthy individuals based on a set of 345 SNP-AIMs showed that both groups had major European ancestral contributions, followed by African and Amerindian (Figure 2). This finding is consistent with the ancestral composition of the Brazilian South-eastern population that exhibited 60–80% of European contribution followed by minor contributions from African and Amerindian ancestral populations (Egan et al., 2017, Ferreira et al., 2006, Muniz et al., 2008, Pena et al., 2009, Rolim et al., 2016).

Although patients and controls have shown a major European genomic contribution, significant differences in the European and African ancestries were detected among the two groups. SLE patients have a higher African contribution (22%) than healthy subjects (13%), while controls showed a European contribution 12% higher than SLE patients. This difference throws light upon the moderate population substructure detected between the case and control groups. The

higher proportion of African component in SLE group seems to disagree with the low incidence of this disease in African populations (Petri, 2002, Pons-Estel et al., 2010). However, it is consistent with previous reports showing that Afro-Americans are more susceptible to SLE than Euro-Americans (Borchers et al., 2010, Danchenko et al., 2006). These results emphasise the diversity of the disease, considered to be a genetically complex autoimmune disorder that exhibits diverse incidence rates according to the ancestry group and presenting heterogeneous clinical and laboratory features.

Comparisons between genetic and self-declared ancestry in SLE patients showed at least 30% of non-declared ancestry background, including African/Amerindian in Whites and European/Amerindian in Blacks. This divergence between the self-declared ancestry and the individual genetic background is in agreement with previous reports describing the genetic structure of the Brazilian population (Cardena et al., 2013, Giolo et al., 2012, Leite et al., 2011, Lins et al., 2011). This finding underscores how crucial it is to evaluate ancestry based on genetic markers in addition to using the perception of individuals about their phenotypic characteristics.

There are differences between the 345 SNP-AIMs and the 12 AIMs panels. The 12 AIMs panel used a Brazilian Amerindian population as a reference, while the 345 SNP-AIMs panel used Amerindian populations of other origins. In addition, the 12 AIMs panel was genotyped by PCR and PCR-RFLP and also includes genetic markers such as *InDels* and *Alu* insertions. In spite of these differences, the results showed a strong concordance between both sets of AIMs, which act as a mutual validation of these two panels and demonstrate the applicability of this methodology to infer ancestry in studies performed by Cytoscan HD array (Table 1). In addition to this validation step, the ancestry analysis using ~70K SNPs from the Cytoscan HD array in comparison with the results obtained using the subset of 345 SNP-AIMs showed

Pearson correlation values > 0.99 for the European, African and Amerindian ancestries (Figure 2).

The 345 SNP-AIMs panel has the potential to be widely used in molecular genetics and cytogenetic research to study diseases whose incidence is affected by ancestry, and has the advantage of using data from the array with no requirement of further experiments for the Cytoscan users. Another noteworthy advantage of using a panel of AIMs to infer ancestry is to insert the information of individual proportions of each parental component as an independent variable in statistical logistic regression models. This is especially suitable to perform corrections for ancestry in case-control studies as demonstrated in this article. We have found significant associations of SNP-AIMs to SLE in non-adjusted models that were corrected after adjustment of the regression models to assess the influence of ancestry, demonstrating a way to efficiently correct for confounding variables in case-control studies. From the set of SNPs selected for case-control analysis, only the Amerindian component had no effect in association of the SNPs adjusted for ancestry, pointing out that the effect of the population substructure is due to differences in the proportions of European and African strata among case and control groups.

In conclusion, spurious associations resulting from population stratification may be circumvented by using the 345 SNP-AIMs panel here described as a straightforward and efficient method to infer ancestry in case-control studies.

Funding

This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Proc No. 2011/23794-7), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Fundação de Apoio ao Ensino, Pesquisa e Assistência do HCFMRP-USP, and

279 Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). FBB (2013/17062-
 280 9) is currently supported by a research fellowship from FAPESP, while ALS (312547/2009-9),
 281 NFC (142439/2011-8), EAD (304931/2014-4) and VLGSL (304455/2012-1) are supported by
 282 CNPq/Brazil.

283 **Conflict of interest statement**

284 The authors have no conflicts of interest to declare.

285 **Acknowledgments**

286 We are grateful to Dr Íscia L Cendes (Faculty of Medical Sciences, University of Campinas)
 287 who provided the healthy samples to Cytoscan HD array and Brazilian Biosciences National
 288 Laboratory (Campinas, SP) for technical support. Study design: FBB, ALS; Data collection:
 289 FBB, NFC, FRT, MCM, TKA; Data analysis: FBB, NFC, AAF; Manuscript preparation: FBB,
 290 ALS, EAD, MS, FRT; Contribution of experimental material: ALS, VLGSL, EAD.

291 **References**

- 292 Alexander, D.H., Novembre, J. & Lange, K. (2009) Fast model-based estimation of ancestry in
 293 unrelated individuals. *Genome Research*, 19, 1655-1664.
- 294 Balding, D.J. (2006) A tutorial on statistical methods for population association studies. *Nature*
 295 *Reviews Genetics*, 7, 781-91.
- 296 Bonilla, C., Shriver, M.D., Parra, E.J., Jones, A. & Fernández, J.R. (2004) Ancestral proportions
 297 and their association with skin pigmentation and bone mineral density in Puerto Rican
 298 women from New York city. *Human Genetics*, 115, 57-68.
- 299 Borchers, A.T., Naguwa, S.M., Shoenfeld, Y. & Gershwin, M.E. (2010) The geoepidemiology
 300 of systemic lupus erythematosus. *Autoimmunity Reviews*, 9, A277-A287.

- 301 Cardena, M.M., Ribeiro-Dos-Santos, A., Santos, S., Mansur, A.J., Pereira, A.C. & Fridman, C.
 302 (2013) Assessment of the relationship between self-declared ethnicity, mitochondrial
 303 haplogroups and genomic ancestry in Brazilian individuals.
- 304 Cardon, L.R. & Palmer, L.J. (2003) Population stratification and spurious allelic association.
 305 *The Lancet*, 361, 598-604.
- 306 Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015) Second-
 307 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4,
 308 7.
- 309 Chakraborty, R. (1985) Genetic microdifferentiation in human and others animal population.
 310 *Indian Anthropological Association*, 171 – 180.
- 311 Coelho, A., Moura, R., Cavalcanti, C., Guimarães, R., Sandrin-Garcia, P., Crovella, S. &
 312 Brandão, L. (2015) A rapid screening of ancestry for genetic association studies in an
 313 admixed population from Pernambuco, Brazil. *Genetics and Molecular Research: GMR*, 14, 2876.
- 315 Danchenko, N., Satia, J. & Anthony, M. (2006) Epidemiology of systemic lupus erythematosus:
 316 a comparison of worldwide disease burden. *Lupus*, 15, 308-318.
- 317 Dang, J., Shan, S., Li, J., Zhao, H., Xin, Q., Liu, Y., Bian, X. & Liu, Q. (2014) Gene–gene
 318 interactions of IRF5, STAT4, IKZF1 and ETS1 in systemic lupus erythematosus. *Tissue*
 319 *Antigens*, 83, 401-408.
- 320 Egan, K.J., Campos Santos, H., Bejamini, F., Duarte, N.E., Horimoto, A.R., Taporoski, T.P.,
 321 Vallada, H., Negrão, A.B., Krieger, J.E. & Pedrazzoli, M. (2017) Amerindian (but not
 322 African or European) ancestry is significantly associated with diurnal preference within
 323 an admixed Brazilian population. *Chronobiology International*, 1-4.

- 324 Enoch, M.-A., Shen, P.-H., Xu, K., Hodgkinson, C. & Goldman, D. (2006) Using ancestry-
 325 informative markers to define populations and detect population stratification. *Journal*
 326 *of Psychopharmacology*, 20, 19-26.
- 327 Farhi, A., Cohen, A.D., Shovman, O., Comaneshter, D., Amital, H. & Amital, D. (2016) Bipolar
 328 disorder associated with rheumatoid arthritis: A case-control study. *Journal of Affective*
 329 *Disorders*, 189, 287-289.
- 330 Ferreira, L.B., Mendes, C.T., Wiezel, C.E.V., Luizon, M.R. & Simões, A.L. (2006) Genomic
 331 ancestry of a sample population from the state of Sao Paulo, Brazil. *American Journal*
 332 *of Human Biology*, 18, 702-705.
- 333 Galanter, J.M., Fernandez-Lopez, J.C., Gignoux, C.R., Barnholtz-Sloan, J., Fernandez-
 334 Rozadilla, C., Via, M., Hidalgo-Miranda, A., Contreras, A.V., Figueroa, L.U. & Raska,
 335 P. (2012) Development of a panel of genome-wide ancestry informative markers to
 336 study admixture throughout the Americas. *PLoS Genetics*, 8, e1002554.
- 337 Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy,
 338 R., Ferreira, R.C. & Nordmark, G. (2009) A large-scale replication study identifies
 339 TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus
 340 erythematosus. *Nature Genetics*, 41, 1228-1233.
- 341 Giolo, S.R., Soler, J.M., Greenway, S.C., Almeida, M.A., De Andrade, M., Seidman, J.,
 342 Seidman, C.E., Krieger, J.E. & Pereira, A.C. (2012) Brazilian urban population genetic
 343 structure reveals a high degree of admixture. *European Journal of Human Genetics*, 20,
 344 111-116.
- 345 Halder, I., Shriver, M., Thomas, M., Fernandez, J.R. & Frudakis, T. (2008) A panel of ancestry
 346 informative markers for estimating individual biogeographical ancestry and admixture
 347 from four continents: utility and applications. *Human Mutation*, 29, 648-658.

- 348 Harley, I.T.W., Kaufman, K.M., Langefeld, C.D., Harley, J.B. & Kelly, J.A. (2009) Genetic
 349 susceptibility to SLE: new insights from fine mapping and genome-wide association
 350 studies. *Nature Reviews Genetics*, 10, 285-290.
- 351 Harley, J.B., Alarcón-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser,
 352 K.L., Tsao, B.P., Vyse, T.J. & Langefeld, C.D. (2008) Genome-wide association scan
 353 in women with systemic lupus erythematosus identifies susceptibility variants in
 354 ITGAM, PTK, KIAA1542 and other loci. *Nature Genetics*, 40, 204-210.
- 355 Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for
 356 the classification of systemic lupus erythematosus. *Arthritis & Rheumatism*, 40, 1725.
- 357 Kosoy, R., Nassir, R., Tian, C., White, P.A., Butler, L.M., Silva, G., Kittles, R., Alarcon-
 358 Riquelme, M.E., Gregersen, P.K. & Belmont, J.W. (2009) Ancestry informative marker
 359 sets for determining continental origin and admixture proportions in common
 360 populations in America. *Human Mutation*, 30, 69-78.
- 361 Leite, T., Fonseca, R., De França, N.M., Parra, E.J. & Pereira, R.W. (2011) Genomic ancestry,
 362 self-reported "color" and quantitative measures of skin pigmentation in Brazilian
 363 admixed siblings. *PloS One*, 6, e27162.
- 364 Lin, D. & Zeng, D. (2011) Correcting for population stratification in genomewide association
 365 studies. *Journal of the American Statistical Association*, 106, 997-1008.
- 366 Lins, T.C., Vieira, R.G., Abreu, B.S., Gentil, P., Moreno-Lima, R., Oliveira, R.J. & Pereira,
 367 R.W. (2011) Genetic heterogeneity of self-reported ancestry groups in an admixed
 368 Brazilian population. *Journal of Epidemiology*, 21, 240-245.
- 369 Lins, T.C., Vieira, R.G., Abreu, B.S., Grattapaglia, D. & Pereira, R.W. (2010) Genetic
 370 composition of Brazilian population samples based on a set of twenty-eight ancestry
 371 informative SNPs. *American Journal of Human Biology*, 22, 187-192.

- 372 Miller, S., Dykes, D. & Polesky, H. (1988) A simple salting out procedure for extracting DNA
373 from human nucleated cells. *Nucleic Acids Research*, 16, 1215.
- 374 Muniz, Y.C.N., Ferreira, L.B., Mendes-Junior, C.T., Wiezel, C.E.V. & Simões, A.L. (2008)
375 Genomic ancestry in urban Afro-Brazilians. *Annals of Human Biology*, 35, 104-111.
- 376 Pena, S.D., Bastos-Rodrigues, L., Pimenta, J. & Bydlowski, S. (2009) DNA tests probe the
377 genomic ancestry of Brazilians. *Brazilian Journal of Medical and Biological Research*,
378 42, 870-876.
- 379 Petri, M. (2002) Epidemiology of systemic lupus erythematosus. *Best Practice & Research*
380 *Clinical Rheumatology*, 16, 847-858.
- 381 Pons-Estel, G.J., Alarcón, G.S., Scofield, L., Reinlib, L. & Cooper, G.S. (2010) Understanding
382 the epidemiology and progression of systemic lupus erythematosus. In: *Seminars in*
383 *Arthritis and Rheumatism*. Elsevier.
- 384 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J.,
385 Sklar, P., De Bakker, P.I. & Daly, M.J. (2007) PLINK: a tool set for whole-genome
386 association and population-based linkage analyses. *The American Journal of Human*
387 *Genetics*, 81, 559-575.
- 388 Rolim, H., Cronemberger, S., Rangel, H., Batista, W.D., Bastos-Rodrigues, L. & De Marco, L.
389 (2016) The Role of Genetic Ancestry in Brazilian Patients With Primary Congenital
390 Glaucoma. *Journal of Glaucoma*, 25, e24-e28.
- 391 Sacks, F.M., Hermans, M.P., Fioretto, P., Valensi, P., Davis, T., Horton, E., Wanner, C., Al-
392 Rubeaan, K., Aronson, R. & Barzon, I. (2014) Association Between Plasma
393 Triglycerides and High-Density Lipoprotein Cholesterol and Microvascular Kidney
394 Disease and Retinopathy in Type 2 Diabetes Mellitus A Global Case-Control Study in
395 13 Countries. *Circulation*, 129, 999-1008.

- 396 Shriner, D., Adeyemo, A., Ramos, E., Chen, G. & Rotimi, C.N. (2011) Mapping of disease-
 397 associated variants in admixed populations. *Genome Biology*, 12, 223.
- 398 Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C.,
 399 Massac, A. & Cameron, N. (2003) Skin pigmentation, biogeographical ancestry and
 400 admixture mapping. *Human Genetics*, 112, 387-399.
- 401 Shriver, M.D., Smith, M.W., Jin, L., Marcini, A., Akey, J.M., Deka, R. & Ferrell, R.E. (1997)
 402 Ethnic-affiliation estimation by use of population-specific DNA markers. *American*
 403 *Journal of Human Genetics*, 60, 957.
- 404 Tang, H., Quertermous, T., Rodriguez, B., Kardia, S.L., Zhu, X., Brown, A., Pankow, J.S.,
 405 Province, M.A., Hunt, S.C. & Boerwinkle, E. (2005) Genetic structure, self-identified
 406 race/ethnicity, and confounding in case-control association studies. *The American*
 407 *Journal of Human Genetics*, 76, 268-275.
- 408 Uddin, M., Thiruvahindrapuram, B., Walker, S., Wang, Z., Hu, P., Lamoureux, S., Wei, J.,
 409 Macdonald, J.R., Pellecchia, G. & Lu, C. (2014) A high-resolution copy-number
 410 variation resource for clinical and population genetics. *Genetics in Medicine*.
- 411 Weir, B.S. & Cockerham, C. (1996) Genetic data analysis II: Methods for discrete population
 412 genetic data. Sinauer Assoc. Inc., Sunderland, MA, USA.

413 **Supporting Information**

414 **Figure S1.** Tri-hybrid ancestral composition in case-control groups using different panels of
 415 ancestry informative markers (AIMs) and data analysis methods. European, African and
 416 Amerindian ancestry proportions in systemic lupus erythematosus (SLE) (A) and controls (B)
 417 using Admix95/Admixture software and 12/345/70K genetic markers. Blue, red and green
 418 represent the average of inferred ancestry from European, African and Amerindian

contributions, respectively. (*) No information about the ancestry determined by the 12 AIMs panel for the control group (PDF).

Table S1. Number of previously reported single nucleotide polymorphism - ancestry informative markers (SNP-AIMs) that differentiated European, African and Amerindian populations, which were present in the Cytoscan HD array (PDF).

Table S2. Panel of 345 single nucleotide polymorphism - ancestry informative markers (SNP-AIMs) selected to infer genetic ancestry in the Cytoscan HD array (PDF).

Table S3. Allele frequency, Hardy-Weinberg equilibrium and F_{st} results for the 345 single nucleotide polymorphism - ancestry informative markers (SNP-AIMs) in the systemic lupus erythematosus (SLE) and control groups (PDF).

Table S4. Self-declared and genetic ancestry of the systemic lupus erythematosus (SLE) patients included in the study (PDF).

File. list_of_345_probes_to_ChAS (TXT).

Table 1. Ancestry inference for systemic lupus erythematosus patients measured by the two different panels of ancestry informative markers (AIMs) using Admix 95 software.

Reference populations	345 SNP-AIMs panel		12 AIMs panel	
	Ancestry estimate	Standard error	Ancestry estimate	Standard error
European	66.21%	0.11	65.56%	3.18
African	21.97%	0.09	20.97%	3.67
Amerindian	11.82%	0.10	13.45%	2.56

Table 2. Logistic regression models examining the association of systemic lupus erythematosus with single nucleotide polymorphism - ancestry informative markers (SNP-AIMs) and candidate SNPs, adjusted for African, European and Amerindian ancestries.

SNP ID	Type	rs dbSNP	Allele (A1A2)	Non-adjusted model		Model 1 (African)		Model 2 (European)		Model 3 (Amerindian)	
				<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)	<i>P</i>	OR (95% CI)
SNP1a	SNP-AIM	rs4926851	CT	0.004	4.4 (1.6 – 12.1)	<i>0.089</i>	2.7 (0.8 – 8.6)	<i>0.109</i>	2.5 (0.8 – 7.7)	0.005	4.3 (1.5 – 11.9)
SNP2a	SNP-AIM	rs11264110	GA	0.019	3.2 (1.3 – 8.9)	<i>0.110</i>	2.3 (0.8 – 6.7)	<i>0.194</i>	2.0 (0.7 – 6.0)	0.030	3.0 (1.1 – 8.4)
SNP3a	SNP-AIM	rs6663966	CT	0.038	2.9 (1.02 – 7.8)	<i>0.409</i>	1.6 (0.5 – 5.1)	<i>0.570</i>	1.4 (0.4 – 4.5)	<i>0.056</i>	2.7 (0.9 – 7.4)
SNP1b	Candidate / ETS1	rs3935795	CT	0.003	5.0 (1.7 – 14.9)	0.007	4.5 (1.5 – 13.7)	0.004	5.3 (1.7 – 16.8)	0.002	6.1 (2.0 – 19.2)
SNP2b	Candidate/ ETS1	rs12277199	CA	0.027	8.1 (1.3 – 64.5)	<i>0.121</i>	4.8 (0.6 – 41.5)	<i>0.094</i>	5.3 (0.7 – 46.0)	0.021	9.0 (1.4 – 73.1)
SNP3b	Candidate/ FAM167A	rs10503422	CT	0.029	2.8 (1.1 – 7.1)	0.017	3.2 (1.2 – 8.9)	0.010	3.7 (1.4 – 10.7)	0.020	3.1 (1.2 – 8.1)
SNP4b	Candidate/ FAM167A	rs7007712	TC	0.033	2.9 (1.1 – 8.0)	0.021	3.3 (1.2 – 9.9)	0.024	3.3 (1.2 – 9.7)	0.037	2.8 (1.1 – 7.9)
SNP5b	Candidate/ PXX	rs59956777	AG	0.001	5.2 (1.9 – 14.6)	0.002	4.9 (1.8 – 14.4)	0.002	5.3 (1.9 – 16.1)	0.001	5.8 (2.1 – 17.2)
SNP6b	Candidate/ JAZF1	rs202	CT	0.019	3.2 (1.3 – 8.9)	0.016	3.5 (1.3 – 10.4)	0.014	3.6 (1.3 – 10.8)	0.018	3.2 (1.2 – 9.0)
SNP7b	Candidate/ JAZF1	rs78276498	AG	0.005	7.4 (1.8 – 32.3)	0.008	7.1 (1.7 – 36.2)	0.013	6.6 (1.5 – 31.2)	0.007	7.0 (1.7 – 30.8)
SNP8b	Candidate/ JAZF1	rs10951192	GC	0.035	2.9 (1.1 – 8.7)	0.031	3.2 (1.2 – 9.8)	<i>0.051</i>	2.8 (0.9 – 8.6)	0.047	2.8 (1.1 – 8.3)
SNP9b	Candidate/ XKR6	rs17725480	TC	0.025	2.9 (1.2 – 7.8)	0.019	3.3 (1.3 – 9.4)	0.011	3.8 (1.4 – 11.3)	0.017	3.3 (1.3 – 9.1)
SNP10b	Candidate/ XKR6	rs17152782	TC	0.013	7.4 (1.5 – 40.4)	<i>0.192</i>	3.3 (0.5 – 21.3)	<i>0.169</i>	3.3 (0.6 – 20.1)	0.010	8.0 (1.6 – 43.9)

OR = Odds ratio; 95% CI = 95% confidence interval; A1: allele 1; A2: allele 2.

(a) SNP-AIM not included in the 345 SNP-AIMs panel; (b) SNP located in candidate gene for SLE susceptibility.

Dependent variable in the logistic regression analysis: SLE; independent variable: A1, covariate: ancestry estimated using all of the 345 SNP-AIMs.

P values > 0.05 given in italic.

Figure 1. Venn diagram showing the number of AIMs able to differentiate between the three population pairs in the panel of 345 SNP-AIMs.

Figure 2. Ancestry inference in systemic lupus erythematosus patients and healthy subjects. A maximum likelihood approach was applied to estimate individual ancestries using the Admixture software ($K=3$), 345 SNP-AIMs (A) and ~70K SNPs from the Cytoscan HD array (B) in European (EUR [CEU–HapMap], $n = 89$), African (AFR [YRI–HapMap], $n = 89$) and Amerindian (AME [Karitiana and Surui–HGDP-CEPH], $n = 45$) as reference populations, case (SLE, $n = 23$) and control subjects (BRZ, $n = 110$). Each vertical row corresponds to one individual. Blue, red and green represent the proportions of inferred ancestry from European, African and Amerindian contributions, respectively. Correlation plots comparing ancestry estimates of ~70K single nucleotide polymorphisms and the subset of 345 ancestry informative markers (AIMs) for European (C), African (D) and Amerindian (E) components in all subjects included in the ancestry analysis.

C – 3 Comprovante de submissão artigo 02

manuscripttrackingsystem		European Journal of Human Genetics	
tracking system home	author instructions	reviewer instructions	help tips logout journal home
Manuscript #	678-17-EJHG		
Current Revision #	0		
Submission Date	14th Sep 17 10:05:32		
Current Stage	Under Consideration		
Title	Copy number variation and loss of heterozygosity in the susceptibility to systemic lupus erythematosus		
Running Title	CNV and LOH profiles in SLE		
Manuscript Type	Article		
Category	Molecular Genetics		
Manuscript Comment	This is the first submission of this manuscript		
Corresponding Author	Aguinaldo Simões (alsimoes@fmrp.usp.br) (Ribeirão Preto Medical School, Univesity of São Paulo)		
Contributing Authors	Fernanda Barbosa , Milena Simioni , Cláudia Wiezel , Fábio Torres , Miriam Moick , Melvin Bonilla , Tania Araujo , Eduardo Donadi , Vera Lúcia Gil-da-Silva-Lopes , Bernardo Lemos		
Abstract	<p>Systemic lupus erythematosus (SLE) is an autoimmune disease with a strong genetic component and etiology characterized by chronic inflammation and autoantibody production. The purpose of this study was to ascertain copy number variation (CNV) and loss of heterozygosity (LOH) in SLE using a case-control design in an admixed Brazilian population. The detection of CNV and LOH was performed using Cytoscan HD array in SLE patients (n = 23) and controls (n = 110). The best candidates were then evaluated by quantitative PCR in a larger SLE (n = 135) and control (n = 200) cohorts, and confirmed that duplications in <i>FCGR3B</i> and <i>ADAM3A</i> genes are indeed more frequent in control populations. In addition, the data shows a synergistic interaction between the two loci with the presence of deletions in both <i>FCGR3B</i> and <i>ADAM3A</i> loci significantly increasing the risk to SLE. Furthermore, heterozygous deletions overlapping the <i>CFHR4</i>, <i>CFHR5</i> and <i>HLA-DPB2</i> genes in SLE patients were detected by the array and further validated using droplet digital PCR. Overall, 21 rare CNVs were identified in SLE patients using a four-step pipeline created for the identification of rare variants. Regarding the copy-neutral LOH regions, no evidence was found that the total number and length of LOH segments are different between SLE patients and control subjects. Finally, we identified a set of LOH segments restricted to the SLE group. The results described here show that novel susceptibility loci to SLE can be found once the distribution of structural variants is analyzed throughout the whole genome.</p>		
Section Editor	Not Assigned		

C – 4 Manuscrito submetido**Copy number variation and loss of heterozygosity in the susceptibility to systemic lupus erythematosus**

Fernanda B Barbosa¹, Milena Simioni², Cláudia E.V. Wiezel¹, Fábio R Torres², Miriam C Molck², Melvin M Bonilla³, Tânia K Araujo², Eduardo A Donadi⁴, Vera L Gil-da-Silva-Lopes², Bernardo Lemos³, Aguinaldo L Simões¹.

¹Department of Genetics, Ribeirão Preto Medical School, USP, Ribeirão Preto, SP, Brazil;

²Department of Medical Genetics, Faculty of Medical Sciences, UNICAMP, Campinas, SP, Brazil;

³Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA;

⁴Division of Clinical Immunology, Department of Medicine, Ribeirão Preto Medical School, USP, Ribeirão Preto, SP, Brazil.

Corresponding author

Aguinaldo Luiz Simões, Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Avenida Bandeirantes, 3900, 14090-190, Monte Alegre, Ribeirão Preto, SP, Brazil.

Tel: +55 16 36023050

Email: alsimoes@fmrp.usp.br

Keywords: Copy number variation, Loss of heterozygosity, Systemic lupus erythematosus, Rare variants, Admixed population.

ABSTRACT

Systemic lupus erythematosus (SLE) is an autoimmune disease with a strong genetic component and etiology characterized by chronic inflammation and autoantibody production. The purpose of this study was to ascertain copy number variation (CNV) and loss of heterozygosity (LOH) in SLE using a case-control design in an admixed Brazilian population. The detection of CNV and LOH was performed using Cytoscan HD array in SLE patients (n = 23) and controls (n = 110). The best candidates were then evaluated by quantitative PCR in a larger SLE (n = 135) and control (n = 200) cohorts, and confirmed that duplications in *FCGR3B* and *ADAM3A* genes are indeed more frequent in control populations. In addition, the data shows a synergistic interaction between the two loci with the presence of deletions in both *FCGR3B* and *ADAM3A* loci significantly increasing the risk to SLE. Furthermore, heterozygous deletions overlapping the *CFHR4*, *CFHR5* and *HLA-DPB2* genes in SLE patients were detected by the array and further validated using droplet digital PCR. Overall, 21 rare CNVs were identified in SLE patients using a four-step pipeline created for the identification of rare variants. Regarding the copy-neutral LOH regions, no evidence was found that the total number and length of LOH segments are different between SLE patients and control subjects. Finally, we identified a set of LOH segments restricted to the SLE group. The results described here show that novel susceptibility loci to SLE can be found once the distribution of structural variants is analyzed throughout the whole genome.

INTRODUCTION

Systemic lupus erythematosus (SLE, MIM 152700) is an autoimmune polygenic disease characterized by local or systemic inflammation from the production of autoantibodies and immune complex deposition in several tissues ¹. SLE has a wide range of clinical manifestations such as malar rash, discoid lesions, nephritis, and arthritis ². SLE is more

prevalent in women than in men at a rate of 9:1, and onset predominantly during childbearing age³. In general, it is less prevalent in Caucasian populations than in African-Americans, African-Caribbean, Asians, and Hispanics⁴.

In addition to the genetic component; hormonal, environmental, epigenetic and immunological factors contribute to the complex etiology of SLE⁵⁻⁷. High heritability and increased concordance rates were identified in monozygotic twins (24–57%) compared with dizygotic twins or full siblings (2–5%), suggesting that SLE has a complex genetic basis with sizable genetic and environmental components^{8,9}. Linkage and genome-wide association studies conducted in cohorts of patients with SLE have confirmed that *HLA* and other loci are associated with SLE¹⁰. This method has also been useful in identifying new candidate single nucleotide polymorphisms (SNPs) correlated with the disease^{11,12}.

In addition to SNPs, genomic segments that vary in copy number in relation to a reference genome (denoted as copy number variations, or CNVs, and typically greater than 50 bp¹³) have been associated with susceptibility to autoimmune diseases, including SLE². Well-documented CNVs that increase risk for SLE include deletions in *C4*¹⁴ and *FCGR3B*¹⁵ genes, while CNVs in other genes were associated with SLE in single-population studied, e.g. *CCL3L1*¹⁶, *TLR7*¹⁷, *DEFB4*¹⁸, *RABGAP1L*¹⁹ and *HLA-DRB5*²⁰.

Loss of heterozygosity (LOH) are defined as long stretches of consecutive homozygous genotypes²¹. The lack of heterozygosity in continuous and uninterrupted stretches of a DNA sequence in diploid genomes reflect segments that are identical-by-descent²². This results from demographic processes that reduce population size and increase the level of homozygosis throughout generations, such as endogamy, population bottleneck, and natural selection²³. Therefore, LOH can occur without obvious pathological consequences as a result of the evolutionary history of human populations and constituting part of the variation found in

healthy individuals²⁴. On the other hand, LOH mapping has been broadly used as a model for investigating Mendelian diseases by homozygous mapping^{25,26} and successfully to identify candidate regions in some multifactorial diseases^{27,28}.

The small number of large-scale studies relating CNVs and LOH to SLE remains a significant gap in the genetic analysis of the disease^{19,29}. Additionally, the ancestral composition of populations can often modify the results of association tests, such that the study of admixed populations can produce different or even conflicting results compared with those reported in the literature³⁰⁻³².

Based on the hypothesis that new SLE-related loci remain to be discovered using CNV and LOH approaches, this study has evaluated the role of these types of genomic variation in SLE through genome-wide screening in Brazilian SLE patients.

MATERIAL AND METHODS

DNA samples and ethical considerations

The case group comprised 135 unrelated SLE patients treated at the Collagen Disease Outpatient Clinic of the University Hospital at the Ribeirão Preto Medical School (HCFMRP, USP) in Brazil. All patients fulfilled the American College of Rheumatology revised criteria for SLE diagnosis^{33,34}. The control cohort includes 200 healthy unrelated subjects resident in São Paulo state, Brazil. DNA was extracted from the blood samples of SLE patients and control subjects using salting-out method³⁵ and QIAamp® DNA Blood Maxi Kit (QIAGEN, Hilden, Germany), respectively. All patients and controls have signed the consent form approved by the Research Ethics Committee of FMRP/USP and FCM/UNICAMP (CAAE: 03199712.0.0000.5440/ CAAE: 03199712.0.3001.5404).

95 ***Cytoscan HD array***

96 The genome-wide human Cytoscan HD array (Affymetrix, CA, USA) was used to detect CNVs
 97 and LOH regions in SLE patients (n = 23) and controls (n = 110) according to the
 98 manufacturer's protocol. Scanned data files were generated using Affymetrix® GeneChip
 99 Command Console software v. 1.2 and analyzed by Affymetrix® Chromosome Analysis Suite
 100 (ChAS) software v. 3.0.

101 ***CNV and LOH detection***

102 To calculate copy number and homozygotic regions throughout the genome the data were
 103 normalized to baseline intensities according to an internal reference model of ChAS software
 104 that comprised 270 HapMap samples and 96 other healthy subjects from BioServe
 105 Biotechnologies (BioServe, Beltsville, USA). CNVs and LOH regions were mapped according
 106 to the human reference sequence version GRCh37/hg19. The number of consecutive probes
 107 required defining each deletion or duplication and was limited to a minimum of 25/ 50
 108 consecutive probes, respectively. The size filter for the identification of each LOH segment
 109 was > 2Mb containing at least 50 consecutive SNPs with homozygous genotypes. To keep
 110 only copy-neutral LOH those segments overlapping deletion-type CNVs were excluded for the
 111 analysis. After variant detection by ChAS, CNV and LOH distribution per subjects and
 112 chromosomes were analyzed using Plink v.1.07³⁶.

113 ***Determination of CNV and LOH regions***

114 Plink was used to evaluate the recurrence of common CNVs, CNV regions (CNVRs), which
 115 are the union of overlapping CNVs among subjects³⁷. Duplications and deletions were
 116 analyzed separately and classified as gain- or loss-type CNVRs. Similar Plink tools were used
 117 to access the loss of heterozygosity segments shared between SLE patients and controls.

118 ***Rare CNVs***

119 *A four-step pipeline was created for the identification of rare variants (population*
 120 *frequency < 1%) based on the Brazilian/HapMap population frequencies, public databases*
 121 *of genomic variants and CNV detection by two different algorithms: ChAS and Nexus Copy*
 122 *Number 8.0 (BioDiscovery, CA, USA) (Figure S1).*

123 ***CNVs located in genes with functional relevance to SLE***

124 Using the *cytoregion* tool from ChAS software, three gene lists were created: (1) genes
 125 overlapping deletion or duplication-type CNVs described in association with SLE, (2) genes
 126 previously associated to SLE in linkage analysis and/or in GWAS, (3) genes related to
 127 autoimmunity. For this analysis, the number of consecutive markers for detecting deletions and
 128 duplications that overlap the genes included in the three lists was reduced from 25/50 to 15/15.

129 ***Validation of CNVs using target-specific methodology***

130 For all copy number genes selected for validation by target-specific methodology, we designed
 131 primers using Primer3Plus³⁸, verified their specificity with *in silico* PCR tool available from
 132 UCSC Genome Browser³⁹, and purchased them from Eurofins Genomics (Louisville, KY,
 133 USA) or Sigma-Aldrich (St Louis, MO, USA). Primer sequences are listed in Table S1. *FOXP2*
 134 or *PAX6* were used as reference genes for diploid copy number.

135 The CNVR encompassing the *ADAM3A* gene was selected for validation using quantitative
 136 real-time PCR (qPCR) in a larger case-control cohort. In addition to *ADAM3A* target, an
 137 individual CNV assay was conducted for *FCGR3B* gene since its coverage is very low in the
 138 Cytoscan HD chip (Table S3). The *ADAM3A* and *FCGR3B* gene copy number genotyping was
 139 performed by SYBR Green-based genomic qPCR in cases (n = 135) and controls (n = 200)
 140 using the StepOne Plus Real-Time PCR System according to the manufacturer's protocol
 141 (Applied Biosystems, CA, USA). All experiments were designed using technical triplicates for

each sample. The reference and target genes for each sample were ran in the same 96-well plate to avoid introducing experimental bias. The copy number of the target gene in each test sample was determined by the $\Delta\Delta\text{CT}$ -based relative quantification method⁴⁰.

Droplet digital PCR (ddPCR) was used as a target-specific methodology to validate three heterozygous deletions (CN = 1) overlapping the *CFHR4*, *CFHR5*, *HLA-DPB2* genes, and a heterozygous duplication (CN = 3) encompassing the *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* and *ST8SIA11* genes (Table S3). The ddPCR experiments were performed according to the manufacturer's protocol (Bio-Rad Laboratories, CA, USA). As the initial step, we treated all genomic DNA samples with *HindIII* restriction enzyme for 2 h at 37 °C and then proceed to EvaGreen ddPCR assay. We calculated copy number using the QuantaSoft Pro software (Bio-Rad Laboratories, CA, USA). The error reported for a single well was the Poisson 95% confidence interval (95% CI). We used the automated clustering analysis for both target and reference and then calculated the final copy number as two times the ratio of target concentration versus reference concentration. A reference sample, expected to be in a diploid status in both target and reference genes was used as an internal control of the reactions.

Ancestry inference

Due to tri-hybrid composition of the Brazilian population, a panel of 345 ancestry informative markers based on SNP data from the array was used to infer the proportion of European, African and Amerindian ancestries of each SLE and control subjects⁴¹. These estimates were used as covariates in logistic regression models for the association between CNV and LOH with SLE susceptibility.

Statistical analysis

Statistical analyses were performed using the computing environment R version 3.1.1 (<http://www.r-project.org/>). The call and size of CNV/LOH was compared between SLE

patients and controls using logistic regression models adjusted for sex and ancestry covariates, while Student's t-test was performed to compare the call and size of CNV/LOH per chromosome between case and control groups.

RESULTS

Characteristics of CNVs

Genomic screening using the Cytoscan HD array found 447 CNVs in 23 patients and 2652 CNVs in 110 controls. An average of 19 CNVs were identified per patient (SD = 10, range 5–55 CNVs), while in controls the average was 24 CNVs (SD = 11, range 7–74 CNVs) (Figure S2). Deletions were 1.6–3.2 times more frequent than duplications, corresponding to 62% of CNVs identified in controls and 76% in SLE patients; however the average deletion size was 3.2–4.0 smaller than the average duplication size both control and case groups, respectively (Figure 1A).

Case-control comparisons with logistic regression models adjusted for sex and ancestry showed that duplications were present in smaller number ($p = 0.001$, OR = 0.7 [95% CI, 0.6–0.9]) in SLE patients than in healthy subjects. No evidence for differences in the total number ($p = 0.552$, OR = 1.0 [95% CI, 0.9–1.0]) and size ($p = 0.341$, OR = 1.0 [95% CI, 1.0–1.0]) of deletions were observed between case and control groups (Figure 1B).

CNVs were detected in all autosomes and in the X chromosome. In both groups, the X chromosome showed the highest number of CNVs, representing an average of 10 X-linked CNVs per SLE patient (SD = 4, range 0–20 CNVs) and 6 CNVs per control (SD = 5, range 0–29 CNVs). On the other hand, the larger CNVs were concentrated on chromosome 14, and showed more CNVs ($p = 0.020$) of larger size ($p = 0.004$) in SLE patients compared to controls. Chromosome 1 also showed a higher number of CNV calls per patient compared to controls ($p = 0.041$) (Figure 2).

Synergic effect of deletion in the FCGR3B and ADAM3A genes

In the CNVRs analysis performed from the Cytoscan HD array data (23 SLE and 110 controls), a deletion ($CN < 2$) partially encompassing the *ADAM5* gene and entirely overlapping the *ADAM3A* gene was identified as a potential candidate for increased susceptibility to SLE ($p = 0.0352$; $OR = 3.3$ [95% CI, 1.0–14.1]) (Figure S3, Table S2). Based on this result, we designed a qPCR assay using the *ADAM3A* as a target gene in order to validate the association in a larger population of SLE patients ($n = 135$) and controls ($n = 200$). The association of the deletion in the *ADAM3A* gene with SLE was not replicated using qPCR in the larger sample population ($p = 0.99$, $OR = 1.0$ [95% CI, 0.6–1.8]). However, in agreement with our expectations we observed that *ADAM3A* duplication was statistically lower in SLE patients than in controls ($p = 1.23 \times 10^{-2}$, $OR = 0.2$ [95% IC, 0.1–0.7]) (Figure S4), suggesting that gains in the *ADAM3A* gene are a protective factor for the development of SLE.

In an independent CNV assay, we showed that deletion in the *FCGR3B* gene is associated with increased susceptibility to SLE ($p = 1.66 \times 10^{-3}$, $OR = 3.6$ [95% CI, 1.7–8.3]), while duplication in this gene is a protective factor for the development of SLE ($p = 1.55 \times 10^{-4}$, $OR = 0.2$ [95% CI, 0.1–0.4]) (Figure S4). Evaluating the simultaneous presence of the deletion in both *FCGR3B* and *ADAM3A* loci, we observed that their joint presence contributes to the increase of susceptibility to SLE ($p = 1.4 \times 10^{-4}$, $OR = 5.9$ [95% CI, 2.5–15.9]) when compared to a deletion only in the *FCGR3B* gene. This suggests a genetic interaction between these loci in SLE. The values presented above correspond to the logistic regression models adjusted for sex and African component. The adjusted models for the European and Amerindian components show similar results to the African one (Figure 3).

Rare CNVs and CNVs overlapping genes with functional relevance to SLE

We applied the four-step pipeline to identify rare CNVs in the set of 447 CNVs detected in the 23 SLE patients. Comparing all CNVs identified in SLE patients with the 2652 CNVs reported for the control group, we identified 88/447 variants not present in the control subjects, i.e. exclusive CNVs of SLE. After filtering using population data of healthy subjects from the HapMap project and then the Database of Genomic Variants, we identified 67/88 CNVs and 49/67 CNVs as exclusive to our SLE sample. As a final step, we used an alternative algorithm (Nexus Copy Number) to detect CNVs in SLE patients. Considering only variants identified using both ChAS and Nexus Copy Number software, our four-step pipeline resulted in the detection of 21 rare CNVs that are exclusive to our SLE sample (Table 1).

Gene list analysis was used to evaluate the presence of genes with functional relevance to SLE within the interval of CNV, and revealed CNVs overlapping 8/153 genes in SLE group (Table 2).

Based on population frequency of CNV, deletions in *CFHR4*, *CFHR5*, *STAT4* and *HLA-DPB2* genes identified using Cytoscan HD array in the rare and functional relevant CNVs were selected for validation by ddPCR. In addition to these selected CNVs, a rare duplication of 649 Kb in size encompassing five genes (*LDHB*, *KCNJ8*, *ABCC9*, *CMAS* and *ST8SIA11*) was also selected for target-specific validation (Figure S3). The genotypes were confirmed for 4/5 CNVs selected for ddPCR validation. Therefore, we described heterozygous deletions (CN = 1) in *CFHR4*, *CFHR5*, *HLA-DPB2* genes, and a heterozygous duplication (CN = 3) encompassing the *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* and *ST8SIA11* genes (Figure 4). Heterozygous deletion in the *STAT4* gene was not confirmed by ddPCR.

Characteristics of LOH

A total of 184 copy-neutral LOH segments (i.e. two alleles were present) in 23 SLE patients were identified in autosomal chromosomes. The distribution of the LOH segments per individual was asymmetrical, resulting in an average of 8 LOH segments per SLE patient (SD = 3, range 2–30 LOH), and the average length of LOH was 29 Mb (SD = 40, range 6–184 Mb) (Figure S2). No significant differences between the call ($p = 0.109$, OR = 1.1 [95% CI, 1.0–1.3]) and length ($p = 0.341$, OR = 1.0 [95% CI, 1.0–1.0]) of LOH were observed in SLE patients compared to controls (Figure S5).

LOH segments were found on all chromosomes, except in chromosomes 18 and 22. SLE patients showed higher call and length of LOH in chromosome 6 ($p = 0.045$, $p = 0.027$) and in chromosome 12 ($p = 0.008$, $p = 0.013$) compared to controls (Figure 2C–D). Furthermore, 163 autosomal segments and 86 segments in X chromosome were shared by two or more subjects. From these, six were shared exclusively among SLE patients (Table S4). Additionally, we reported other 17 other LOH segments exclusive for SLE patients when compared with Brazilian controls and European (CEU) and African (YRI) HapMap populations (Table 3).

CNV and LOH profiles in SLE

In order to merge the copy number and loss of heterozygosity data, the circular representation of these variants is shown in the Figure 5. This representation highlights higher (X chromosome) and lower (chromosome 18) chromosomal density of variants in the genome. In addition, it shows the genomic location of rare CNVs and LOH segments exclusive to SLE.

DISCUSSION

Despite remarkable progress in the identification of loci-specific CNVs associated with SLE, questions regarding the role of structural variation in genetic variability and SLE susceptibility have remained^{19,42,43}. Here we present the first genome-wide CNV/LOH study of SLE patients

in a tri-hybrid population. As crucial step to obtain reliable association results for admixed cohorts^{41,44}, we performed adjustment for population stratification for all CNV and LOH statistical comparisons between case and control groups. We reported the synergistic effect of common multi-allelic CNVs increasing the risk for SLE compared to the variation in a single locus, further supporting the involvement of multi-loci deletions in the etiology of the disease. Our study incorporates two CNV calling algorithms and added data of other healthy subjects in external populations and databases with the aim of increasing stringency and improving the sensitivity and specificity of detecting rare CNVs throughout the genome.

The higher total number of deletions in whole-genome in relation to duplications and the smaller average size of losses compared to gains found in this study for both case and control groups, were also reported for Korean women SLE patients and their respective controls¹⁹, as well as for healthy subjects from the Database of Genomic Variants¹³. Additionally, our data showed that the total burden of duplications in genome is significantly high in controls compared to SLE patients, supporting that the presence of multiple gains of DNA segments through the genome may be a protective factor for SLE development. Taking together, these findings that found no evidence of differences in CNV distribution in disease and non-disease cohorts or regarding the burden of gains as a protective factor against the disease development suggest that the evolution history of populations over the generations exerts a stronger impact in the pattern of distribution of structural variants through the whole-genome compared to the presence of pathologic conditions. The excess of X-linked CNVs observed here also corroborates methodologically similar reports, suggesting that there is predominance of variants in the X chromosome over the other chromosomes for both disease and control groups^{29,45}. This underscores the need to search for loci-specific CNVs as causal variants in the

susceptibility and severity of complex diseases for CNV-phenotype associations, instead of focusing in the total burden of structural variants.

The putative association of low copy number of *FCGR3B* gene with an increased risk for SLE⁴⁶ previously identified in African, Chinese and Caucasian populations was replicated in our Brazilian cohort. The data corroborates the association of this gene with the disease even in an admixed population such as the one presented here. As *FCGR3B* is involved in the recruitment of polymorphonuclear neutrophils to sites of inflammation and clearance of immune complexes⁴⁶, losses in this gene could result in the reduction of neutrophil trafficking to inflammatory lesions and in a decrease in the ability to control immune response⁴⁷. In addition to the 3.6-fold increase in the susceptibility of developing SLE observed in the single *FCGR3B* deletion, an additive effect was observed when there was in a deletion of both *FCGR3B* and *ADAM3A* genes, leading to a 5.9-fold increase risk for SLE. Similar synergistic effect of losses in three other loci encompassing the *RABGAP1L*, *C4*, and a region on chromosome 10q21 with no genes in the interval, which resulted in a 5.5-fold increase in the risk for developing SLE compared to deletions in any loci¹⁹. The observations are in agreement with the hypothesis that the combined effect of multiple dosage-sensitive genomic regions may lead to the predisposition to the disease.

Previous reports associated SLE with SNPs¹⁰ and CNVs mapped in the major histocompatibility complex region, i.e., duplications in *HLA-DRB5*²⁰ and deletions in complement 4 (C4) gene⁴⁸. Here we report for the first time identified in SLE patients three heterozygous deletions overlapping HLA (*HLA-DPB2*) and complement-related genes (*CFHR4* and *CFHR5*). The deletions were identified using microarray screening followed by target-specific validation. Although the ability to attribute pathogenicity to a particular CNV remains limited, deep analysis of candidate risk loci harboring losses in the *HLA-DRB2* and

other HLA genes could provide one possible update for the unexplained genetic disease susceptibility. SNPs in complement factor H (CFH) genes, a key regulator of the alternative complement pathway, have also been associated with SLE⁴⁹. Our findings suggest new insights into the pathogenic mechanisms of complement factor H in SLE involving changes in copy number in the factor H-related genes, highlighting the homeostatic balance between *CFH*-*CFHR* genes as critical to maintain regulation in complement activation⁴⁹. Since *CFHR4* gene plays a key role in regulating complement activation and opsonization on biological surfaces by interacting with C-reactive protein⁵⁰, deletion in this gene could lead to reduced protein binding and thus would limit its ability to inhibit inflammation, facilitating SLE development. The involvement of *CFHR5* in renal diseases, e.g. *CFHR5* nephropathy⁵¹, has particular interest in view of the SLE patient who present the rare deletion in *CFHR5* here identified has lupus nephritis, suggesting that losses in this gene may compromise renal functioning and indicating that this variation may be related to lupus nephritis development.

Regarding the whole-genome LOH profile in SLE, our data revealed no evidence that the total number and length of LOH segments are different between case and controls. As LOH pattern is still poorly studied in autoimmune diseases, it is difficult to compare the present study with others⁵². For other complex disorders, several studies reported inconsistent case-control association of LOH burden with increased risk for the disease, e.g., schizophrenia^{53,54} and Alzheimer's disease⁵⁵⁻⁵⁷, leading to a lack of consensus about the cumulative effect of multiple regions of homozygosity in disease susceptibility. Aiming to evaluate the contribution of specific loci in homozygous state for SLE pathogenesis, we used an approach based on the identification of LOH segments present exclusively in SLE patients. The lack of reference databases reporting LOH data for healthy or pathologic conditions⁵⁸ was circumvented by using external populations from HapMap project in addition to the Brazilian controls as criteria

for the exclusion of LOH regions commonly present in healthy subjects. In view of genomic regions where rare LOH occur may be enriched for loci with critical function ²¹, the set of exclusive LOH regions and segments found in SLE phenotypes revealed by our approach represents the first step in LOH analyzes attempting the identification of candidate loci harboring recessive mutations and risk haplotypes for SLE.

Here we show that novel susceptibility loci for SLE can be identified using large-scale evaluation of CNVs and LOH. The detection of rare and common CNVs in functionally relevant genes may elucidate how lupus is triggered and clarify the relationship between clinical manifestations and biological pathways that underlie disease progression. The evaluation of the fine-scale architecture of CNV regions, as well as the prediction of pathogenicity of long segments encompassing several variants found in homozygosity, may contribute to understanding how risk loci harboring CNVs and LOH segments acts on the etiology of SLE.

Acknowledgments

This research was supported by grants from *Fundação de Amparo à Pesquisa do Estado de São Paulo* (FAPESP, Proc No. 2011/23794-7), *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, *Fundação de Apoio ao Ensino, Pesquisa e Assistência do HCFMRP-USP*, and *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq). FBB (2013/17062-9, 2016/10306-8) is currently supported by a research fellowship from FAPESP, while ALS (312547/2009-9), EAD (304931/2014-4) and VLGSL (304455/2012-1) are supported by CNPq/Brazil.

We are grateful to Dr Íscia L Cendes (Faculty of Medical Sciences, State University of Campinas) who provided the healthy samples to Cytoscan HD array and Brazilian Biosciences National Laboratory (Campinas, SP) for technical support.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Supplementary information is available at European Journal of Human Genetics's website.

References

1. Cui Y, Sheng Y, Zhang X: Genetic susceptibility to SLE: recent progress from GWAS. *Journal of autoimmunity* 2013; **41**: 25-33.
2. Mohan C, Putterman C: Genetics and pathogenesis of systemic lupus erythematosus and lupus nephritis. *Nature reviews Nephrology* 2015; **11**: 329-341.
3. Margery-Muir AA, Bundell C, Nelson D, Groth DM, Wetherall JD: Gender balance in patients with systemic lupus erythematosus. *Autoimmun Rev* 2017; **16**: 258-268.
4. Pons-Estel GJ, Alarcón GS, Scofield L, Reinlib L, Cooper GS: Understanding the epidemiology and progression of systemic lupus erythematosus. *Seminars in arthritis and rheumatism* 2010; **39**: 257-268.
5. Kamen DL: Environmental influences on systemic lupus erythematosus expression. *Rheumatic Disease Clinics of North America* 2014; **40**: 401-412.
6. Yang ML, Gee AJP, Gee RJ *et al*: Lupus autoimmunity altered by cellular methylation metabolism. *Autoimmunity* 2012.
7. Rhodes B, Vyse T: The genetics of SLE: an update in the light of genome-wide association studies. *Rheumatology* 2008; **47**: 1603-1611.
8. Amur S, Parekh A, Mummaneni P: Sex differences and genomics in autoimmune diseases. *Journal of autoimmunity* 2012.
9. Alarcon-Segovia D, Alarcón Riquelme ME, Cardiel MH *et al*: Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in

- 376 1,177 lupus patients from the GLADEL cohort. *Arthritis & Rheumatism* 2005; **52**:
 377 1138-1147.
- 378 10. Armstrong D, Zidovetzki R, Alarcn-Riquelme M *et al*: GWAS identifies novel SLE
 379 susceptibility genes and explains the association of the HLA region. *Genes and*
 380 *immunity* 2014.
- 381 11. Sun C, Molineros JE, Looger LL *et al*: High-density genotyping of immune-related loci
 382 identifies new SLE risk variants in individuals with Asian ancestry. *Nat Genet* 2016;
 383 **48**: 323-330.
- 384 12. Alarcon-Riquelme ME, Ziegler JT, Molineros J *et al*: Genome-Wide Association Study
 385 in an Amerindian Ancestry Population Reveals Novel Systemic Lupus Erythematosus
 386 Risk Loci and the Role of European Admixture. *Arthritis & rheumatology (Hoboken,*
 387 *NJ)* 2016; **68**: 932-943.
- 388 13. Zarrei M, MacDonald JR, Merico D, Scherer SW: A copy number variation map of the
 389 human genome. *Nature Reviews Genetics* 2015.
- 390 14. Yih Chen J, Ling Wu Y, Yin Mok M *et al*: Effects of complement C4 gene copy number
 391 variations, size dichotomy, and C4A deficiency on genetic risk and clinical presentation
 392 of systemic lupus erythematosus in east asian populations. *Arthritis & Rheumatology*
 393 2016; **68**: 1442-1453.
- 394 15. Lee YH, Bae SC, Seo YH *et al*: Association between FCGR3B copy number variations
 395 and susceptibility to autoimmune diseases: a meta-analysis. *Inflammation research :*
 396 *official journal of the European Histamine Research Society [et al]* 2015; **64**: 983-991.
- 397 16. Mamtani M, Rovin B, Brey R *et al*: CCL3L1 gene-containing segmental duplications
 398 and polymorphisms in CCR5 affect risk of systemic lupus erythematosus. *Annals of the*
 399 *rheumatic diseases* 2007.

- 400 17. Garcia-Ortiz H, Velazquez-Cruz R, Espinosa-Rosales F, Jimenez-Morales S, Baca V,
 401 Orozco L: Association of TLR7 copy number variation with susceptibility to childhood-
 402 onset systemic lupus erythematosus in Mexican population. *Ann Rheum Dis* 2010; **69**:
 403 1861-1865.
- 404 18. Zhou XJ, Cheng FJ, Lv JC *et al*: Higher DEFB4 genomic copy number in SLE and
 405 ANCA-associated small vasculitis. *Rheumatology (Oxford, England)* 2012; **51**: 992-
 406 995.
- 407 19. Kim JH, Jung SH, Bae JS *et al*: Deletion variants of RABGAP1L, 10q21. 3, and C4 are
 408 associated with the risk of systemic lupus erythematosus in Korean women. *Arthritis &*
 409 *Rheumatism* 2013; **65**: 1055-1063.
- 410 20. Wu L, Guo S, Yang D *et al*: Copy number variations of HLA-DRB5 is associated with
 411 systemic lupus erythematosus risk in Chinese Han population. *Acta biochimica et*
 412 *biophysica Sinica* 2014; **46**: 155-160.
- 413 21. Szpiech ZA, Xu J, Pemberton Trevor J *et al*: Long Runs of Homozygosity Are Enriched
 414 for Deleterious Variation. *The American Journal of Human Genetics* 2013; **93**: 90-102.
- 415 22. Browning BL, Browning SR: Improving the accuracy and efficiency of identity-by-
 416 descent detection in population data. *Genetics* 2013; **194**: 459-471.
- 417 23. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg NA, Li JZ: Genomic
 418 patterns of homozygosity in worldwide human populations. *The American Journal of*
 419 *Human Genetics* 2010; **91**: 275-292.
- 420 24. Bennett RJ, Forche A, Berman J: Rapid mechanisms for generating genome diversity:
 421 whole ploidy shifts, aneuploidy, and loss of heterozygosity. *Cold Spring Harbor*
 422 *perspectives in medicine* 2014; **4**: a019604.

- 423 25. Harville HM, Held S, Diaz-Font A *et al*: Identification of 11 novel mutations in 8 BBS
424 genes by high-resolution homozygosity mapping. *Journal of medical genetics* 2009.
- 425 26. Bocquet B, Marzouka NaD, Hebrard M *et al*: Homozygosity mapping in autosomal
426 recessive retinitis pigmentosa families detects novel mutations 2013.
- 427 27. Gamsiz ED, Viscidi EW, Frederick AM *et al*: Intellectual disability is associated with
428 increased runs of homozygosity in simplex autism. *Am J Hum Genet* 2013; **93**: 103-
429 109.
- 430 28. Gandin I, Faletra F, Faletra F *et al*: Excess of runs of homozygosity is associated with
431 severe cognitive impairment in intellectual disability. *Genetics in medicine : official
432 journal of the American College of Medical Genetics* 2015; **17**: 396-399.
- 433 29. Singh N, Traisak P, Martin KA, Kaplan MJ, Cohen PL, Denny MF: Genomic alterations
434 in abnormal neutrophils isolated from adult patients with systemic lupus erythematosus.
435 *Arthritis research & therapy* 2014; **16**: R165.
- 436 30. Shriner D, Adeyemo A, Ramos E, Chen G, Rotimi CN: Mapping of disease-associated
437 variants in admixed populations. *Genome Biol* 2011; **12**: 223.
- 438 31. Cardon LR, Palmer LJ: Population stratification and spurious allelic association. *The
439 Lancet* 2003; **361**: 598-604.
- 440 32. Liu J, Lewinger JP, Gilliland FD, Gauderman WJ, Conti DV: Confounding and
441 heterogeneity in genetic association studies with admixed populations. *American
442 journal of epidemiology* 2013; **177**: 351-360.
- 443 33. Hochberg MC: Updating the American College of Rheumatology revised criteria for
444 the classification of systemic lupus erythematosus. *Arthritis & Rheumatology* 1997; **40**:
445 1725-1725.

- 446 34. Cohen A, Reynolds W, Franklin E *et al*: Diagnostic and therapeutic criteria committee
447 of The American Rheumatism Association. Preliminary criteria for the classification of
448 systemic lupus erythematosus. *Bull Rheum Dis* 1971; **21**: 643.
- 449 35. Miller S, Dykes D, Polesky H: A simple salting out procedure for extracting DNA from
450 human nucleated cells. *Nucleic acids research* 1988; **16**: 1215.
- 451 36. Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome
452 association and population-based linkage analyses. *The American Journal of Human*
453 *Genetics* 2007; **81**: 559-575.
- 454 37. Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human
455 genome. *Nature* 2006; **444**: 444-454.
- 456 38. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA: Primer3Plus,
457 an enhanced web interface to Primer3. *Nucleic acids research* 2007; **35**: W71-W74.
- 458 39. Tyner C, Barber GP, Casper J *et al*: The UCSC Genome Browser database: 2017 update.
459 *Nucleic Acids Research* 2016: gkw1134.
- 460 40. Livak KJ, Schmittgen TD: Analysis of relative gene expression data using real-time
461 quantitative PCR and the 2- $\Delta\Delta$ CT method. *methods* 2001; **25**: 402-408.
- 462 41. Barbosa FB, Cagnin NF, Simioni M *et al*: Ancestry informative marker panel to
463 estimate population stratification using genome-wide human array. *Annals of human*
464 *genetics* 2017.
- 465 42. Yim SH, Jung SH, Chung B, Chung YJ: Clinical implications of copy number variations
466 in autoimmune disorders. *The Korean journal of internal medicine* 2015; **30**: 294-304.
- 467 43. Ptacek T, Li X, Kelley J, Edberg J: Copy number variants in genetic susceptibility and
468 severity of systemic lupus erythematosus. *Cytogenetic and genome research* 2008; **123**:
469 142-147.

- 470 44. Martin LS, Eskin E: Population Structure in Genetic Studies: Confounding Factors and
471 Mixed Models. *bioRxiv* 2016; 092106.
- 472 45. Kidd JM, Cooper GM, Donahue WF *et al*: Mapping and sequencing of structural
473 variation from eight human genomes. *Nature* 2008; **453**: 56-64.
- 474 46. Yuan J, Zhao D, Wu L *et al*: FCGR3B copy number loss rather than gain is a risk factor
475 for systemic lupus erythematosus and lupus nephritis: a meta-analysis. *International*
476 *journal of rheumatic diseases* 2015.
- 477 47. Rahbari R, Zuccherato LW, Tischler G *et al*: Understanding the Genomic Structure of
478 Copy-Number Variation of the Low-Affinity Fcγ Receptor Region Allows
479 Confirmation of the Association of FCGR3B Deletion with Rheumatoid Arthritis.
480 *Human mutation* 2017; **38**: 390-399.
- 481 48. Li N, Zhang J, Liao D, Yang L, Wang Y, Hou S: Association between C4, C4A, and
482 C4B copy number variations and susceptibility to autoimmune diseases: a meta-
483 analysis. *Sci Rep* 2017; **7**: 42628.
- 484 49. Zhao J, Wu H, Khosravi M *et al*: Association of genetic variants in complement factor
485 H and factor H-related genes with systemic lupus erythematosus susceptibility. *PLoS*
486 *genetics* 2011; **7**: e1002079.
- 487 50. Cantsilieris S, White SJ, Richardson AJ, Guymer RH, Baird PN: Comprehensive
488 Analysis of Copy Number Variation of Genes at Chromosome 1 and 10 Loci Associated
489 with Late Age Related Macular Degeneration. *PloS one* 2012; **7**: e35255.
- 490 51. Gale DP, Pickering MC: Regulating complement in the kidney: insights from CFHR5
491 nephropathy. *Disease models & mechanisms* 2011; **4**: 721-726.
- 492 52. Ku CS, Naidoo N, Teo SM, Pawitan Y: Regions of homozygosity and their impact on
493 complex diseases and traits. *Hum Genet* 2011; **129**: 1-15.

- 494 53. Lencz T, Lambert C, DeRosse P *et al*: Runs of homozygosity reveal highly penetrant
495 recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences* 2007;
496 **104**: 19942-19947.
- 497 54. Johnson EC, Bjelland DW, Howrigan DP *et al*: No Reliable Association between Runs
498 of Homozygosity and Schizophrenia in a Well-Powered Replication Study. *PLoS Genet*
499 2016; **12**: e1006343.
- 500 55. Nalls M, Guerreiro R, Simon-Sanchez J *et al*: Extended tracts of homozygosity identify
501 novel candidate genes associated with late-onset Alzheimer's disease. *Neurogenetics*
502 2009; **10**: 183-190.
- 503 56. Sims R, Dwyer S, Harold D *et al*: No evidence that extended tracts of homozygosity are
504 associated with Alzheimer's disease. *American Journal of Medical Genetics Part B:*
505 *Neuropsychiatric Genetics* 2011; **156**: 764-771.
- 506 57. Ghani M, Reitz C, Cheng R *et al*: Association of Long Runs of Homozygosity With
507 Alzheimer Disease Among African American Individuals. *JAMA neurology* 2015; **72**:
508 1313-1323.
- 509 58. Hauck F, Magerus-Chatinet A, Vicca S *et al*: Somatic loss of heterozygosity, but not
510 haploinsufficiency alone, leads to full-blown autoimmune lymphoproliferative
511 syndrome in 1 of 12 family members with FAS start codon mutation. *Clinical*
512 *Immunology* 2013; **147**: 61-68.

Table 1. Description of rare copy number variation identified in systemic lupus erythematosus (SLE) patients from the four-step pipeline application.

Genomic location	Genes	Type	Size (Kb)	SLE ID	Pop freq
chr1:196964969-196987806	<i>CFHR5</i>	Del	23	SLE005	–
chr2:34219274-34313413	–	Del	94	SLE009	< 1%
chr2:36932842-36964864	<i>VIT</i>	Del	32	SLE010	–
chr3:65517399-65595008	<i>MAG11</i>	Dup	78	SLE015	< 1%
chr5:96625630-96865984	–	Del	240	SLE014	–
chr6:152303218-152311895	<i>ESR1</i>	Del	9	SLE003	< 1%
chr7:11983316-12078163	–	Dup	95	SLE005	< 1%
chr7:85522989-85709099	–	Del	186	SLE003	< 1%
chr7:114209440-114215739	<i>FOXP2</i>	Del	6	SLE023	–
chr7:146859536-146868357	<i>CNTNAP2</i>	Del	9	SLE003	–
chr8:30079990-30202711	<i>MIR548O2</i>	Dup	123	SLE018	–
chr9:137354612-137529641	–	Dup	175	SLE023	< 1%
chr12:21796951-22445614	<i>LDHB, KCNJ8, ABCC9, CMAS, ST8SIA1</i>	Dup	649	SLE018	< 1%
chr13:93378273-93386333	<i>GPC5</i>	Del	8	SLE023	–
chr13:94506273-94515199	<i>GPC6, GPC6-AS2</i>	Del	9	SLE023	< 1%
chr13:104521177-104647267	–	Del	126	SLE013	–
chr14:67111441-67148873	<i>GPHN</i>	Dup	37	SLE011	–
chr14:67193022-67217571	<i>GPHN</i>	Del	25	SLE016	< 1%
chr22:25066486-25112325	–	Del	46	SLE006	< 1%
chrX:15723836-15831401	<i>CA5B, INE2, ZRSR2</i>	Dup	108	SLE018	–
chrX:52833688-52904643	<i>XAGE5, XAGE3</i>	Dup	71	SLE017	–

chr = chromosome; *Del* = deletion; *Dup* = duplication; *ID* = identification; *Pop freq* = population frequency.

Table 2. Description of copy number variation identified in systemic lupus erythematosus (SLE) patients overlapping genes with functional relevance to SLE.

Genomic location	CNV			Presence of CNVs (Del/Dup)				
	Genes	Type	Size (Kb)	SLE (n=23)	BRZ (n=110)	YRI (n=89)	CEU (n=85)	Pop freq.
chr1:196827841-196877037	<i>CFHR4</i>	Del	49	1/0	2/1	0/4	0/1	> 1% / > 1%
chr1:196964969-196987806	<i>CFHR5</i>	Del	23	1/0	0/0	1/0	0/0	< 1% / < 1%
chr2:191904170-191921930	<i>STAT4</i>	Del	18	1/0	0/0	0/0	0/0	< 1% / < 1%
chr6:33083147-33095011	<i>HLA-DPB2</i>	Del	12	1/0	0/0	0/0	0/0	< 1% / < 1%
chr6:29837187-29874629	<i>HLA-H</i>	Del	37	1/0	8/4	5/0	2/0	> 1% / > 1%
chr12:99791778-99798721	<i>ANKS1B</i>	Del	7	6/0	25/0	9/0	8/1	> 1% / < 1%
chr8:39247097-39386952	<i>ADAM3A</i>	Del	140	19/0	14/51	9/0	34/10	~50% / ~20%
chrX:153334931-153335831	<i>MECP2</i>	Dup	1	0/7	0/28	0	0	> 1% / > 5%

chr = chromosome; *Del* = deletion; *Dup* = duplication; *ID* = identification; *Pop freq* = population frequency; *BRZ* = Brazilian controls; *YRI* = Yoruba (HapMap project); *CEU* = Utah residents with Northern and Western European ancestry (HapMap project).

Table 3. Description of loss of heterozygosity (LOH) segments exclusive to the systemic lupus erythematosus (SLE) group compared with Brazilian controls, African and European populations from the HapMap project.

Genomic location	Gene count	OMIM gene count	Size (Mb)	SLE ID
chr19:48866213-50927358	127	78	2.1	SLE002
chr4:92342885-94701185	2	1	2.4	SLE006
chr10:85433274-87951821	10	3	2.5	SLE011
chr6:103924199-106577507	8	6	2.7	SLE014
chr4:183503610-190921709	50	20	7.4	SLE016
chr7:27844434-30165896	17	9	2.3	SLE016
chr10:7112075-14412084	33	17	7.3	SLE016
chr12:130805662-133778166	34	21	3.0	SLE016
chr20:3698495-9587591	42	28	5.9	SLE016
chr21:14386012-16673128	13	6	2.3	SLE016
chr2:221534426-235094377	112	70	13.6	SLE017
chr9:7789576-14075095	7	3	6.3	SLE017
chr15:22752398-26129249	110	18	3.4	SLE017
chr21:26092452-34399401	85	25	8.3	SLE017
chr6:144131236-147039527	15	11	2.9	SLE019
chr8:141498752-146292734	113	61	4.8	SLE022
chr12:64797404-66857938	18	12	2.1	SLE022

chr = chromosome; *ID* = identification; *OMIM* = Online Mendelian Inheritance in Man.

Figure 1. Deletion and duplication profiles in the systemic lupus erythematosus (SLE) patients and Brazilian controls. Distribution of copy number variation (CNVs), deletions (del) and duplications (dup) according to the size ranges in SLE patients and Brazilian healthy individuals (BRZ) (A). Notched box plot comparing the distribution of call (B) and size (C) of deletions and duplications in SLE patients and controls. The boxes represent the interquartile range with black lines indicating the median of the data and the leaked circles symbolizing outliers. The notches exhibit a 95% confidence interval around the median. According to Chambers (1983), if there is no overlap between the notches, there is evidence that the differences between the groups evaluated are significant, as we observed in the call and size of duplications between cases and controls.

Figure 2. Chromosome distribution of copy number variation (CNV) in systemic lupus erythematosus (SLE) patients. Violin plot (R package: vioplot v. 0.2) illustrating the distribution of the data and its probability density for call (A) and size (B) of the total number of CNVs per chromosome, and the call (C) and size (D) of LOH segments. The thicker black bars in the center represent the interquartile range, while the thin black line extended from it represents the 95% confidence intervals around the median (white dot).

Figure 3. Copy number variation in *ADAM3A* and *FCGR3B* genes in systemic lupus erythematosus (SLE). The risk of SLE in relation to the copy number (CN) < 2 (deletions at *ADAM3A*, *FCGR3B*, or 2 loci) and CN > 2 (duplications at *ADAM3A*, *FCGR3B*, or 2 loci), was estimated relative to subjects without any variation in both loci (diploid status at 2 loci) as a reference. The point estimates values represent odds ratios with 95% confidence interval according to adjusted logistic regression models for sex and African (A, D), European (B, E) and Amerindian (C, F) ancestries.

Figure 4. Copy number variation genotypes obtained by droplet digital PCR. Charts showing the copy number (CN) for each SLE patient, confirming heterozygous deletions of *CFHR4* in SLE002 patient (A), *CFHR5* in SLE005 patient (B), HLA-DPB2 in SLE019 patient (C), and heterozygous duplication involving the *LDHB*, *KCNJ8*, *ABCC9*, *CMAS* and *ST8SIA1* genes in the SLE018 patient (D). In all cases, the reference gene (*FOXP2*) shows invariable diploid status in the subjects analyzed. The reference sample confirmed the diploid copy number for both target and reference genes.

Figure 5. Circo plot showing the distribution of copy number variation (CNV) in the genome of systemic lupus erythematosus (SLE) patients. The largest circos represent the 22 autosomal chromosomes and the X chromosome according to the GRCh37/hg19 version of the human genome, followed by the set of CNVs identified in SLE patients, rare CNVs in SLE patients, loss of heterozygosity (LOH) segments identified in SLE patients and LOH segments restricted to SLE group. Deletions are represented in red bars, duplications in blue bars and LOH regions in green bars.